REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

Public Reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comment regarding this burden estimates or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188,) Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 12, 2005	Final Technical 25 Jun 99 - 24 Aug 04

4. TITLE AND SUBTITLE
Deep Subwavelength Optical Nanolithography

5. FUNDING NUMBERS
DAAD19-99-1-0196

6. AUTHOR(S)
S. R. J. Brueck

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of New Mexico
Center for High Technology Materials
1313 Goddard SE, MSC04 2710
Albuquerque, NM 87106

8. PERFORMING ORGANIZATION
314661FT

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
39817.1-EL-MUR

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

Interferometric lithography, the interference of a small number of coherent optical beams, is a powerful technique for the fabrication of a wide array of samples. The techniques and limits of interferometric lithography are discussed with particular attention to the smallest scales achievable. With immersion techniques, the smallest pattern size for a single exposure is a half pitch of $\lambda/4n$ were λ is the optical wavelength and n is the refractive index of the immersion material. Currently with a 193-nm excimer laser source and H_2O immersion, this limiting dimension is ~ 34 nm. With nonlinear spatial frequency multiplication techniques, this limit is extended by factors of 1/2, 1/3, etc. - extending well into the nanoscale regime. Interferometric lithography provides an inexpensive, large-area capability as a result of its parallelism. Imaging interferometric lithography provides an approach to arbitrary structures with comparable resolution. Imaging interferometric microscopy provides an alternative super-resolution technique with application to lithographic mask inspection. Numerous application areas include: nanoscale epitaxial growth for semiconductor heterostructures; nanofluidics for biological separations; nanophotonics including distributed feedback and distributed Bragg reflectors, 2D and 3D photonic crystals, metamaterials and negative refractive index materials for enhanced optical interactions.

14. SUBJECT TERMS
Interferometric lithography, immersion lithography, nanoscale epitaxial growth, nanofluidics, nanophotonics, metamaterials.

15. NUMBER OF PAGES
452

16. PRICE CODE
NSP

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Final Report

for the period

June 1, 1999 to August 30, 2004

ARO/MURI:

Deep Subwavelength Optical Nanolithography

Submitted to: Dr. Dwight Woolard
Program Manager
Army Research Office

Grant Number: DAAD19-99-1-0196

Submitted by: Professor S. R. J. Brueck
Principal Investigator
Center for High Technology Materials
University of New Mexico

Date Submitted: May 12, 2005

Abstract

Interferometric lithography, the interference of a small number of coherent optical beams, is a powerful technique for the fabrication of a wide array of samples. The techniques and limits of interferometric lithography are discussed with particular attention to the smallest scales achievable. With immersion techniques, the smallest pattern size for a single exposure is a half pitch of $\lambda/4n$ were λ is the optical wavelength and n is the refractive index of the immersion material. Currently with a 193-nm excimer laser source and H_2O immersion, this limiting dimension is ~ 34 nm. With nonlinear spatial frequency multiplication techniques, this limit is extended by factors of 1/2, 1/3, etc. - extending well into the nanoscale regime. Interferometric lithography provides an inexpensive, large-area capability as a result of its parallelism. Imaging interferometric lithography provides an approach to arbitrary structures with comparable resolution. Imaging interferometric microscopy provides an alternative super-resolution technique with application to lithographic mask inspection. Numerous application areas include: nanoscale epitaxial growth for semiconductor heterostructures; nanofluidics for biological separations; nanophotonics including distributed feedback and distributed Bragg reflectors, 2D and 3D photonic crystals, metamaterials and negative refractive index materials for enhanced optical interactions.

Final Technical Report

Deep Subwavelength Optical Nanolithography

I INTRODUCTION

Nanoscience and nanotechnology are increasingly important research directions because of the exciting new physics and chemistry that become accessible at the nanoscale and because of the accompanying potential for new products and processes that will have a major impact on the world's technological future. The new science involves both the increased importance of surfaces (e.g. increased surface/volume ratio) and of the chemical/steric effects that are the basis of much of molecular biology, as well as the quantum phenomena that become dominant as the material scale becomes comparable to the wavefunctions of elementary excitations: electrons, holes, photons, spin waves, magnetic excitations, and others.

Optical lithography is well established as the manufacturing technology of choice for the semiconductor integrated circuit industry which is today well into the nanoscale with ~ 65-nm gate lengths in large-volume manufacturing production. With recent developments such as immersion lithography, it now appears that the hegemony of optical lithography has a significant duration yet to run, despite the continuing refrain that the end of optical lithography as we know it is neigh upon us. In this context, it is worth noting that predictions of the end of optical lithography have held steady at "two generations out" for roughly the past 30 years. The dual purposes of this article are to discuss the limits to optical lithography on very general grounds and to point out the opportunities for exciting nanoscience research that are made possible by its capabilities.

Any discussion of tools for nanoscience research necessarily involves consideration of the associated costs – in resources and in time. The research community never has resources for everything and must make choices based on affordability and productivity. Thus, for example, e-beam lithography can produce structures at scales smaller than optical lithography and with an almost complete pattern flexibility (within the limits imposed by proximity effects) and is a staple of nanoscience research. Nonetheless, there are many applications that require large areas (many cm^2 of nanopatterned material) where e-beam lithography is not a viable approach as a result of its serial point-by-point writing modality and the consequent long times involved in addressing the large number of pixels in a large-area nanoscale image. A linear pixel resolution of 20 nm corresponds to a density of 2.5×10^{11} cm^{-2}! This issue only gets worse as the patterning gets deeper into the nanoscale. As an aside, this is a powerful argument applying to any technique, such as the various approaches to maskless lithography, that requires storage and transfer of information on an individual pixel basis, the data transfer demands are very difficult, even with today's advanced computing and communication technologies.

Optical lithography, in contrast, is a parallel writing technique. Traditional optical lithography uses a mask-based approach along with optical reduction to ameliorate the demands on both the mask fabrication and on the optical system. Once the mask is fabricated, all of the information on the mask is transferred onto the wafer in the lithography step. The economics of the integrated circuit industry has put a premium on throughput performance. Current lithography tools expose ~ 85 wafers per hour. Each 300-mm diameter wafer contains ~ 125 die with an area of 22x36 mm^2 = 8.2 cm^2. This corresponds to a sustained information data rate onto the wafer of ~ 1 THz!

Nonetheless, the cost of optical lithography remains an issue for nanotechnology both at the research and the early-stage commercialization phases. The impressive and very capable lithography tools used by the integrated circuit industry are very expensive both in initial (~ $20 million) and in operating costs (typical mask set costs for a modern microprocessor exceed $1 million). Clearly these numbers only make sense in a high volume, high product-value manufacturing context, and are out of reach for a typical research venue.

Fortunately, many applications require only a periodic or quasi-periodic pattern and a much simpler laboratory scale technology, interferometric lithography, based on the interference of a small number (most often two) coherent laser beams can produce useful patterns over large areas, and volumes, with considerable, but not total, pattern flexibility, and with dimensions that today are approaching 20-nm scales. Combining multiple interferometric exposures and mix-and-match with lower-resolution, laboratory-scale optical lithography and with limited use of higher-resolution e-beam lithography dramatically expands the available range of patterns.

Nonlinearities at all stages of the patterning process can be used to further extend the spatial scales. In simple terms, the aerial intensity image in interferometric lithography is a simple $[1+\cos(2kx\sin\theta)]$ functional dependence, where $k = 2\pi n/\lambda$ is the optical wavevector in the medium above the resist and θ is the angle of incidence in the same medium. After development of the photoresist, a square wave lineshape is typically achieved, described by the functional dependence $\sum_l \left[\frac{\sin(2lka\sin\theta)}{2lka\sin\theta}\right]\cos(2lkx\sin\theta)$ where a is the critical dimension or *CD* (width of the line). This expression clearly contains spatial frequencies well above those in the aerial image and provides justification for the claim that *there is no fundamental limit to optical lithography* [1]. These nonlinear concepts, discussed in more detail in Section II.4, are likely to extend the reach of optical lithography by another factor of at least two, to scales of ~ 10 nm with a 193-nm source. Further similar extensions by integral factors, 3, 4, … are possible but will be increasingly difficult as a result of alignment, process, uniformity and edge roughness issues. Another direction is the use of directed self-assembly – using molecular moieties as rulers to proceed further into the nanoscale while retaining the capabilities of optical lithography to hierarchically span the range of dimensions from cm's down to ~ 10 nm.

The concepts of interferometric lithography and examples of results are presented in section II starting from the simplest case of two-beam interference. Particular emphasis is placed on the recent re-emergence of an old idea – immersion interferometric lithography. This section concludes with a discussion of the limits of interferometric lithography. In section III, a very brief discussion of the extension of these ideas to a full imaging solution is presented. This is already the topic of many more extensive treatments and cannot be fully developed in the available space. The remainder of the report contains a survey of active research areas for the application of these technologies. Examples are drawn from nanoscale epitaxial crystal growth, nanofluidics/nanobiological applications, nanomagnetics, and nanophotonics. Details are included in the attached articles. The treatment presented here is necessarily very high-level. Nonetheless, the juxtaposition of this wide range of application areas serves to make the major point – optical and especially interferometric lithography are enabling technologies that allow research into, and even manufacturing of, a wide variety of exciting nanoscience and nanotechnology – at an affordable cost that is broadly accessible.

II INTERFEROMETRIC LITHOGRAPHY (IL)

Interferometric lithography is the use of a small number of coherent optical beams incident from different directions on a 2D (thin photoresist film << λ) or 3D (thick photosensitive layer >> λ) to produce an interference pattern whose intensity distribution is recorded in the photosensitive layer and is later transferred (developed) by thermal and chemical processes. In the literature, this concept has been variously referred to as holographic lithography, interference lithography, and interferometric lithography.

Interferometric lithography (IL) is a well-established concept that has largely tracked the application of lasers to lithography [2]-[6]. Early work was based on cw, single mode lasers such as the Ar^+ line at 488 and the Cd^+ line at 442 nm. With the commercial development of i-line lithography (365 nm), attention shifted to the cw Ar^{++} lines at 364 and 352 nm and to the 355-nm third-harmonic of a YAG laser to take advantage of commercial resist developments. As leading-edge commercial lithography shifted to KrF (248 nm) wavelengths, the intracavity doubled Ar-ion line at 244 nm was used. Today, the IC industry is dominantly using ArF excimer lasers (193 nm) and work has been reported both with fifth-harmonic YAG (213 nm) and with line-narrowed ArF lasers at 193 nm. The very small bandwidth tolerance of the high resolution lenses used in modern lithography has forced the development of line-narrowed excimer lasers with sub-pm bandwidth. These lasers have longitudinal coherence lengths of several cm and are more than adequate for interferometric application.

II.1 Two beam interferometric lithography

The basic concept of two-beam interference is shown schematically in Fig. 1. Two coherent optical plane waves with the same polarization are symmetrically incident on a photosensitive layer coated onto a substrate at angles of $\pm\theta$. The intensity pattern in space is simply given by:

$$\begin{aligned} I(x,z) &= |E|^2 \left|\hat{e}_1 e^{ikx\sin\theta} + \hat{e}_2 e^{-ikx\sin\theta}\right|^2 = |E|^2\left[2 + \left(\hat{e}_1\cdot\hat{e}_2^{*} e^{i2kx\sin\theta} + \hat{e}_1^{*}\cdot\hat{e}_2 e^{-i2kx\sin\theta}\right)\right] \\ &= 2|E|^2\left[1+\sigma_{pol}\cos(2kx\sin\theta)\right] \end{aligned} \tag{1}$$

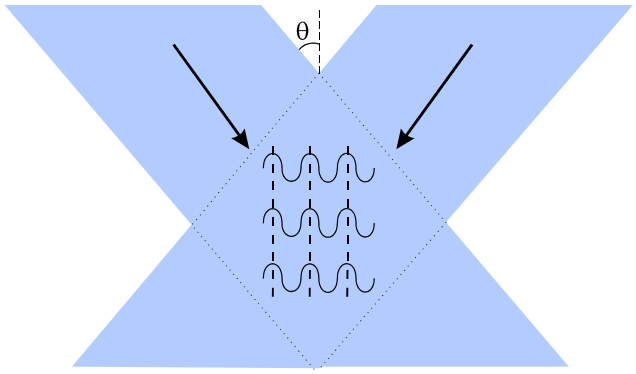


Fig. 1: Schematic representation of two beam interference. Coherent laser beams are symmetrically incident from the left and the right. The period of the interference pattern is $\lambda/[2n\sin\theta] = \lambda/2NA$. The standing wave pattern exists throughout the overlap between the beams as long as this overlap distance is shorter than the longitudinal coherence length of the laser beams and the wafer can be placed anywhere inside this coherence volume.

where $k = 2\pi n/\lambda$ with n the refractive index of the medium (1 for air) and λ the optical wavelength and $\hat{e}_i$ is the unit polarization vector for each wave. The period of the interference pattern along the x-direction is $\lambda/2n\sin\theta$ and σ_{pol} = [1, $\cos(2\theta)$] for (TE, TM) polarization. By Snell's law this period is invariant across any $\theta = \pi/2$ layer boundaries such as the lithography stack; this is not true for the contrast. TE polarization always gives the maximum contrast, while the contrast is reduced for TM polarization as a result of the π-phase shift between the $\hat{e}_x$ and $\hat{e}_z$ contributions to the interference intensity. For evaluating the polarization dependence, it is important to use the propagation

angles inside the photosensitive layer. The contrast reduction for TM polarization becomes significant for steep angles in the resist and will be discussed in detail in the immersion discussion below. For now take $n = 1$; it is relatively easy experimentally to reach $\sin\theta \sim 1$ (for example at a 75° angle of incidence, $\sin\theta = 0.97$ so the limiting period of IL is $\gtrsim \lambda/2$ and extends below 100 nm at a wavelength of 193 nm.

Two experimental arrangements for IL are shown in Fig. 2. The top set-up (a) is a simple corner cube arrangement (Fresnel mirror) where the right and left halves of the beam are folded onto each other using a 90° geometry. This arrangement has been used quite successfully with laser sources with a high transverse coherence such as single-mode TEM_{00} Ar-ion laser beams. Expanding the beam allows quite large areas –especially at periods corresponding to incidence angles θ of ~ 45° where the sizes of the beams incident on the sample and the mirror are equal. At extremely large or small periods, the available area becomes limited by the required size of the folding mirror and alternative experimental configurations are preferable.

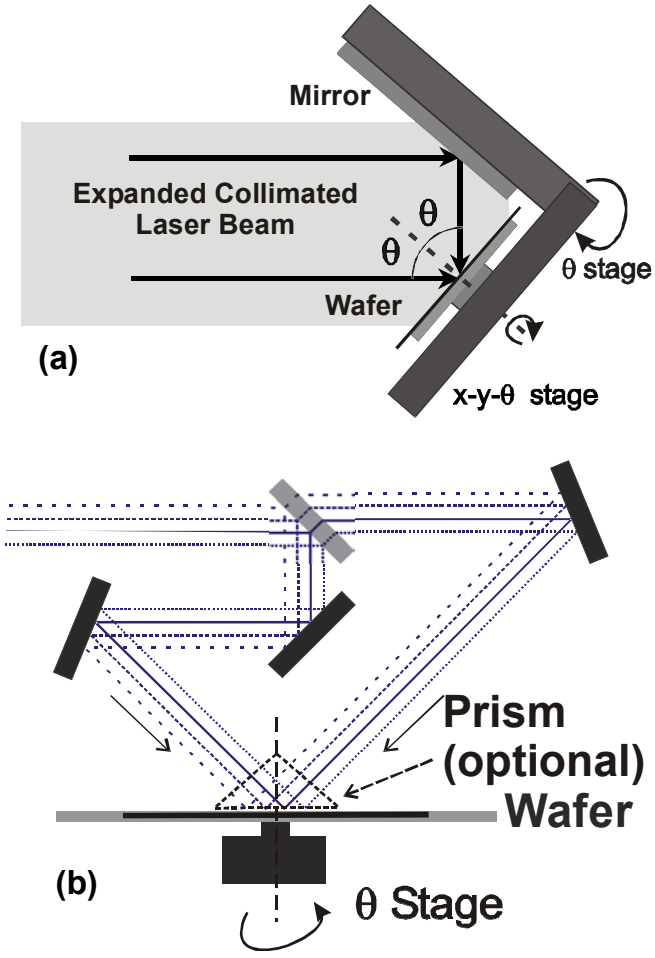


Fig. 2: Experimental arrangements for IL.

Excimer lasers exhibit a large number of transverse modes and consequently a very low transverse coherence. The experimental arrangement in Fig. 2a is not suitable in this case, since the two halves of the beam are not, in general, mutually coherent. The interferometer arrangement shown in Fig. 2b resolves this problem by splitting the input beam in half and folding it onto itself as indicated by the various line styles in the figure; to maintain contrast, any offset has to be within the transverse coherence length. There remains the requirement of a significant longitudinal coherence (narrow spectral width) since the distances from the beamsplitter to the image plane vary for the two beams across the field. Fortunately, today's very narrow linewidth lithography-optimized excimer lasers have linewidths << 1 pm (longitudinal coherence length $\sim \lambda^2/\Delta\lambda \gtrsim 4$ cm, making large-area (at least several cm^2) IL straightforward.

Diffractive beam splitters offer an achromatic interferometric solution that has been used with larger linewidth lasers, for example free-running excimer lasers without line narrowing, to provide large area patterns [7], [8]. The basic idea is to maintain both the angular relationship of the two beams that gives the interferometric pattern and also equal path lengths for the two beams across the pattern area to provide the necessary self-coherence. A grating beam splitter provides the necessary capability to meet both of these requirements. Generally a phase grating is used optimized to result in maximum coupling into the ±1 diffraction orders. These two orders

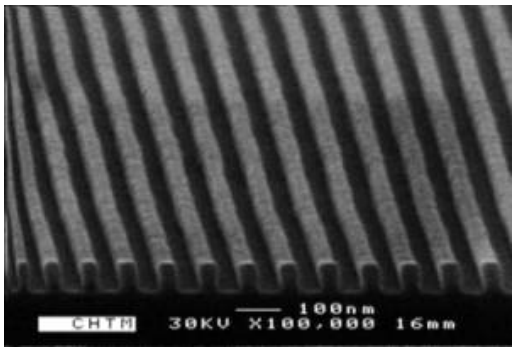


Fig. 3: 45-nm half-pitch pattern in photoresist. (Resist supplied by AZ Electronic Materials)

are then recombined either with a second grating or with a pair of mirrors to provide the final aerial image pattern.

II.2 Immersion interferometric lithography

Recently, there has been resurgence of interest in immersion lithography, the use of a suitable fluid between the imaging lens and the wafer [9]-[12]. This allows an improvement in the optical resolution by a factor of the refractive index of the immersion fluid, provided that the lens is suitably modified. At 193 nm, water is a very attractive immersion fluid with a refractive index of 1.44 at 193 nm [13]. The simple addition of a prism to the interferometric arrangement of Fig. 2b allows investigation of immersion interferometric lithography. In the simplest configuration, a drop of water is placed on the wafer and the wafer and water are raised till the water contacts the uncoated prism and forms a high quality optical component. More elaborate automatic fill and dry systems will be used in manufacturing applications.

An example of a 45-nm half-pitch immersion IL pattern is shown in Fig. 3. After the immersion exposure the sample was baked to activate the chemically amplified resist and developed in a standard basic solution. More details are provided elsewhere [14], [15].

Polarization is a very important parameter at the steep angles in the resist enabled by immersion [16]. For TE polarization, as indicated by Eq. (1), the fringe visibility $[(I_{max} - I_{min})/(I_{max} + I_{min})]$ is unity independent of the half pitch. However, for TM polarization, the fringe visibility decreases as the half-pitch is decreased as a result of competition between the out-of-phase E_x^2 and E_z^2 contributions to the intensity. At an interference angle of 45° in the resist, these two terms are equal and opposite in phase and the visibility goes to zero. For larger interference angles, the line-space pattern is inverted. Fig. 4 shows the simple calculation of the fringe visibility for the parameters indicated. The cusp at ~ 40 nm corresponds to the inversion of the pattern – the interchange of lines and spaces. The minimum half-pitch available with a 193 nm source and water as the immersion liquid is 33.5 nm; if a higher index liquid becomes available, the glass of the lens sets the limit at 31.3 nm. With a higher index glass as the final lens element, smaller half pitch patterns are available. Clearly, the reduction in the fringe contrast and the line:space inversion will require some form of polarization control for imaging these small half-pitch patterns.

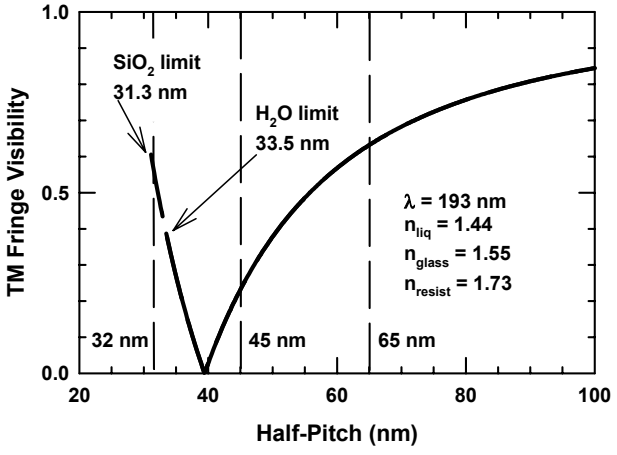


Fig. 4: Aerial image contrast in photoresist as a function of half-pitch.

II.3 Multiple-Exposure interference – Complex Two Dimensional Patterns

Simple two-beam interference produces an array of uniformly spaced parallel lines covering the exposure area, e. g. a

simple grating. By combining multiple such gratings, structures of significant complexity can be constructed [17]. Fourier analysis tells us that any pattern consists of a summation of periodic patterns. A seemingly small, but important, difference between IL and a Fourier expansion is the DC level that accompanies each exposure, *cf.* Eq. (1), since the exposure intensity must be positive for all positions. The consequence is that the DC level increases with the number of exposures and for a large number of exposures, the result is a perfect aerial image of the desired pattern riding atop a large, uniform exposure level. This loss of contrast in the image is an inevitable consequence of summing incoherently related images, and is the same reason that the contrast is lower in incoherent imaging than it is in coherent imaging [18]. Today's photoresists are quite nonlinear, giving an approximately sigmoidal response such that every location above (below) a certain threshold is removed (retained) on development. To a point, this nonlinear response can ameliorate this uniform exposure issue and restore the pattern. However, eventually the uniform exposure becomes too large relative to the image information and pattern details are lost.

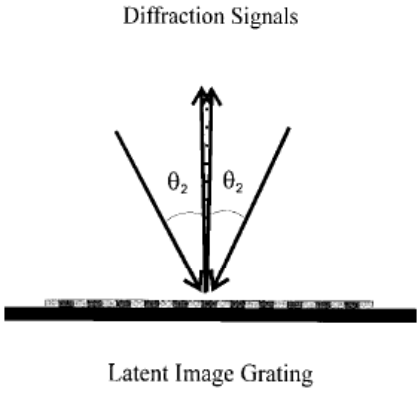


Fig. 5: Experimental configuration for projection moiré alignment. The two beams are incident on the photoresist image produced in the first exposure, or on a predefined grating pattern. The angles are adjusted so that there is only a single fringe across the illuminated area (sets the period), and the sample translation is adjusted so that the fringe is at maximum brightness (sets the phase).

Notwithstanding this complexity limitation, there are many useful structures readily available with multiple exposure techniques. The simplest, and perhaps the most useful, is the square array 2D pattern obtained by carrying out one IL exposure, rotating the sample by 90°, and performing a second exposure. For low exposure doses in a positive photoresist, such that neither exposure alone is sufficient to clear the resist during developing, the result is a 2D array of holes; at higher exposure doses where each exposure clears, the result is a 2D array of posts. For a negative-tone resist, the occurrence of holes and posts are reversed. Details of the pattern, such as the sidewall angle, are sensitive to the exposure dose and need to be optimized in any application [19].

Another simple, but important, application of multiple exposures is to reduce the density of a pattern while retaining the small size capability that is inherent in small-pitch patterns. Again, the idea is quite simple, expose a first pattern at a period $p_1 = p$ with, for example, equal line:space dimensions and a critical dimension (CD) or pattern size of $p/2$, then expose the same area again at a period of $p_2 = p/2$. For a positive photoresist, the result is a pattern with a period of p and a CD of $p/4$. Two constraints that have to be achieved for this approach to be successful are matching of the periods and adjusting the alignment between the two patterns. The linear dimension of the final pattern L, perpendicular to the lines over which the patterns are in phase, depends critically on the deviation of the two periods $\delta = 2p_2\text{-}p_1$ and is given by $L \sim p_1 p_2 / \delta$. Fortunately, by using diffraction from the latent image of the first exposure (or from a predefined grating pattern alignment mark) it is possible to derive a feedback signal to allow setting both the

frequency and the phase of the second exposure with the necessary precision to cover a large area, as is illustrated in Fig. 5 [20]. The basic concept is to monitor the interference between two diffracted beams, one from each input beam over across the exposure area. This can either be done at low intensity over the entire area, or at higher intensity with the sacrifice of a small portion of the final pattern using large area masks. The goal is to set the input angles on the second exposure so that a single fringe of the interference pattern extends over the entire exposure area (this sets the period, e. g. $\delta \to 0$) and then to adjust the phase of that pattern (light or dark – this sets the relative phase of the two patterns). As a result of the moiré spatial amplification of the pitch differential (small pitch mismatches translate into large, easily measured, distances) this alignment technique has general applicability across many applications in nanoscale fabrication.

II.4 Multiple-beam interference – 2D and 3D Patterns

Adding more input beams to an IL exposure is another direction toward increasing the flexibility of the technique and the complexity of the patterns that can be achieved. The simplest example is to use four beams, two oriented at $\pm\theta$ in the x-direction and two at $\pm\theta$ in the y-direction. If orthogonal polarizations are used for the two directions, this is identically equivalent to the process for defining a square array in the previous section, with higher throughput since the two exposures are simultaneous rather than sequential and without the need for a rotation of the wafer.

If all of the interfering beams are inclined at the same angle to the surface normal, the resulting aerial image is again independent of the position of the wafer, however, if the beams are at different angles, the aerial image is no longer invariant along the z-direction and the placement of the wafer becomes important. For a small number of beams, the pattern is periodic along z, as the number of beams increases to a continuum, the periodicity along z tends towards infinity and the familiar result of a single focal plane for an imaging system results. A simple, but useful, example is the addition of a normal incidence beam along with the two offset beams of two-beam interference. Assuming TE polarization, the straightforward generalization of Eq. 1 is:

$$\begin{aligned} I(x,z) &= |E|^2 \left| e^{-ikz} + \left(e^{ikx\sin\theta} + e^{-ikx\sin\theta} \right) e^{-ikz\cos\theta} \right|^2 \\ &= |E|^2 \left[3 + 4\cos(kx\sin\theta)\cos\left[kz(1-\cos\theta)\right] + 2\cos(2kx\sin\theta) \right] \end{aligned} . \tag{2}$$

For z-positions where the second term vanishes, the aerial image is very similar to that of two-beam interference with an additional DC bias. However, at z-positions, such as $z = 0$, where the second term is at a maximum, only alternate lines survive. The result is a set of lines for a 1D exposure with a linewidth comparable to the dense pattern, but with a twice the separation, e.g. a sparse array. For some applications, for example for field-emitter tips for displays, there are important advantages to the sparse pattern [19]. Fig. 6 shows the aerial images for (a) a two-beam, interference at a period of $2p$ from Eq. (1); (b) the three-beam interference for $z = 0$ (spaced at $2p$ but with a linewidth comparable to $p/2$), and (c) the three-beam interference where the second term vanishes (period of p). The smaller pattern feature size for the three beam exposure is evident.

The variation in the *z*-direction also offers the possibility of patterning a full 3D pattern in a single exposure or set of exposures. This possibility will be discussed in more detail in Sect. *IV.4.1* as part of the discussion of photonic crystals.

II.5 Mix-and-Match Approach to Complex Structures

Another straightforward strategy is to use IL to provide large-area periodic patterns and then to use another lithographic technique to add either a small area customization (for example a defect in a photonic crystal) or large area aperiodic structures (for example contact pads and integration with standard microfabrication. In a preliminary investigation of this approach, narrow lines (~ 200 nm) produced by IL were placed atop a sparse array of much larger lines (2 μm) produced by conventional lithography for electrical test structures [21]. Fig. 7 shows the result. Fig. 7(a) shows a SEM of the final structure, the narrow line produced by IL is shown atop the wider pad defined by optical lithography, the next line over misses the pad and has no effect on the measurement. Fig. 7(b) shows the variation of the alignment across the 1-cm wide die. There is a period mismatch along with random errors. The period mismatch, which corresponds to a $\Delta p/p \sim 6\times10^{-5}$, could be eliminated with an adjustment of the IL pitch, the random errors, with a 3σ of only ~ 140 nm, likely arise from placement errors in the optical lithography mask.

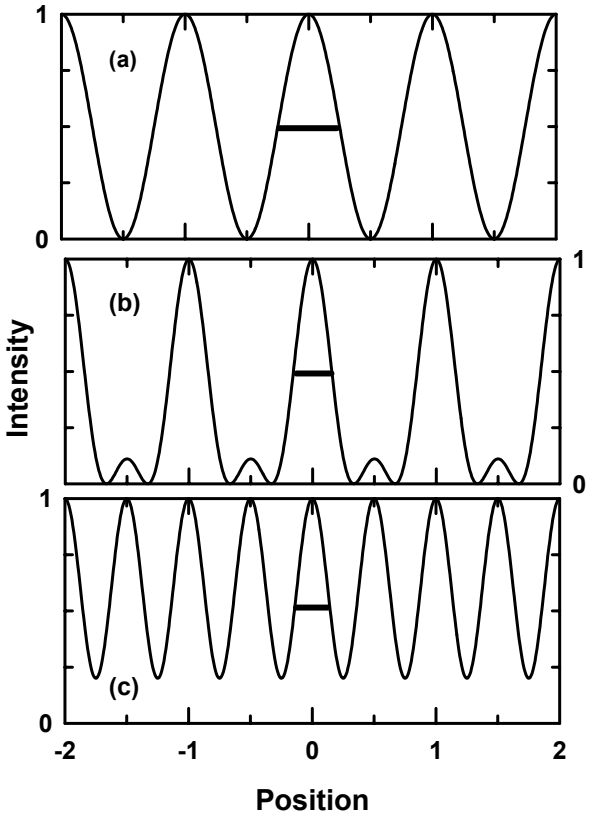


Fig. 6: Aerial image intensity: (a) for a two-beam exposure at twice the period and (b) and (c) at two positions in *z* for a three-beam exposure. The 50% linewidths are indicated. The three beam exposure offers a smaller linewidth with the same period as the two-beam exposure.

Fig. 8 shows another mix-and-match combination of optical and interferometric lithographies. This was an alternative approach to address the sparse pattern issue discussed above in the context of field-emitter displays [19]. For this integration process, a two lithography-step sequence was developed using a sacrificial silicon dioxide layer. A first lithography step, either optical, OL, as in (a) or interferometric, IL, as in (b) was carried out and the results were etched into the sacrificial layer. Then, a second lithography step, using the other lithography technique was carried out to isolate every other hole in the pattern. Alignment was carried out with a grating surrounding the pattern. In (a) a period and phase adjustment very similar to that described above was carried out. In (b) the alignment was performed using standard, pre-printed alignment marks.

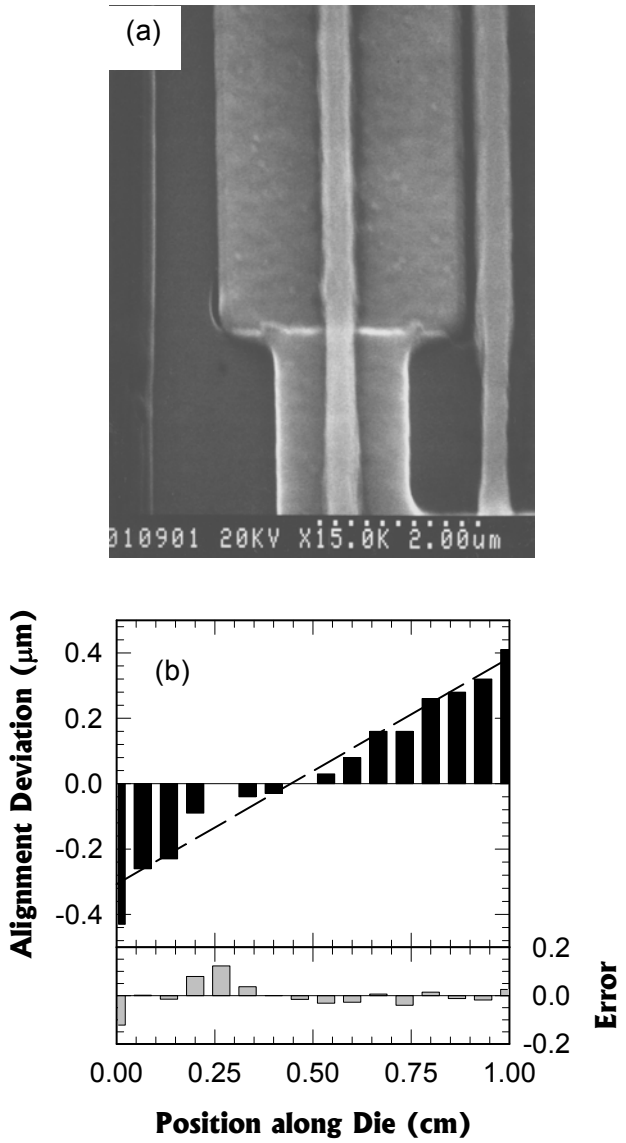


Fig. 7: Demonstration of the integration of OL and IL for an electrical test structure pattern. (a) SEM of the placement of a narrow line defined by IL (200 nm on a 2 μm pitch) atop a wider electrical test structure pattern defined in a separate OL step. (b) Variation of the placement across a 1-cm wide die. The variation consists of a slight misadjustment of the period (~ 6×10^{-5}, dashed line in the fig.) and a random error shown in the bottom part of the figure with a 3σ of 140 nm.

II.6 Nonlinear process in lithography and spatial frequency multiplication

As noted in the introduction, the spatial frequencies of the developed pattern can extend well beyond the $2/\lambda$ bandpass limits of optics as a result of nonlinear processes in the exposure, development and processing – e.g. resulting in transforming a sinusoidal aerial image to a square wave developed resist pattern. Modern resists are engineered to be highly nonlinear – that is to exhibit a sharp threshold between exposed and unexposed regions. Indeed, this is an all important characteristic that has allowed the development of many of the resolution enhancement techniques such as optical proximity correction and phase shift masks that are routinely employed by the integrated circuit industry.

These nonlinear concepts are likely to extend the reach of optical lithography by another factor of at least two, or to scales of ~ 20- to 10-nm with a 193-nm source. Further similar extensions by integral factors, 3, 4, … are possible but will be increasingly difficult as a result of process, uniformity and edge roughness issues.

In the previous section, we showed that the optical system band pass limits the one-dimensional pattern density in a single exposure to ~ λ/NA, while the nonlinearities in the processing result in higher harmonics extending in spatial frequency well beyond the diffraction limits of Eq. 1. These nonlinearities can be used to extend the available pattern density [22], leading to the claim that "there is no fundamental limit to optical lithography – there are only process control, uniformity and surface roughness issues [1]." The key is to apply the nonlinearities <u>*between*</u> exposure steps and combine images in such as way as to reduce or eliminate the lower frequency components, leaving the higher density of structures. This is variously know as spatial period division or spatial frequency multiplication depending on whether the focus is on real space (division) or frequency space (multiplication).

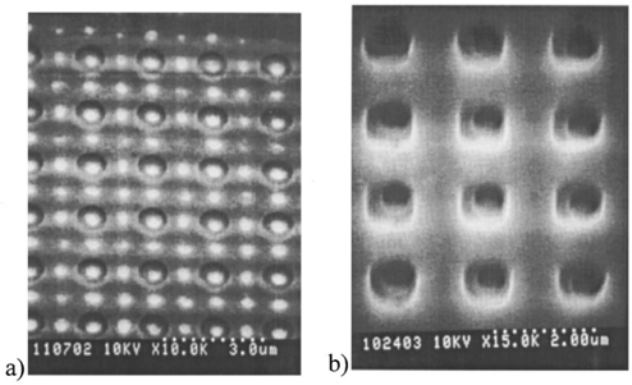


Fig. 8: Sparse array of holes produced by a combination of optical and interferometric lithographies: (a) OL first, IL second; (b) IL first, OL second. In both cases the first pattern

A one-dimensional example of a spatial period division is shown in Fig. 9 to illustrate the general concept. The top panel Fig. 9a shows a photoresist pattern atop a sacrificial layer on a substrate. The period of the pattern is limited by the optics to $\lambda/2NA$; the *CD*, in contrast, is not limited by the optics and is chosen to be less than $\lambda/8NA$. This does not violate any optical limits and is, indeed, common practice in integrated circuit manufacturing where the transistor gate length is less than ½ the pattern period (e.g. 50-nm gates for a 180-nm period in today's microprocessors). In Fig. 9b, the photoresist pattern has been transferred into the sacrificial layer. In Fig. 9c, a second photoresist layer has been spun, exposed and developed. The period and *CD* is the same as in the first case, but the pattern has been shifted by $\lambda/4NA$ to interpolate the second pattern with the first. Finally this pattern is also transferred into the sacrificial layer (Fig. 9d), that layer is then used as a mask to transfer the total pattern into the substrate layers, and finally the sacrificial layer is removed leaving a final pattern at a period of $\lambda/4NA$, twice as dense as the single exposure optical limit.

The mathematics follows simply:

$$\begin{aligned} T(x) &= \mathcal{R}[I_1(x)] \oplus \mathcal{R}[I_2(x)] \\ &= \frac{a}{2}\left\{ \sum_l \frac{\sin\left(\frac{2\pi l NA a}{\lambda}\right)}{\frac{2\pi l NA a}{\lambda}} \left[1+(-1)^l\right] \right\} e^{i\frac{4\pi l NA}{\lambda}x} = a\sum_l \frac{\sin\left(\frac{4\pi l NA a}{\lambda}\right)}{\frac{4\pi l NA a}{\lambda}} e^{i\frac{8\pi l NA}{\lambda}x} \end{aligned} \tag{3}$$

The result is the combination of two exposures with the nonlinearities applied to each exposure independently. The symbol used to represent the combination is left ambiguous; in this case the combination is simple addition. In another case, for example the use of two resist layers sensitive at different wavelengths, it might be multiplication. The $p/4$ phase shift of the second exposure results in a factor of $(-1)^l$ in the Fourier coefficients, with the net result that all of the odd terms cancel leaving a $\sin(x)/x$ pattern at twice the original pitch or half the period.

This spatial frequency multiplication concept applies equally to arbitrary patterns as to the periodic grating pattern discussed here. The parsing of the pattern into two masks becomes a simple coloring problem, with the design rule that no two features on a mask can be closer than twice the final pitch (adjusted for the system magnification). An important difference between the arbitrary and periodic pattern cases is that for a periodic pattern the aerial image contrast is uniform across the pattern and all that is necessary to ensure a uniform final pattern is to arrange for a uniform illumination; in contrast, for an arbitrary image the aerial image contrast inherently varies across the pattern and additional measures, such as optical proximity correction, to create a more uniform image contrast to maintain the uniformity of the nonlinear processes are necessary.

In principle, this process can be repeated multiple times, for integer spatial frequency multiplications – 3, 4, … The alignment requirements, the image uniformity requirements, and the

difficulties of minimizing edge roughness of the resulting patterns will ultimately provide a practical limit [23].

This discussion has so far omitted the important issue of alignment between these multiple exposures. For the periodic pattern case, the image being printed is a large grating, as is well known, the angular resolution of a grating scales as the number of lines in the grating so that alignment schemes based on moiré interference, as discussed above, offer an approach to alignment [24]. Optical aberrations and other contributions to wavefront distortion will introduce positional errors and limit the ability to interpolate patterns. These issues need to be investigated before the ultimate feasibility of these spatial frequency multiplication schemes can be assessed.

II.7 Limits of interferometric lithography

The semiconductor industry maintains a 15 year plan, the International Technology Roadmap for Semiconductors (ITRS) [25], that projects forward and tries to match industry manufacturing requirements (scaled with a Moore's law 0.7 linear dimension shrink every three years) with available and projected technologies. Table I compares the industry requirements as defined in the ITRS with the scales available from interferometric lithography and from imaging interferometric lithography (discussed below).

Table I: ITRS Requirements and (Imaging) Interferometric Lithography Capabilities

Year of Vol. Manufac.	Half-Pitch (nm)	IL Approach	IL Limit (nm) (periodic)	IIL Limit (nm) (arbitrary)
2007	65	Air immersion	48	64
2010	45	H_2O immersion	34	45
2013	32	HI* immersion	27	36
		Air immersion& Freq. Doubling	24	32
2016	22	H_2O immersion & Freq. Doubling	17	23
2019	16	HI* immersion & Freq. Doubling	12	16
		H_2O immersion & Freq. Tripling	12	16

Gray squares are ITRS roadmap requirements.
* HI – High Index immersion (assuming both an immersion liquid index and a glass index of 1.8 @ 193 nm).

For periodic patterns, as used in development of many Si manufacturing processes, the 65-nm hp (half-pitch) requirements are met with IL without the need for immersion techniques. For the 45-nm hp pitch node, water immersion is suitable as demonstrated in Sect. II.2 above. Two alternatives are indicated for reaching the 32-nm hp node: use of a higher-index matching fluid (and a higher-index final glass element as well), here assumed as an index of 1.8, or a frequency-doubling step from the 65-nm node. Of course, frequency doubling from the 45-nm H_2O immersion result is possible as well. Spatial frequency doubling from the 45-nm hp node will access the 22 nm node. Finally, the 16 nm half-pitch node can be accessed either by frequency doubling from the high index version of the 32-nm node, or by frequency tripling from the 45-nm hp H_2O immersion node. This table gives the very simply derived optical limits. As discussed above, issues of line-edge roughness, alignment, and wavefront distortion are significant and must be carefully investigated to determine ultimate feasibility. Nonetheless, this simple analysis is sufficient to justify the claim that *there are no fundamental limits to optical lithography*! Imaging issues (the right most column in Table I) will be briefly discussed in Sect. III.

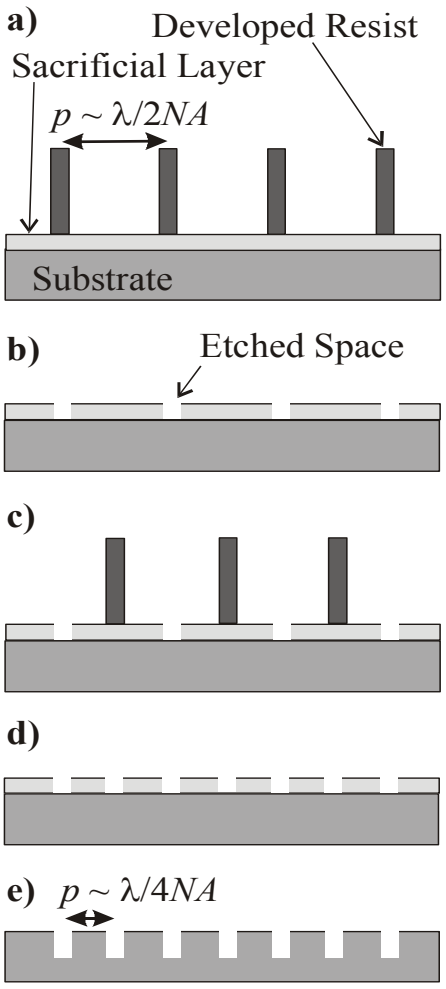


Fig. 9: Illustration of process flow for spatial period division. a) Expose a develop a line:space pattern with the line width < $p/4$ with p the pitch. b) Transfer this pattern into a sacrificial layer. c) Expose and develop a second photoresist layer with the same pattern, but shifted by $p/2$. d) Transfer this pattern into the sacrificial layer. and e) Transfer the composite pattern into the wafer resulting in a spatial frequency doubled pattern at $p/2$.

II.8 Directed self-assembly

While the above statement about the optical limits of lithography is correct, the issues associated with line-edge roughness, photoresist mechanical properties, alignment, and wavefront flatness will get continuously more difficult. Self-assembly is a very attractive alternative for extending the spatial scales to sub-nm dimensions. Directed self-assembly using the properties of nanoscale particles and of molecular moieties is an attractive alternative as the lithographic scale decreases below ~ 20- to 10-nm that both addresses the need for smaller features (extending to the nm regime and below) and retains the exquisitely wide scale – from cm to nm of existing optical lithography [26].

One example of a directed self-assembly process is shown in Fig. 10 which shows ~ 50-nm silica particles assembled into trenches etched into a Si substrate [27], [28]. This is a very flexible process that can be used to produce nanochannels with controlled porosity and even multilevel structures as shown in Fig. 11.

III IMAGING-INTERFEROMETRIC LITHOGRAPHY

Optical lithography is the only technology that has been used for volume manufacturing throughout the history of the integrated circuit. The parallel nature of lithography in which an entire image is transferred in a single exposure is the primary reason for this dominance. It now seems apparent that optical lithography – with the extension of water immersion - will continue as the primary manufacturing technology at least through the 45-nm half-pitch node.

Arbitrary images, in principle containing as many spatial frequencies as there are pixels in the image, are significantly more difficult to produce than the simple, single spatial frequency periodic images of IL. The scale of features produced is smaller than the resolution limits of traditional imaging, and many very sophisticated techniques have been introduced to extend the range of feature sizes that can be produced. Optical proximity correction, adding additional sub-resolution features to the mask to impact the diffraction limited image, off-axis illumination, and phase shift masks have been extensively discussed in the literature and are not covered here [29]-[31].

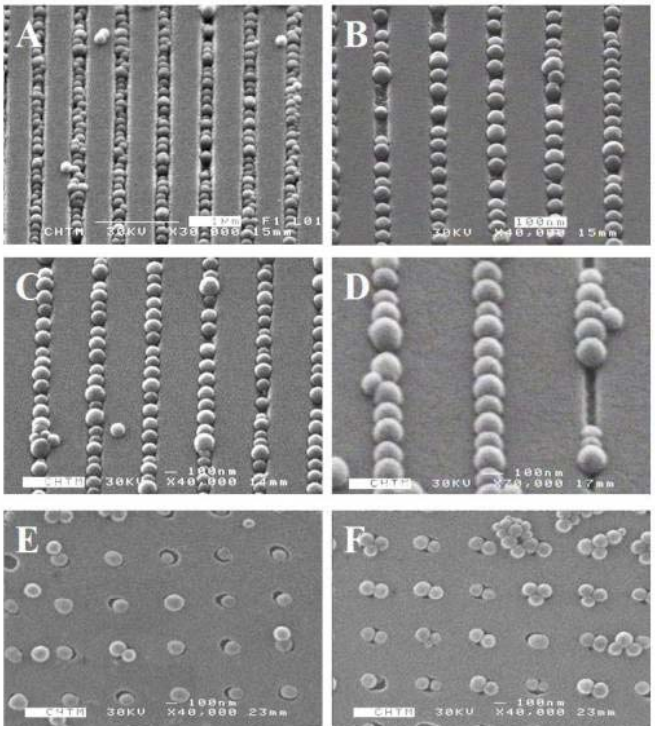


Fig. 10: Examples of directed self-assembly of ~ 50-nm diameter silica particles into patterns defined in an SiO_2 layer. The particles are deposited by spin-coating a colloidal suspension of particles onto an IL patterned wafer. Both 1D and 2D patterns are shown. The perfection of the deposition is impacted by the particle size distribution.

Imaging interferometric lithography (IIL)[32]-[34] is an approach to resolution enhancement that includes both extreme off-axis illumination (tilt of a coherent illumination source to the edge of the image pupil plane) with spatial filters to assure uniform coverage of frequency space. This ensures that the maximum spatial frequency available to the optical system is accessed (corresponding to the interference between coherent beams originating across the diameter of the pupil). In order to print these frequencies in both the *x*- and *y*-spatial directions of the image, both the offset direction and the pupil filter have to be rotated by 90°. Finally, to restore a telecentricity to the image so that deviations from precise focus can be tolerated, additional exposures with illumination from the opposite sides of the pupil are required. This can all be automated within the tool so that exposure times are not impacted by these multiple exposures. An important advantage of IIL for hyper-*NA* immersion exposures is that the polarization of the source can be adjusted at each exposure to ensure that the optimal (TE) polarization is used for both the *x*- and *y*-offsets. This is more difficult to accommodate with phase-shift mask approaches to accessing the limits of image frequency space.

Both OPC and PSM techniques significantly complicate the mask and increase mask costs that are becoming comparable to or even larger than tool costs in a cost-of-ownership calculation. IIL uses simpler binary masks and minimizes the need for OPC, this can be an important cost savings – particularly for small lots as in ASIC production or during initial product development cycles with frequent mask replacements.

The resolution limits for arbitrary patterns are not as small as for periodic patterns as a result of the need to capture the pattern information in the sidebands around the fundamental half-pitch frequency [35]. This reduction in resolution is pattern dependent, a factor of 13% reduction was assumed for the final column of Table I. Fig. 12 shows simulations for some of the nodes in Table I. For the 65- and 45-nm hp node, the fidelity of the image is quite good [36]. For the nodes involving spatial frequency multiplication, the variation of the contrast across the image causes more distortion of the pattern. This could be addressed with additional OPC features; some relatively minor OPC has already been applied for the modeling of the 22-nm hp node in Fig. 12 [37]. The important result is that there is a path for extending 193-nm based optical lithography all the way to the 16 nm node with relatively small changes to current optical lithography. Investigations are currently underway towards a number of less evolutionary alternatives: EUV; mask-

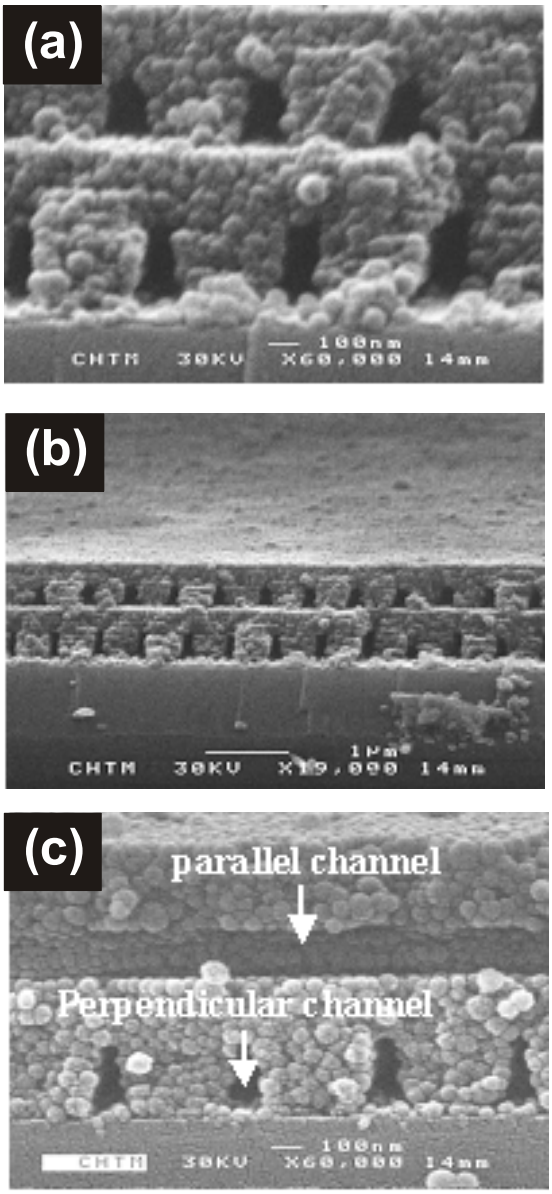


Fig. 11: Multilevel directed self-assembly of silica nanoparticles to form multilevel nanofluidic channels.

less; and nanoimprint lithographies. Ultimately, the manufacturing decision will be based on economics – the most cost effective solution – including all factors such as throughput and yield - will be adopted by the industry.

IV APPLICATIONS OF INTERFEROMETRIC LITHOGRAPHY

Periodic patterns are suitable – and even desirable - for many experimental nanoscience investigations. The obvious example is a photonic crystal for which periodicity is the primary attribute. In many cases an important requirement is on the overall sample area. For example, for the study of nanoscale epitaxial crystal growth, it is critical to have large area samples for detailed characterization, since many semiconductor material probes require large-area samples.

Interferometric lithography provides a unique capability for the inexpensive fabrication of large-area samples with nanoscale features for a wide variety of nanoscience investigations. Because IL is maskless, it is a simple matter to change periodicities, pattern symmetries, line:space ratios and other pattern details. These periodic patterns find extensive applicability in nanoscale epitaxial growth, in investigations of nanofluidics and of interaction of biomolecular species at the nanoscale, in nanomagnetics, and in nanophotonics among others. Space does not permit a full recitation of each of these areas. A brief overview of each of these areas is presented to illustrate the broad range of investigations enabled by interferometric lithography.

IV.1 Nanoscale Epitaxial Growth

Epitaxial growth of semiconductors continues to have an enormous impact on the development of information technology. The traditional paradigm of epitaxial crystal growth begins with a nearly perfect crystalline template (the substrate) upon which new materials are deposited in atomic registry with exquisite atomic level control in the growth direction, and with no control whatsoever in the transverse dimensions. When materials with a lattice mismatch of more than ~0.1% are combined, a large strain energy builds up and a prohibitively high defect density re-

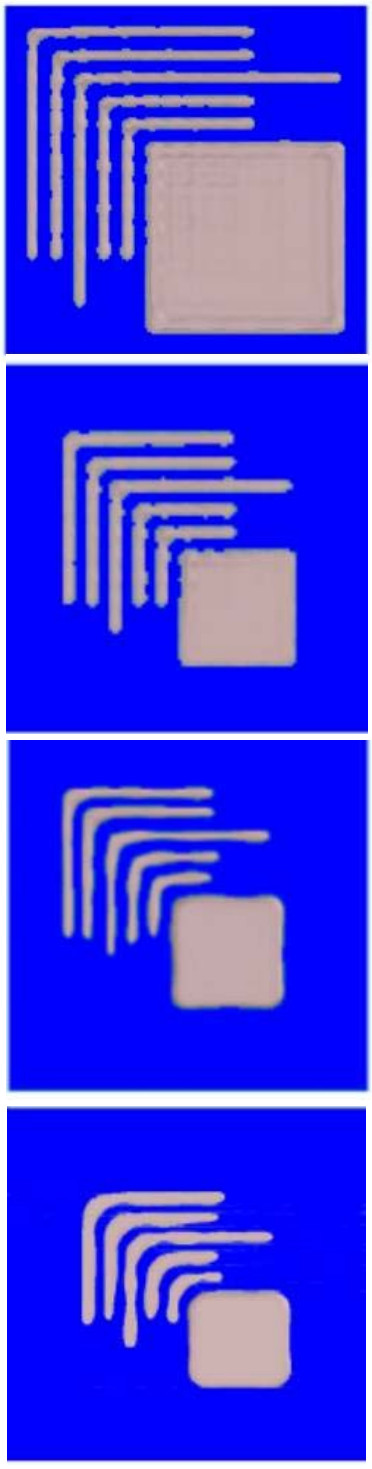


Fig. 12: Simulation of the upcoming ITRS nodes with 193-nm based optical lithography.

sults. The situation is entirely different if the links between the two materials are in the form of an array of "bridges" of nanoscale dimensions [38], [39]. In this case 3D strain mechanisms are enabled that are forbidden in traditional, large-area epitaxy and the lattice mismatch strain can be localized to the growth interface. Furthermore, during nanoscale growth there are significant modifications to fundamental kinetics and energetics of the crystal growth that lead to new growth modalities that are not available in large area growth, and to the nucleation and propagation of defects that allow defects to be localized at the growth interface. The dimensionless parameters that operate in nanoscale growth are the ratio of the linear pattern dimension to an adatom diffusion length and to the critical thickness for the generation of defects in heterostructure growth.

Patterning can involve masking of the growth surface with a second usually amorphous material such as SiO_2, or the exposing of different crystal faces on the substrate with selectivity coming from preferential growth on one class of facets. An example of growth through a SiO_2 mask is shown in Fig. 13 that shows a sequence of results for the growth of InGaAs on GaAs [40]. Panel (a) shows the starting substrate and 45-nm thick SiO_2 growth mask patterned at a pitch of 355 nm. Panel (b) is a SEM surface after growth of a 50-nm GaAs buffer layer and a 200-nm $In_{0.06}Ga_{0.94}As$ layer. Except for the incompletely filled hole, which was a rare occurrence across the 2 cm^2 sample, the surface is remarkably smooth and featureless showing good coalescence between the isolated growth areas. Panel (c) shows a low resolution cross section TEM. The SiO_2 mask, the GaAs buffer layer growth and the InGaAs growth are clearly seen. There are no evident long range defects such as threading dislocations that would be expected for blanket growth, even though the InGaAs layer thickness is three times the critical thickness for

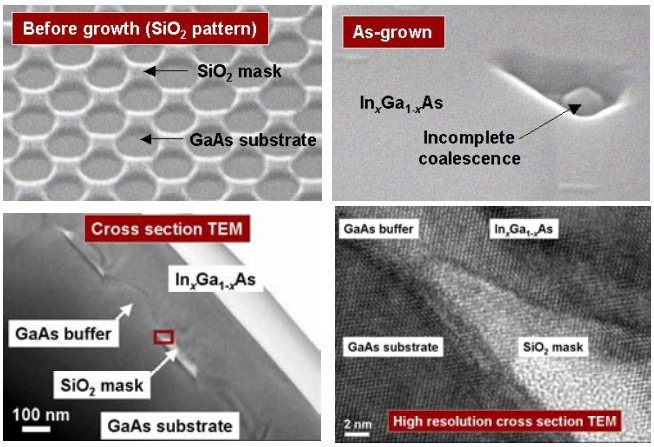


Fig. 13: Nanoscale epitaxial growth of InGaAs on GaAs over a thin SiO_2 mask. The top two panels are SEM pictures showing the intial SiO_2 mask (a) and the smooth featureless overgrowth c) Low resolution TEM showing the overgrown mask and the GaAs buffer and InGaAs overgrowth. d) High resolution TEM showing the coherent interface between the growths and the absence of any long range defects. This InGaAs layer is three times the critical thickness for blanket epitaxy.

this lattice mismatch. Finally, panel (d) is a high resolution TEM of the area within the small square in the center of (c). The various films are clearly identified and no extended defect is evident. Both x-ray diffraction and photoluminescence measurements indicate that the InGaAs film is almost fully relaxed. This confirms the initial premise that nanoscale 3D growth provides strain-relief mechanisms that are not available for traditional blanket-film growth.

An example of selective growth using crystal facets is shown in Fig. 14 which shows the growth of GaN on a nanoscale textured v-groove Si substrate that exhibits both (100) and {111} faces [41], [42]. Fig. 14a shows the starting substrate with a pitch of 355 nm. Within a period, two (100) faces and two equivalent {111} faces are exposed. Fig. 14b shows the result of ~ 75 nm of GaN growth. The GaN has nucleated in the hexagonal phase on the two {111} faces as show by the opposing white arrows in the figure. There are voids where no material grows at the centers of both (100) faces, clearly evincing the selective nucleation on the {111} surfaces. There is a large density of stacking faults (fluctuations between hexagonal and cubic phases) that provide strain relief of the large lattice mismatch between Si and GaN. After the growth proceeds some small amount, defect-free hexagonal GaN is obtained in two separated nanocrystals located on each Si {111} face. As the growth proceeds, these nanocrystals coalesce, cutting of the material flow to the bottom facet and leaving the void. Symmetry dictates that the GaN crystal cannot maintain the two opposed c-axes and as a result a cubic symmetry region is established, initially again with a significant density of localized stacking faults. Fig. 14c shows a mapping of these regions. Fig. 14d is a high resolution TEM in the central region clearly showing the stacking fault and cubic GaN regions. This spatial phase separation is a demonstration of the radically new and unanticipated growth modalities available with nanoscale growth.

IV.2 Nanofluidics

Microfluidics has become an important research area, particularly for biological separations and other investigations. Very complex systems are being investigated [43]. Nanofluidic systems offer an exciting new domain. This results because the Debye screening length in biologically relevant fluids is ~ 5- to 50-nm, thus in a nanofluidic system the screening from the walls impacts the entire fluid volume. In contrast to the situation in microfluidic systems where the bulk of the fluid volume is neutral, in nanofluidic channels the entire fluid is charged. This makes it possible to vary the fluid conditions locally along the channel and also strongly modifies electrokinetic transport since the electroosmotic component extends fully across the channel. In this condition

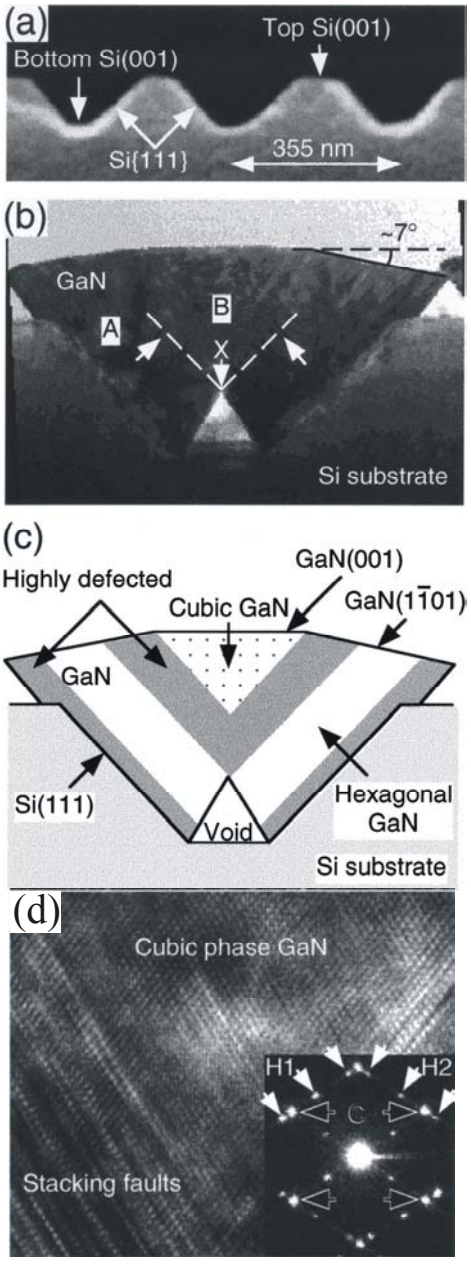


Fig. 14: GaN growth (MOCVD) on a nanoscale patterned {111}-(100) textured Si surface. (a) Shows the patterned substrate with two opposing {111} faces and two (100) faces in each 355-nm period. (b) Shows the result of a growth of ~75 nm of GaN. The GaN nucleates in the hexagonal phase on the two {111} faces with c-axes directed perpendicular to the Si face. (c) Map showing the material evolution. As the two GaN nanocrystals come together, symmetry dictates that a cubic region is formed. There are regions of stacking faults both at the Si:GaN interfaces and again at the hexagonal:cubic interfaces. (d) High-resolution TEM of the cubic region.

the fluid flow is modified from the usual plug flow conditions and new phenomena are observed [44], [45].

Fig. 15a shows an integrated nanofluidic flow system [46]. Integrated refers to the multiple hierarchical length scales from mm for the pipette interfaces to ~ 100 μm for the microchannels that run between the pipette ports and the ~ 100-nm wide nanochannels which are ~ 1 cm long. Thus the fabrication scales for this chip extend over a range of 10^5! The nanochannels were defined with interferometric lithography, etched into a Si wafer using RIE and then oxidized to isolate the fluid from the Si and to further narrow the nanochannel dimensions (SiO_2 has a ~ 40% larger volume than the Si it replaces). The remainder of the structure was fabricated with traditional optical lithography since there are no small features nor is there a critical alignment between the two phases of the fabrication. This is a good example of the mix-and-match fabrication discussed above.

IV.3 Nanoscale Magnetic Effects

There has been substantial interest in the magnetic properties of individual single-domain nanomagnetic structures as a result of their potential for increased storage density as compared with traditional multi-domain microstructures [47], [48]. Interferometric lithography is a natural technology for the definition of large areas of these patterned magnetic materials [49]-[52]. Interest in this area has been stimulated by the development of lithographically patterned multilayer elements including magnetic tunnel junction (MTJ) [53] and pseudo-spin-valve (PSV) [54], [55] elements that are key components in high-density magnetic random access memory (MRAM) architectures.

Superconducting films also show interesting phenomena in the presence of a periodic perturbation such as an array of holes.

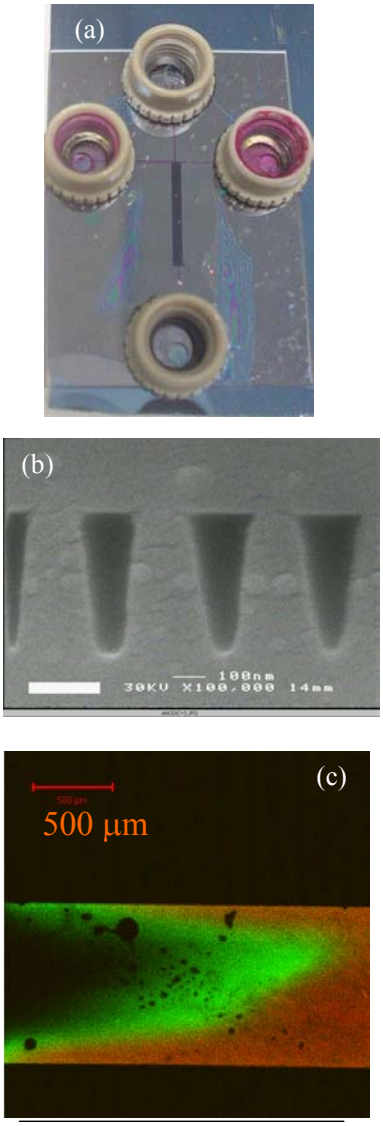


Fig. 15: (a) Optical micrograph of an integrated nanofluidic chip consisting of pipette ports, microchannels in a T configuration, and ~ 2000 parallel nanochannels (~ 1 cm long running from top to bottom in the figure). (b) SEM micrograph of a cross section of the nanochannel array. The channels are etched ~ 1 µm deep into Si, oxidized to provide SiO_2 isolation for the fluid, and capped with a pyrex cover slip using anodic bonding. (c) Optical micrograph showing electrokinetic separation of two dyes with different charge states.

The holes act as pinning sites for the quantized magnetic flux lines associated with a superconducting film. At low magnetic fields, the flux lines are all associated with these pinning sites and the magnetization shows distinctive structure when the flux density is commensurate with the density of the hole array or is an integral multiple of the hole density so that the flux lattice forms a superstructure locked to the hole array [56]. For a large-area sample, mismatches of the crystal structure between grains nucleated in different regions of the sample, quite analogous to grain boundaries in crystal growth, are observed [57].

IV.4 Nano-optics and Nanophotonics

Nanophotonics is an obvious application area for IL. Often the requirement is for large areas of periodic or quasi-periodic structures. Several immediately evident areas include 1D, 2D and 3D photonic crystals as well as more complex periodic structures for metamaterial applications such as enhanced transmission and field concentration in arrays of nanoscale metallic structures (holes and toroids) and negative index metamaterials.

IV.4.1 Distributed Feedback/Distributed Bragg Reflectors/Photonic Crystals

Distributed feedback (DFB) and distributed Bragg reflector (DBR) lasers require simple 1D gratings. The typical periodicity requirement for a first order grating is $\lambda/2n$ where λ is the laser wavelength and n is the modal index. For a GaAs laser, $\lambda \sim$ 860 nm and $n \sim 3.5$, so $p \sim 125$ nm. As demonstrated earlier in this paper, immersion IL at 193 nm can extend to as small as $p \sim 67$ nm, covering the full range of periods of interest. Fig. 16 shows a typical result for a mid-IR DFB laser [58]. Over 30 dB of side-mode suppression was achieved in this initial, unoptimized experiment.

Two-dimensional photonic crystals in slab geometries have many applications from waveguiding [59] and photonic-crystal lasers [60] to photonic crystal light extraction [61]. The

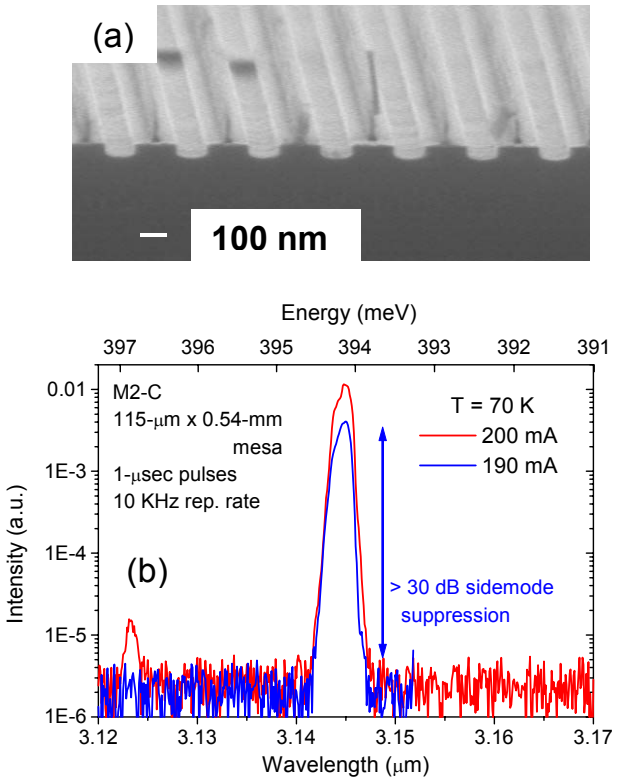


Fig. 16: (a) 465-nm period distributed feedback grating etched into a interband cascade InAs/GaSb laser structure. (b) High resolution output spectrum showing 30 dB side-mode suppression.

required pitch for each of these structures is similar to that for the DFB lasers discussed above, since the same wavelengths and materials are of interest. For photonic crystal LED light extraction, the wavelengths are shorter (~ 460 nm for blue LEDs) and the refractive index is somewhat lower (~ 2.5) so the required pitch is as small as 92 nm; again well within the capabilities of immersion IL.

There are several approaches to the fabrication of 2D structures using IL, perhaps the easiest is to use multiple 2-beam exposures with appropriate rotation of the wafer between exposures as discussed above in Sect. II.3. Fig. 17 shows an example of a 2D hexagonal pattern mask defined in a Ni etch mask deposited onto a GaN wafer with a hexagonal lattice parameter of 405 nm. Pattern transfer becomes a significant issue as the scale of the pattern is reduced. Typically, the thickness of the photoresist layer has to be reduced along with the pattern size to avoid photoresist collapse issues on developing that result from surface tension forces as the resist layer dries. Ultimately, this may force the adoption of a dry developing scheme or a new class of resist with more structural rigidity (higher Young's modulus). Various multi-level resist schemes incorporating a hard mask layer are also possible approaches to this issue.

There are at least two approaches to 3D structures. Structures can be built up in a layer-by-layer fashion very analogous to traditional semiconductor manufacturing. In particular, there has been quite a bit of attention devoted to a woodpile structure, with alternating layers of long bars oriented in the *x*- and *y*- directions with each sequential layer in the same direction offset by ½ the pitch [62]. Most often the fabrication has used standard semiconductor processing technologies [63]-[64]. Alternatively, using multiple-beam, noncoplanar interferometric lithography the entire 3D photonic crystal pattern can be produced in a single series of exposures in a thick photosensitive material [65]-[67]. All 14

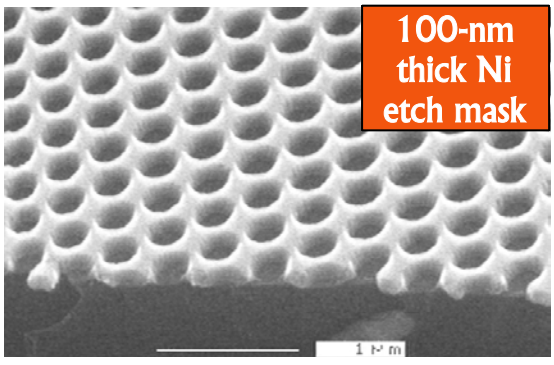


Fig. 17: 2D patterned Ni etch mask at a hexagonal lattice constant of 405 nm.

Bravais lattices and many other space groups are accessible with adjusting the number, intensity and polarizations of the various beams [68], [69]. This interferometric technique has been extended to the fabrication of compound (inter-penetrating) lattices such as the woodpile structure and the diamond lattice [66], [70].

IV.4.2 Metamaterials and Negative Refractive Index

Metals naturally provide a negative permittivity at frequencies below the plasma frequency ω_p as a consequence of the free electron response function given by

$$\varepsilon_{plasma} = 1 - \frac{\omega_p^2}{\omega(\omega + i\gamma)} \tag{4}$$

where γ is the electronic scattering frequency. There is no comparable naturally occurring negative magnetic susceptibility. However, LC resonant (tank) circuits provide a region of negative permeability around their resonance frequency. If an ensemble of such resonant structures with dimensions much less than the wavelength is aggregated into a material with inter-structure distances of much less than a wavelength, then an effective medium results with a negative permeability over a limited frequency range. The permeability in this case is dependent upon the properties of the individual resonances and is independent of the details, such as periodicity, of their arrangement into a material. Such a material is classified as a metamaterial. Its permittivity (ε) and permeability (μ) result from the fabricated resonant structures.

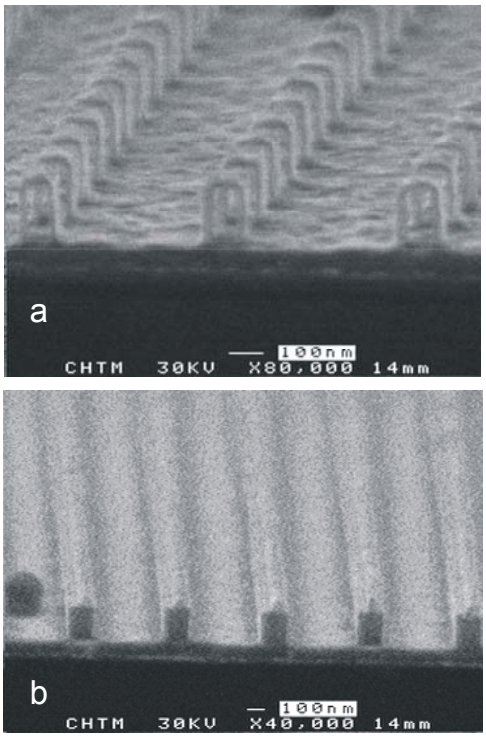


Fig. 18: SEM pictures of the arrays of 2D (a) and 1D (b) staple-shaped nanostructures. The 1D samples are different from the 2D sample in that the gold "staples" are continuous in the *y*-direction and the loops are filled with Si_3N_4.

Fig. 18 shows an example of magnetic metamaterial structures fabricated using IL [71]. The major feature is a loop formed by shadow evaporation over a 1D patterned Si_3N_4 sacrificial layer with a 600-nm pitch. In the top panel, this has been segmented into an array of "staples" and the nitride has been selectively removed. The bottom panel shows a 1D array with the nitride still in place. Fig. 19 shows the measured (top) and simulated (middle) responses for a number of structures with different structural dimensions, but with the same pitch. The response is clearly dependent on the individual structural parameters satisfying the definition of a metamaterial. Finally, the bottom panel shows the permeability derived from the model for the highest frequency resonance. A region of negative Re(μ) is clearly established. Other reports of negative permeability at infrared frequencies [72] have relied on e-beam lithography to define the smallest feature, and consequently have been restricted to small samples. In contrast, this

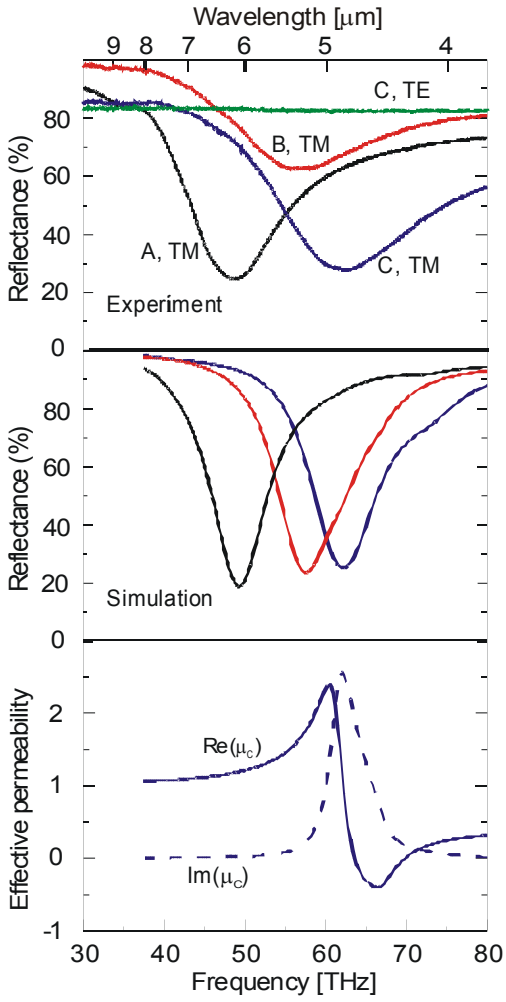


Fig 19: Experimental (top) and simulation (middle) results for the reflectivity as a function of frequency for three different structures. The resonance frequency is a function of the nanostructure parameters and is independent of array period, which is the same for all three structures. The resonance is observed only for TM polarization, i.e. incident magnetic field coupled into the inductive structure. Only one TE polarization result is shown. The bottom panel shows the effective permeability extracted from the model for the highest frequency resonance.

process relies on deposition (of a dielectric spacer and of the gold film) to define the smallest features and samples are several cm^2 in size, limited by convenience of handling rather than by cost and time of pattern definition. This illustrates one of the major advantages of a parallel optical lithography approach to nanophotonic structures.

There is great interest in the fabrication of negative index materials [73] – combining a negative permeability, as shown above, with a negative permittivity usually from electron plasma response of a metal film. The interest in this topic has increased dramatically as a result of the prediction of important new effects, including diffractionless imaging, in such a material [74]. A particularly simple realization of such a material, shown in Fig. 20, is a metal/dielectric/metal/substrate multilayer stack perforated by a 2D hole array [75]. For the incident polarization indicated in the cartoon, the magnetic response – very similar to that discussed above – results from the distributed inductance/capacitance associated with the two separated metal films. Fig. 21 shows the measurement and modeling of the refractive index, clearly showing a negative index over a limited wavelength range. Importantly, in this experiment, the refractive index was obtained uniquely from measurements of both the amplitude and the phase of both the film reflectance and transmission. For the simulation the only adjustable parameter was the electronic scattering frequency in the metal that was taken as approximately 3× the literature value for bulk Au based on fitting the transmission measurements. There is remarkably good agreement between the measurement and the modeling. There remains significant loss in this material with the $\mathrm{Im}(n) > |\mathrm{Re}(n)|$ across the resonance. Detailed investigation shows that for this structure while the permeability shows a resonance response, its real part does not become negative and the negative index is a consequence of the combination of negative real part of the permittivity and a positive imaginary part of the permeability at the resonance. More optimal

structures have been designed that promise a significant improvement in the quality of the negative index material (|Re(n)/Im(n)|).

Interferometric lithography is especially attractive for the fabrication of this class of negative index materials. As has been amply demonstrated earlier, IL is ideally suited to the fabrication of large arrays of 2D structures such as the pattern of holes. The fabrication in this case was simply to create a 2D post array, etch through the underlying ARC layer, and then to deposit the three films sequentially and dissolve away the photoresist to leave the final structure. Sample dimensions were ~ 2×2 cm^2. This simplicity of fabrication of large areas of materials offers promise both for the detailed study and optimization of these effects and also for manufacturing of large areas of nanoscale structures for practical applications.

V CONCLUSIONS

Nanoscale fabrication is a critical aspect of any nanotechnology. Optical lithography is the only large-scale manufacturing technique that the integrated circuit industry has ever used for pattern definition. Despite continuing predictions of its demise, optical lithography continues to be extended, most recently with immersion techniques that gain by the refractive index of the material between the lens and the photoresist – initially H_2O with an index of 1.44 at 193 nm. It now appears likely that a hyper-NA (NA > 1) optical lithography will address the industry needs for the 45-nm half-pitch node. Further extensions are possible and are being actively investigated for even finer lithographic patterns.

While this is good news for the semiconductor industry, the cost of these optical lithography tools, ~ $20 to $30 million, makes them inaccessible to a typical research environment. Much of this cost is associated with the need of the industry to produce fully arbitrary patterns at very high throughputs. Current steppers also require quite expensive masks (~ $ 1- to 2-million for a microprocessor mask set) to define the patterns and are capable of a throughput of ~ 85 wafers per hour, each with ~ 125 dies (22×33 mm^2) across a 300-mm diameter wafer; this corresponds to an average sustained data throughput rate of ~ 10^{12} bit/s.

Fortunately, nanoscience research often requires less flexibility – for example periodic patterns, and always requires less throughput. The needs of much of nanoscience research can be met at a substantially lower cost. Interferometric lithography, the interference of a small number of coherent laser beams, readily produces large-area patterns (several cm^2) at the same and even smaller scales than are available from commercial steppers. With an ArF laser source at 193 nm and H_2O immersion,

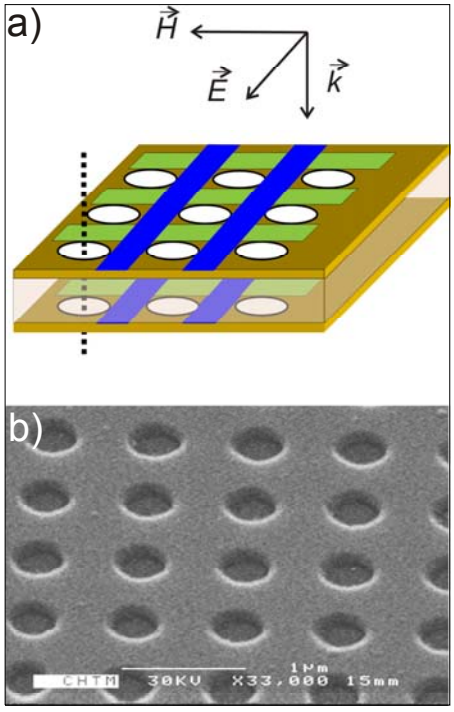


Fig. 20: a) Negative index material at 2 μm consisting of a Au/Al2O3/Au film stack on a glass substrate perforated with a 2D array of holes at a pitch of 0.8 μm. The green shaded regions contribute to the permeability while the blue shaded regions are responsible for the negative permittivity. b) SEM micrograph of the structure.

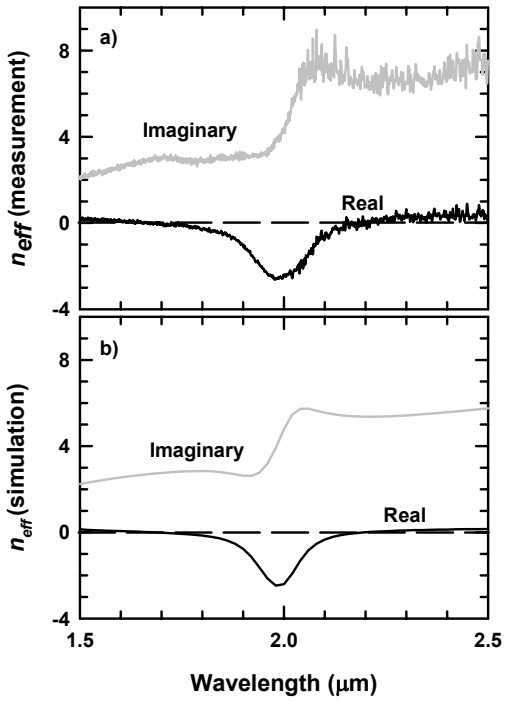


Fig. 21: Measured (a) and simulated (b) real and imaginary parts of the refractive index for the structure of Fig. 15. There is a region of negative index around 2 μm. The maximal real part of the negative index is around –2 while the imaginary part is 3 to 4, indicating a significant loss.

the smallest half-pitch (equal line:space) pattern accessible extends to 34 nm. Spatial frequency multiplication, taking advantage of nonlinearities in the processing, allows extensions by a factor of two or more, extending this capability ultimately to below 10 nm. Importantly, IL is a maskless approach to lithography that greatly extends flexibility without requiring expensive masks.

Several different techniques were discussed that extend the flexibility of IL. Multiple IL exposures, multiple-beam IL, and mix-and-match with both high resolution (e-beam) and with lower resolution (proximity optical lithography) all offer complimentary capabilities. Taken together, these techniques provide a facile suite of techniques applicable to a broad range of nanoscience applications, with a resolution capability that rivals standard nanofabrication techniques such as electron-beam lithography, but with a large-area capability that often allows additional diagnostic techniques that are not available for very small samples – and of course can sustain volume manufacturing.

Directed self-assembly, the combination of self-assembly to define features extending to molecular dimensions while retaining the organization across multiple length scales from nm to cm offered by optical lithography, is a promising future direction. Ultimately, this provides a pathway to a sub-nm patterning regime.

IL provides the limiting resolution of optical lithography, but the periodicity poses a problem for some applications, especially for the integrated circuit industry. Imaging interferometric lithography (IIL) was introduced as an extension of IL to arbitrary patterns. Essentially, IIL is a combination of extreme off-axis illumination, to reach the highest spatial frequencies permitted by a given optical system, along with imaging pupil plane filters, to avoid multiple counting of low frequency information and thereby to balance the image frequency response. With the addition of the concept of spatial frequency multiplication, IIL offers a route for optical lithography to the 16-nm half-pitch limit of the current industry manufacturing roadmap extending to 2019.

Numerous application areas for nanoscale patterns were discussed. Nanoscale heteroepitaxy provides a new paradigm for heterostructure crystal growth. The use of nanoscale seeds for nucleation of many small crystals provides new capabilities for the relief of lattice-mismatch strain. The important dimensionless parameters are (1) the ratio of the pattern linear dimension to the adatom mean-free path so that an adsorbed atom can sample the entire growing crystal for a preferred, low-energy incorporation site, and (2) the ratio of the pattern linear dimension to the criti-

cal thickness before the generation of defects in large-area heterostructure growth. These techniques are applicable to both molecular-beam epitaxy (MBE) and metal-organic vapor-phase epitaxy (MOVPE), the dominant growth techniques for advanced heterostructure devices. In the case of MBE, selective area growth, faceting, and lateral overgrowth and coalescence – all of which are rarely, if ever, observed for large area ($\gtrsim$ 1-μm scale) patterning – were clearly observed, illustrating the fundamental difference afforded by nanoscale patterning. Two examples were given: (1) growth of a relaxed InGaAs layer on a GaAs substrate using a 2D SiO_2 masking layer; and (2) spatial phase separation of the cubic and hexagonal phases of GaN grown on a (100)/{111} patterned Si wafer.

Nanofluidic transport and separation, particularly of biomolecules, is another exciting application for nanoscale patterning. Here the important dimensionless parameter is the ratio of the linear pattern dimension to the Debye screening length in the ionic fluids important for biology. A double layer of opposite charges forms at the wall of a nanochannel filled with an ionic fluid. Electrokinetic (electroosmosis and electrophoresis) transport is impacted by the extent of this double layer. This Debye length is of the order of 10- to 100-nm for most situations relevant to biology, which sets the scale differentiating nanochannels from the much more widely studied microchannels where the screening length is a small fraction of the Debye length. An integrated chip was demonstrated that combines features ranging from 50 nm to 1 cm – a five order of magnitude range! Initial flow measurements were presented demonstrating the separation of two dyes with different charge states.

Magnetic storage has long been migrating towards smaller features and higher densities and is now entering an era of single-domain, nanoscale, multilayer structures. Again, interferometric lithography is a powerful fabrication technique for the production of large areas and for ultimate manufacturing of devices and circuits such as magnetic random-access memories (MRAM).

Nanooptics and nanophotonics are natural applications for IL. Distributed feedback and distributed Bragg reflector lasers, requiring nanoscale one-dimensional patterning have long been a mainstay for the spectroscopic control of diode lasers. Recently, there has been extended interest in the use of 2D patterning for photonic crystals to control optical properties. This extends from photonic crystal waveguides with missing or modified rows of holes (posts) to guide optical beams to the use of photonic crystals for light extraction from high-index, semiconductor-based LEDs. Three-dimensional photonic crystals can be fabricated both with layer-by-layer techniques and by using multiple IL beams with varying azimuthal angles. Metamaterials and plasmonics are another area of great current interest. Using simple IL procedures, large areas of resonant magnetically coupled resonators (tank circuits) with resonances throughout the mid-infrared spectral region were demonstrated. Very recently, a large-area negative index material at 2 μm was demonstrated using a relatively simple IL implementation.

Hopefully by now the message is clear. Interferometric lithography provides a readily available, inexpensive, large-area route to nanoscale patterning of great importance to many areas in nanotechnology. Combining IL with self-assembly techniques in “directed self-assembly” extends the reach of patterning capabilities truly to molecular scales. There is a vast range of possibilities across many different fields and much to be learned from the application and extension of the techniques presented here.

[1] S. R. J. Brueck, "There are no fundamental limits to optical lithography," International Trends in Applied Optics, International Commission on Optics, pp. 85-110, SPIE Press (2002).
[2] M. J. Beesley and J. G. Castledine, "The use of photoresist as a holographic recording medium," *Appl. Opt.* **9**, 2720 (1970).
[3] R. A. Bartolini, "Characteristics of relief phase holograms recorded in photoresists." *Appl. Opt.* **13**, 129-139 (1974).
[4] L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll, "Generation of periodic surface corrugations," *Appl. Opt.* **17**, 1165 (1978).
[5] S. H. Zaidi and S. R. J. Brueck, "High aspect-ratio holographic photoresist gratings," *Appl. Opt.* **27**, 2999 – 3002 (1988).
[6] M. L. Schattenberg, E. H. Anderson, and H. I. Smith, "X-ray/VUV transmission gratings for astrophysical and laboratory applications," Phys. Scr. **41**, 13-20 (1990).
[7] A. Yen, E. H. Anderson, R. A. Ghanbari, M. L. Schattenburg and H. I. Smith, "Achromatic holographic configuration for 100-nm-period lithography," *Appl. Opt.* **31**, 4540-4545 (1992).
[8] B. W. Smith, A. Bourov, H. Kang, F. Cropanese, Y. Fan, N. V. Lafferty and L. V. Zavyalova, "Water immersion lithography at 193 nm," *Jour. Microlith., Microfab. and Microsys.* **3**, 44-51 (2004).
[9] C. V. Shank, and R. V. Schmidt, "Optical technique for producing 0.1-μm periodic surface structures," *Appl. Phys. Lett.* **23**, 154-156 (1973).
[10] W. T. Tsang and S. Wang, "Simultaneous exposure and development technique for making gratings in positive photoresist," *Appl. Phys. Lett.* **23**, 196-198 (1974).
[11] J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, "Liquid immersion deep-ultraviolet interferometric lithography," *Jour. Vac. Sci. Technol.* **B17**, 3306-3309 (1999).
[12] M. Switkes and M. Rothschild, "Immersion lithography at 157 nm," *Jour. Vac. Sci. Technol.* **B19**, 2353-2356 (2001).
[13] J. H. Burnett and S. G. Kaplan, "Measurement of the refractive index and thermo-optic coefficient of water near 193 nm," *Jour. Microlith., Microfab. and Microsys*. **3**, 68-72 (2004).
[14] A. K. Raub and S. R. J. Brueck, "Deep UV immersion interferometric lithography," *Proc. SPIE* **5040**, 667-678 (2003).
[15] Alex K. Raub, A. Frauenglass, S. R. J. Brueck, Will Conley, Ralph Dammel, Andy Romano, Mitsuru Sato, William Hinsberg, "Imaging capabilities of resist in deep-UV liquid immersion interferometric lithography," *Jour. Vac. Sci. Technol.* **22**, 3459-3464 (2004).
[16] A. Biswas and S. R. J. Brueck, "Simulation of the 45-nm half-pitch node with 193-nm immersion lithography," *Jour. Microlith., Microfab., and Microsys.* **3**, 35-43 (2004).
[17] S. H. Zaidi and S. R. J. Brueck, "Multiple-Exposure Interferometric Lithography," *Jour. Vac. Sci. Technol.* **B11**, 658-666 (1993).
[18] J. W. Goodman, Introduction to Fourier Optics, 2nd Ed., (McGraw Hill, New York, 1996).
[19] X. Chen, S. H. Zaidi, S. R. J. Brueck and D. J. Devine, "Interferometric Lithography of Sub-Micrometer Sparse Hole Arrays for Field-Emission Display Applications," *Jour. Vac. Sci. Technol.* **B14**, 3339-3349 (1996).
[]20 S. H. Zaidi, A. Frauenglass and S. R. J. Brueck, "Moiré Interferometric Alignment and Overlay Techniques," *Proc. SPIE* **2196**, 371-382 (1994).
[21] S. H. Zaidi, S. R. J. Brueck, T. Hill and R. N. Shagam, "Mix-and-match interferometric and optical lithographies for nanoscale structures," *Proc. SPIE* **3331**, 406-413 (1998).
[22] S. H. Zaidi and S. R. J. Brueck, "Nonlinear Processes to Extend Interferometric Lithography," *Proc. SPIE* **3676**, 371-378 (1999).
[23] J. A. Hoffnagle, W. D. Hinsberg, F. A. Houle and M. I. Sanchez, "Use of Interferometric Lithography to Characterize the Spatial Resolution of a Photoresist Film," *Jour. Photopolymer Sci. and Tech.* **16**, 373-379 (2003).
[24] S. S. H. Naqvi, S. H. Zaidi, S. R. J. Brueck and J. R. McNeil, "Diffractive Techniques for Lithographic Process Monitoring and Control," *J. Vac. Sci. Technol.* **B12**, 3600-3606 (1994).
[25] www.public.itrs.net
[26] A. Hatzor, and P. S. Weiss, "Molecular Rulers for Scaling Down Nanostructures," *Science* **291**, 1019-1020 (2001).

[27] D. Xia, D. Li, A. Biswas and S. R. J. Brueck, "Directed Self-Assembly of Silica Nanoparticles on Nanoscale Patterned Surfaces by Spin-Coating," *Adv. Matls* **16**, 1427 - 1432 (2004).
[28] D. Xia and S. R. J. Brueck, "Lithographically directed deposition of silica nanoparticles using spin coating," *Jour. Vac. Sci. Technol.* **22**, 3415-3420 (2004).
[29] A. K-K. Wong, Resolution Enhancement Techniques in Optical Lithography, (SPIE Press, Bellingham, WA, USA, 2001).
[30] H. J. Levinson, Principles of Lithography, 2nd Ed, (SPIE Press, Bellingham, WA, USA, 2005).
[31] For a compilation of much of the original literature, see Selected Papers on Resolution Enhancement Techniques in Optical Lithography, F. M. Schellenberg, ed. SPIE Milestone Series **178** (Spie Press, Bellingham, WA, USA, 2004).
[32] X. Chen and S. R. J. Brueck, "Imaging interferometric lithography - approaching the resolution limits of optics," *Opt. Lett.* **24**, 124-126 (1999).
[33] X. Chen and S. R. J. Brueck, "Experimental comparison of off-axis illumination and imaging interferometric lithography," *Jour. Vac. Sci. Technol.* **B17**, 921-929 (1999).
[34] A. Frauenglass, S. Smolev, A. Biswas, S.R.J. Brueck, "244-nm imaging interferometric lithography," *Jour. Vac. Sci. Technol.* **22**, 3465-3469 (2004).
[35] S. R. J. Brueck and X. Chen, "Spatial frequency analysis of optical lithography resolution enhancement techniques," *Jour. Vac. Sci. Technol***. B17**, 908-921 (1999).
[36] A. Biswas and S. R. J. Brueck, "Simulation of the 45-nm half-pitch node with 193-nm immersion lithography," *Jour. Microlith., Microfab. and Microsys.* **3**, 35-43 (2004).
[37] S. R. J. Brueck and A. M. Biswas, "Extension of 193-nm immersion optical lithography to the 22-nm half-pitch node," *Proc. SPIE* ***5*377**, 1315-1322 (2004).
[38] S. Luri and E. Suhir, "New approach to the high quality epitaxial growth of lattice-mismatched materials," *Appl. Phys. Lett.* **49**, 140-142 (1986).
[39] D. Zubia, S. H. Zaidi, S. R. J. Brueck and S. D. Hersee, "Nanoheteroepitaxial growth of GaN on Si by organometallic vapor phase epitaxy," *Appl. Phys. Lett.* **76**, 858-860 (2000).
[40] S. C. Lee, L. R. Dawson, B. Pattada, S. R. J. Brueck, Y-B. Jiang and H. F. Xu, "A strain-relieved, dislocation-free $In_xGa_{1-x}As$ (100) heterostructure by nanoscale patterned growth," *Appl. Phys. Lett.* **85**, 4181-4183 (2004).
[41] S. C. Lee, X. Y. Sun, S. D. Hersee, F. Xu and S. R. J. Brueck, "Spatial phase separation of GaN selectively grown on a nanoscale faceted Si surface, *Appl. Phys. Lett.* **84**, 2079-2082 (2004).
[42] S. C. Lee, B. Pattada, X. Y. Sun, S. D. Hersee, Y.-B. Jiang, H. Xu, and S. R. J. Brueck, "Nanoscale spatial phase modulation of GaN on a 355-nm period array of V-grooves fabricated in a Si(001) substrate," *IEEE Jour. of Quant. Electron.* **41**, 596-605 (2005).
[43] D. J. Harrison, K. Fluri, K. Seiler, Z. Fan, C. S. Effenhauser and A. Manz, "Micromachining a miniaturized capillary electrophoresis-Based Chemical Analysis System on a Chip," *Science* **261**, 895-897 (1993).
[44] H. Daiguji, P. Yang and A. Majumdar, "Ion transport in nanofluidic channels," *Nano Lett.* **4**, 137-142 (2004).
[45] D. Stein, M. Kruithof, and C. Dekker, "Surface-Charge-Governed Ion Transport in Nanofluidic Channels," *Phys. Rev. Lett*. **93**, 035901 (2004).
[46] M. J. O'Brien, P. Bisong, L. K. Ista, E. M. Rabinovitch, A. L. Garcia, S. S. Sibbett, G. P. Lopez and S. R. J. Brueck, "Fabrication of an integrated nanofluidic chip using interferometric lithography," *Jour. Vac. Sci. Technol.* **21**, 2941-2945 (2003).
[47] S. Y. Chou, M. S. Wei, P. R. Krauss and P. B. Fischer, "Single-domain magnetic pillar array of 35 nm diameter and 65 Gbits/in^2 density for ultrahigh density quantum magnetic storage," *J. Appl. Phys.* **76**, 6673-6675 (1994).
[48] S. Y. Chou, Patterned magnetic nanostructures and quantized magnetic disks," *Proc. IEEE* **85**, 652-671 (1997).
[49] G. Meier, M. Kleiber, D. Grundler, D. Heitmann and R. Wiesendanger, "Vertical polarization of quantum magnets in high density arrays of nickel dots with small height-to-diameter ratio," *Appl. Phys. Lett.* **72**, 2168-2170 (1998).
[50] M. A. M. Haast, J. R. Schuurbuis, L. Abelmann, J. C. Lodder and Th. L. Popma, "Reversal mechanism of submicron patterned CoNi/Pt multilayers," *IEEE Trans. Magn.* **34**, 1006-1008 (1998).
[51] M. Theilen, S. Kirsch, H. Weinforth, A. Carl and E. F. Wassermann, "Magnetization reversal in nanostructured Co/Pt multilayer dots and films," *IEEE Trans. Magn.* **34**, 1009-1011 (1998).

[52] C. A. Ross, S. Haratani, F. J. Castaño, Y. Hao, M. Hwang, M. Shima, J. Y. Cheng, B. Vögeli, M. Farhoud, M. Walsh and H. I. Smith, "Magnetic behavior of lithographically patterned particle arrays," *Jour. Appl. Phys.* **91**, 6848-6853 (2002).

[53] S. S. P. Parkin, K. P. Roche, M. G. Samant, P. M. Rice, R. B. Beyers, R. E. Scheuerlin, E. J. O'Sullivn, S. L. Brown, J. Bucchigano, D. W. Abraham, Y. Lu, M. Rooks, P. L. Trouilloud, R. A. Wanner, and W. J. Gallagher, "Exchange-biased magnetic tunnel junctions and application to nonvolatile magnetic random access memory" *Jour. Appl. Phys*. **85**, 5828-5833 (1999).

[54] B. A. Everitt, A. V. Pohm, R. S. Beech, A. Fink, and J. M. Daughton, "Pseudo spin valve MRAM cells with sub-micrometer critical dimension," *IEEE Trans. Magn. 34*, 1060-1062 (1999),

[55] X. Zhu, P. Grütter, Y. Hao, F. J. Castaño, S. Haranti, C. A. Ross, B. Vögeli, and H. I. Smith, "Magnetization switching in 70-nm-wide pseudo-spin-valve nanoelements," *Jour. Appl. Phys.* **93**, 1132-1136 (2003).

[56] V. Metlushko, U. Welp, G. W. Crabtree, R. Osgood, S. D. Bader, L. E. DeLong, Z. Zhang, S. R. J. Brueck, B. Ilic, K. Chung, and P. J. Hesketh, "Interstitial flux phases in a superconducting niobium film with a square lattice of artificial pinning centers," *Phys. Rev.* **B60**, 12585 (1999).

[57] S. B. Field, S. S. James, J. Barentine, V. Metlushko, G. Crabtree, H. Shtrikman, B. Ilic and S. R. J. Brueck, "Vortex configurations, matching, and domain structure in large arrays of artificial pinning centers," *Phys. Rev. Lett.* **88**, 067003 (2002).

[58] J. L. Bradshaw, J. D. Bruno, J. T. Pham, D. E. Wortman, S. Zhang and S. R. J. Brueck, "Single-longitudinal-mode emission from interband cascade DFB laser with a grating labricated by interferometric lithography," *Proc. IEE-Optoelectronics* **150**, 288-292 (2003).

[59] A. Yariv, Y. Xu, and S. Mookherjea, "Transverse Bragg resonance laser amplifier," *Opt. Lett.* **28**, 176-178 (2003).

[60] O. Painter, R. K. Lee, A. Scherer, A, Yariv, J. D. O'Brien, P. D. Dapkus, and I. Kim, "Two-dimensional photonic band-gap defect mode laser ," *Science* **284**, 819-1821 (1999),

[61] J. J. Wierer, M. R. Krames, J. E. Epler, N. F. Gardner, J. R. Wendt, M. M, Sigalas, S. R. J. Brueck, D. Li, and M. Shagam, "III-Nitride LEDs with photonic crystal structures," *Proc. SPIE* **5739**, 102-107 (2005).

[62] K. M. Ho, C. T. Chan, C. M. Soukoulis, R. Biswas, and M. Sigalas, "Photonic band-gaps in 3-dimensions – new layer-by-layer periodic structures," *Solid State Commun.* **89**, 413-416 (1994).

[63] N. Yamamoto, S. Noda, and A. Sasaki, "New realization method for three-dimensional photonic crystal in the optical wavelength region: Experimental Consideration," *Jpn. Jour. Appl. Phys* **36**, 1907-1911 (1997).

[64] S. Y. Lin, J. G. Fleming, D. L. Heatherington, B. K. Smith, R. Biswas, M. M. Sigalas, W. Zubrzycki, S. R. Kurtz, and J. Bur, "A three dimensional photonic crystal operating at infrared wavelengths," *Nature (London)* **394**, 251-253 (1998).

[65] M. Campbell, D. N. Sharp, M. T. Harrison, R. G. Denning, and A. J. Turberfield, "Fabrication of photonic crystals for the visible spectrum by holographic lithography," *Nature (London)* **404**, 53-56 (2000).

[66] S. Shoji, H. B. Sun and S. Kawata, "Photofabrication of wood-pile three-dimensional photonic crystals using four-beam laser interference," *Appl. Phys. Lett.* **83**, 608-610 (2003).

[67] X. Wang, J. F. Xu, H. M. Shu, Z. H. Zeng, Y. L. Chen, H. Z. Wang, Y. K. Pang and W. Y. Tam, "Three-dimensional photonic crystals fabricated by visible light holographic lithography," *Appl. Phys. Lett.* **82**, 2212-2214 (2003).

[68] L. Z. Cai, X. L. Yang and Y. R. Wang, "Formation of three-dimensional periodic microstructures by interference of four noncoplanar beams," *Jour. Opt. Soc. Amer.* **A19**, 2238-2244 (2002).

[69] C. K. Ullal, M. Maldovan, M. Wohlgomuth and E. L. Thomas, "Triply periodic bicontinuous structures through interference lithography: a level set approach," *Jour. Opt. Soc. Amer.* **A20**, 948-954 (2003).

[70] Y-C. Zhong, S-A. Zhu and H-Z. Wang, "Fabrication of compound lattice by holographic lithography," *Chin. Phys. Lett*. **22**, 369-373 (2005).

[71] S. Zhang, W. Fan, A. Frauenglass, B. Minhas, K. J. Malloy and S. R. J. Brueck, "Demonstration of Mid-Infrared Resonant Magnetic Nanostructures Exhibiting a Negative Permeability," *Phys. Rev. Lett.* **94**, 037402 (2005).

[72] T. J. Yen, W. J. Padilla, N. Fang, D. C. Vier, D. R. Smith, J. B. Pendry, D. N. Basov, and X. Zhang, "Terahertz Magnetic Response from Artificial Materials," *Science* **303**, 1494-1496 (2004).

[73] V. G. Veselago, "The electrodynamics of substances with simultaneously negative values of ε and μ," Sov. Phys. Usp. **10**, 509-514 (1968).
[74] J. B. Pendry, "Negative refraction makes a perfect lens," *Phys. Rev. Lett.* **85**, 3966-3969 (2000).
[75] S. Zhang, W. Fan, N. C. Panoiu, K. J. Malloy, R. M. Osgood and S. R. J. Brueck, "Demonstration of an infra-red negative-index material," Postdeadline paper presented at Nanophotonics for Information Systems OSA Topical Meeting, April 15, 2005.

REPRINTS

Reprints of publications supported by the ARO/MURI program are included in this appendix.

1. Lithographic Imaging

Ch. J. Schwarz, A. V. V. Nampoothiri, J. C. Jaspara, W. Rudolph, and S. R. J. Brueck, *Demonstration of Two-Photon Lithography,* Jour. Vac. Sci. Technol. **B19**, 2362-2364 (2001).

S. R. J. Brueck, *There are No Fundamental Limits to Optical Lithography,* International Trends in Applied Optics, International Commission on Optics, pp. 85-110, SPIE Press (2002).

D. V. Korobkin and E. Yablonovitch, *Two-Fold Spatial Resolution Enhancement by Two-Photon Exposure of Photographic Film*, Opt. Engr. 41, 1729-1732 (2002)

A. K. Raub and S. R. J. Brueck, *Deep UV Immersion Interferometric Lithography*, Proc. SPIE **5040**, Optical Microlithography XVI, 667-678 (2003).

Eric S. Wu, Balu Santhanam and S. R. J. Brueck, *Grating Analysis of Frequency Parsing Strategies for Imaging Inteferometric Lithography*, Proc. SPIE **5040**, Optical Microlithography XVI, 1267-1283 (2003).

A. Biswas and S. R. J. Brueck, *Simulation of the 45-nm Half-Pitch Node with 193-nm Immersion Lithography*, Jour. Microlithography, Microfabrication, and Microsystems **3**, 35-43 (2004)

Thanis M. Tridhavee, Balu Santhanam and S. R. J. Brueck, *Optimization and apodization of aerial images at high NA for imaging interferometric lithography,* Proc. SPIE **5377**, 1544–1554 (2004).

A. K. Raub, A. Frauenglass, S. R. J. Brueck, W. Conley, R. Dammel, A. Romano, M. Sato and W. Hinsberg, *Deep-UV Immersion Interferometric Lithography,* Proc. SPIE **5377**, 306-318 (2004).

Abani. M. Biswas, A. Frauenglass, and Steven R. J. Brueck, *Simulation of the 45-nm Half-Pitch Node with 193-nm Immersion Lithography,* Proc. SPIE **5377**, 1579-1586 (2004).

S. R. J. Brueck and Abani. M. Biswas, *Extension of 193-nm Immersion Optical Lithography to the 22-nm Half-Pitch Node,* Proc. SPIE **5377**, 1315-1322 (2004).

Alex K. Raub, A. Frauenglass, S. R. J. Brueck, Will Conley, Ralph Dammel, Andy Romano, Mitsuru Sato, William Hinsberg, *Imaging Capabilities of Resist in Deep-UV Liquid Immersion Interferometric Lithography*, Jour. Vac. Sci. Technol. **B22**, 3459-3464 (2004).

A. Frauenglass, S. Smolev, A. Biswas, S.R.J. Brueck, *244-nm Imaging Interferometric Lithography,* Jour. Vac. Sci. Technol. **B22**, 3465-3469 (2004).

2. Microscopy Imaging

Christian J. Schwarz, Yuliya Kuznetsova and S. R. J. Brueck, *Demonstration of Imaging Interferometric Microscopy*, Proc. SPIE **5038**, Metrology, Inspection and Process Control XVII 1-12 (2003).

C. J. Schwarz, Y. Kuznetsova and S. R. J. Brueck, *Imaging Interferometric Microscopy,* Optics Lett. **28**, 1424-1426 (2003).

3. Lithography Stage Development

Lu Xiaoming and Hy D. Tran, *Self-Calibration For 2-Dimensional Precision Stages Based On Circle Closure Principle*, in *Proc. ASPE*, Scottsdale, AZ, October 2000.

P.T. Ton-Nu, J. Ouyang, and H. D. Tran, *Multipoint Temperature Control Using Thermoelectric Modules*, in Proc. ASPE, Scottsdale, AZ, October 2000.

Ajit Borundia and Hy D. Tran. *An experimental investigation of heat pipes at low power inputs,* Proc. ASPE 2002.

Xiaoming Lu and Hy. D Tran, *2-D Self-Calibration for Scale Based Metrology in Nanolithography*, Proc. ASPE 2002.

L. Xiaoming and H. D. Tran, *A Fuzzy Logic Based Adaptive Feedforward PI Controller for Nanometer Positioning,* Proc. ASPE Annual Meeting 2003.

A. Abdo, G. Nellis, A. Wei, M. El-Morsi, R. Engelstad ,S. R. J. Brueck and A. Neumann, *Optimizing the fluid dispensing process for immersion lithography,* Jour. Vac. Sci. Technol. **B22**, 3454-3458 (2004).

4. Lithography Metrology

P. C. Logofatu and J. R. McNeil, *Scatterometry: a metrology for subwavelength surface-relief gratings*, in Metrology, inspection, and process control for microlithography XV, N. T. Sullivan, ed., Proc. SPIE **4344**, 472-483 (2001).

P. C. Logofatu and J. R. McNeil, *Measurement precision of optical scatterometry*, in Metrology, inspection, and process control for microlithography XV, N. T. Sullivan, ed., Proc. SPIE **4344**, 447-453(2001).

P. C. Logofatu, *Analysis of phase measurement error for null generalized ellipsometry using the phase compensator*, Journal of optoelectronics and advanced materials **3**, 45-49 (2001).

Petre C. Logofătu, *Simple Method for Determining the Fast Axis of a Wave Plate,* Optical Engineering 41, 3316-3318 (2002).

Petre. C. Logofătu, *Phase Modulation Scatterometry,* Appl. Opt. 41, 7187-7192 (2002).

P. C. Logofătu, *Sensitivity Analysis of Grating Parameter Estimation*, Appl. Opt. 41, 7179-7186 (2002).

Petre Catalin Logofatu, *UV scatterometry*, Proc SPIE **5038**, 208-214 (2003)

5. Lithography Applications – Nanoscale Epitaxial Growth

D. Zubia, S. H. Zaidi, S. R. J. Brueck and S. D. Hersee, *Nanoheteroepitaxial Growth of GaN on Si by Organometallic Vapor Phase Expitaxy*, Appl. Phys. Lett. **76**, 858-860 (2000).

D. Zubia, S. H. Zaidi, S. D. Hersee and S. R. J. Brueck, *Nanoheteroepitaxy – A Nanofabrication Route to Improved Epitaxial Growth*, Jour. Vac. Sci. Technol. B18 3514-3520 (2000).

D. Zubia, S. Zhang, R. Bommena, X. Sun, S. R. J. Brueck and S. D. Hersee, *Initial Nanoheteroepitaxial Growth Stages of GaAs on Si(100) by OMVPE*, Jour. Cryst. Matl. 30, 812-816 (2001)

D. Zubia, S. Zhang, R. Bommena, X. Sun, S. R. J. Brueck and S. D. Hersee, *Initial Nanoheteroepaxial Growth of GaAs on Si (100) by OMVPE,* Jour. of Electronic Matls. **30**, 812-816 (2001)

S-C. Lee, K. J. Malloy and S. R. J. Brueck, *Nanoscale Selective Growth of GaAs by Molecular Beam Epitaxy*, Jour. Appl. Phys. **90**, 4163-4168 (2001).

S-C. Lee, L. R. Dawson, K. J. Malloy and S. R. J. Brueck, *Molecular Beam Epitaxial Growth of One-Dimensional Rows of InAs Quantum Dots on Nanoscale-Patterned GaAs*, Appl. Phys. Lett. **79**, 2630-2632 (2001).

S. R. J. Brueck, S. D. Hersee and D. A. V. Zubia, special section editors, *Introduction to the Feature Section on the Growth of Heterostructure Materials on Nanoscale Substrates*, IEEE Jour of Quantum Electronics **38**, 973-974 (2002).

S. D. Hersee, D. Zubia, X. Sun, R. Bommena, M. Fairchild, S. Zhang, D. Burckel, A. Frauenglass and S. R. J. Brueck (invited paper), *Nanoheteroepitaxy for the Integration of Highly Mismatched Semiconductor Materials,* IEEE Jour. of Quantum Electronics **38**, 1017-1028 (2002).

S. C. Lee, L. R. Dawson and S. R. J. Brueck, *Dynamical Faceting and Nanoscale Lateral Growth of GaAs by Molecular Beam Epitaxy,* Jour. Crystal Growth **240**, 333-339 (2002).

S. C. Lee, A. Stintz, and S. R. J. Brueck, *Nanoscale Limited Area Growth of InAs Islands on GaAs(001) by Molecular Beam Epitaxy,* Jour. of Appl. Phys. **91**, 3282-3288 (2002).

Q. Li, S. M. Han, S. R. J. Brueck, S. Hersee, Y-B. Jiang and H. Xu, *Selective Growth of Ge on Si(100) through Vias of SiO2 Nanotemplate using Solid Source Molecular Beam Epitaxy*, Appl. Phys. Lett. **83**, 5032-5034 (2003).

X-Y. Sun, M. Fairchild, R. Bommena, S. D. Hersee, D. Burckel, A. Frauenglass and S. R. J. Brueck, *Defect Reduction Mechanisms in the Nanoheteroepitaxy of GaN on SiC*, Jour. Appl. Phys. **95**, 1450-1454 (2004).

S. C. Lee, X. Y. Sun, S. D. Hersee and S. R. J. Brueck, *Growth of GaN on a Nanoscale Faceted Si Substrate by Metal-Organic Vapor Phase Epitaxy*, Proc. 2003 Int'l Conf. On Compound Semiconductors.

S. C. Lee, X. Y. Sun, S. D. Hersee, F. Xu and S. R. J. Brueck, *Spatial Phase Separation of GaN Selectively Grown on a Nanoscale Faceted Si Surface,* Appl. Phys. Lett. **84**, 2079-2082 (2004).

S. C. Lee and S. R. J. Brueck, *Equilibrium Crystal Shape of GaAs in Nanoscale Patterned Growth*, J. of Appl. Phys. **96**, 1214-1218 (2004).

S. Birudavolu, N. Nuntawong, G. Balakrishnan, Y. C. Xin, S. Huang, S. C. Lee, S. R. J. Brueck, C. P. Hains and D. L. Huffaker, *Selective Area Growth of InAs Quantum Dots formed on a Patterned GaAs Substrate*, App. Phys. Lett. **85**, 2337-2339 (2004).

S. C. Lee, L. R. Dawson, and S. R. J. Brueck, *Heteroepitaxial Selective Growth of $In_xGa_{1-x}As$ on SiO_2-Patterned GaAs(001) by Molecular Beam Epitaxy,* Jour. Appl. Phys. **96**, 4856 – 4865 (2004).

S. C. Lee, L. R. Dawson, B. Pattada, S. R. J. Brueck, Y.-B. Jiang and H. F. Xu, *A Strain-Relieved, Dislocation-Free $In_xGa_{1-x}As$ (100) Heterostructure by Nanoscale Patterned Growth,* Appl. Phys. Lett. **85**, 4181-4183 (2004).

S. C. Lee and S. R. J. Brueck, *Nanoscale Two-Dimensional Patterning on Si (001) by Large-Area Interferometric Lithography and Anisotropic Wet Etching,* Jour. Vac. Sci. Technol. **B22**, 1949 - 1952 (2004)

S. C. Lee, X. Y. Sun, S. D. Hersee and S. R. J. Brueck, *Phase Control of GaN on Si by Nanoscale Faceting in Metal-Organic Vapor-Phase Epitaxy,* J. Cryst. Growth **272**, 2-8 (2004).

S. C. Lee, B. Pattada, X. Y. Sun, S. D. Hersee, Y.-B. Jiang, H. Xu, and S. R. J. Brueck, *Nanoscale spatial phase modulation of GaN on a V-Grooved Si Substrate – Cubic Phase GaN on Si(001) for Monolithic Integration,* IEEE Jour. of Quant. Electron. **41**, 596-605 (2005).

6. Lithography Applications - Nanophotonics

S. R. J. Brueck, *Radiation from a Dipole Embedded in a Dielectric Slab*, Jour. Sel. Top. in Quantum Electron. JSTQE-6, 899-910 (2000)

B. K. Minhas, W. Fan, K. Agi, S. R. J. Brueck and K. J. Malloy, *Metallic Inductive and Capacitive Grids – Theory and Experiment,* Jour. Opt. Soc. Amer. **A19**, 1352-1359 (2002).

Sharma, S. H. Zaidi, P. Logofãtu and S. R. J. Brueck, *Optical and Electrical Properties of Nanostructured Metal-Silicon-Metal Detectors,* IEEE Jour. of Quant. Electron. **38**, 1651-1660 (2002).

S. R. J. Brueck, V. A. Smagley and P. G. Eliseev, *Radiation from a Dipole Embedded in a Multilayer Dielectric Slab*, Phys. Rev. **E68**, 036608 (2003)

J. L. phase, J. D. Bruno, J. T. Pham, D. E. Wortman, S. Zhang and S. R. J. Brueck, *Single-Longitudinal-Mode Emission from Interband Cascade DFB Laser with a Grating Fabricated by Interferometric Lithography,* Proc. IEE-Optoelectronics **150**, 288-292 (2003).

Shuang Zhang, Wenjun Fan, A. Frauenglass, B. K. Minhas, K. J. Malloy and S. R. J. Brueck, *Fabrication of 1D and 2D Vertical Nanomagnetic Resonators,* Jour. Vac. Sci. Technol. **B22,** 3327-3330 (2004).

Shuang Zhang, Wenjun Fan, A. Frauenglass, Babar Minhas, K. J. Malloy and S. R. J. Brueck, *Mid-Infrared Resonant Magnetic Nanostructures Exhibiting a Negative Permeability,* Phys. Rev. Lett. **94**, 037402 (2005).

Wenjun Fan, Shuang Zhang, Babar Minhas, Kevin J. Malloy and S. R. J. Brueck, *Enhanced Infrared smission through Subwavelength Coaxial Metallic Arrays,* Phys. Rev. Lett. **94**, 033902 (2005).

7. Lithography Applications - Nanomagnetics

S. B. Field, S. S. James, J. Barentine, V. Metlushko, G. Crabtree, H. Shtrikman, B. Ilic and S. R. J. Brueck, *Vortex Configurations, Matching and Domain Structure in Large Arrays of Artificial Pinning Centers,* Phys. Rev. Lett. **88**, 067003-1-3 (2002).

V. Metlushko, U.Welp, G.W. Crabtree, R. Osgood, S.D. Baker, L.E. DeLong, Zhao Zhang, S.R.J. Brueck, B. Ilic, K.Chung, P.J. Hesketh, *Interstitial Flux Phases in a Superconducting Niobium Film with a Square Lattice of Artificial Pinning Centers, Phys Rev B, Vol 60, No 18, 585-588, November 1999.*

8. Lithography Applications - Nanofluidics

M. J. O'Brien, P. Bisong, L. K. Ista, E. M. Rabinovitch, A. L. Garcia, S. S. Sibbett, G. P. Lopez and S. R. J. Brueck, *Fabrication of an Integrated Nanofluidic Chip using Interferometric Lithography*, Jour. Vac. Sci. Technol. **21**, 2941-2945 (2003).

9. Lithographically-Directed Self-Assembly

Deying Zia and S. R. J. Brueck, *A Facile Approach to Directed Assembly of Patterns of Nanoparticles using Interference Lithography and Spin-Coating,* Nanoletters **4**, 1295-1299 (2004).

Deying Xia, Dong Li, Abani Biswas and S. R. J. Brueck, *Directed Self-Assembly of Silica Nanoparticles on Nanoscale Patterned Surfaces by Spin-Coating,* Advanced Materials **16**, 1427 - 1432 (2004).

Deying Xia and S. R. J. Brueck, *Lithographically Directed Deposition of Silica Nanoparticles using Spin Coating,* Jour. Vac. Sci. Technol. **22**, 3415-3420 (2004).

10. Lithography Applications – Etch Diagnostics

H. Gerung, C. J. Brinker, S. R. J. Brueck and S. M. Han, *In Situ Real-Time Monitoring of Profile Evolution during Plasma Etching of Mesoporous Low-Dielectric-Constant SiO_2,* J. Vac. Sci. Technol. **A23**, 347-354 (2005).

11. Lithography Applications – Deposition Strain

S. J. Hearne, S. C. Seel, J. A. Florio, C. W. Dyck, W. Fan and S. R. J. Brueck, *Quantitative Determination of Tensile Stress Creation during Island Coalescence using Selective-Area Growth*, Jour. of Appl. Phys. **97**, 083530 (2005).

Patents

Based on work partially funded under the MURI program was issued.

U. S. Patent 6,596,377 B1 issued July 22, 2003
Thin-Film Product and Method of Forming
Stephen D. Hersee, David Zubia, Steven R. J. Brueck, and Saleem H. Zaidi

U. S. Patent 6,685,841 issued February 3, 2004
Nanostructured Devices for Separation and Analysis: Application to Biological Membranes
G. P. Lopez, S. R. J. Brueck, L. Ista, Michael O'Brien, and, Stephen D. Hersee

Demonstration of Two-Photon Lithography

Ch. J. Schwarz[a,b], A.V.V. Nampoothiri[a,c], J.C. Jasapara[d], W. Rudolph[a],and S.R.J. Brueck[a,e]
Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106

Two-photon exposure is demonstrated using ~12 fs duration pulses at a center wavelength of 800 nm and a standard I-line photoresist. Using direct-writing with a focused Gaussian beam, developed resist features as small as 300 nm, compared with the ~ 1 μm focal spot diameter, have been achieved. The experiments reveal substantial differences (~ 10^4) in the two-photon absorption coefficient for multiple pulse trains (100 MHz) and for single pulses. A simple model involving photoexcitation of an intermediate conformation of the photoactive compound and relaxation back to the ground state can explain this observation. The resolution enhancement of $1/\sqrt{2}$ resulting from the two-photon absorption is demonstrated by writing two lines as close as possible to each other with both one- and two-photon absorption mechanisms.
Journal of Vacuum Science and Technology B, to be published Nov./Dec. 2001

Introduction

The continuing demands for smaller feature sizes in microelectronics are driving optical lithography ever further into the sub-wavelength regime. Many different resolution enhancement techniques, including off-axis illumination, optical proximity correction, phase-shift masks and imaging interferometric lithography have been demonstrated. These advances, along with the continuing migration to shorter wavelength sources and to higher NA lenses, have been successful in extending the manufacturing use of optical lithography to dimensions well beyond the expectations of even a few years ago.

It has long been known[1-3] that nonlinear processes, such as two-photon absorption in which the photoresist responds to the square of the intensity, offer possibilities of higher resolution. This can be understood from elementary arguments. Consider an interferometric lithography arrangement where two coherent laser beams are incident on a photoresist layer at angles of $\pm\theta$. Then the aerial image intensity is $I(x) \propto \cos^2(kx\sin\theta) \propto [1+\cos(2kx\sin\theta)]/2$, where $k = 2\pi/\lambda$ with λ the optical wavelength, which is the appropriate expression for one-photon absorption. For two-photon absorption, the resist responds to $I^2(x) \propto \cos^4(kx\sin\theta) \propto [3+4\cos(2kx\sin\theta)+\cos(4kx\sin\theta)]/8$. The higher harmonic content is evident in the last term of this expression. Plotting these expressions shows that the density of peaks in the response is the identical in both cases, but that the linewidth is reduced in the two-photon case. Yablonovitch et al. have proposed[2] a scheme to eliminate the lower spatial frequency term and realize the full factor of two enhancement in resolution inherent in these equations. The photoresist exposure and development process is, in itself, highly nonlinear, thus higher harmonics of the fundamental frequency are routinely achieved in one-photon lithography (e. g. vertical sidewalls). Exploitation of all of these nonlinear processes offers significant potential for extending the resolution range of optical lithography yet further into the subwavelength domain.

In this paper we present results of direct writing on a photoresist using focused Gaussian beams. The possible enhancement in the area density of structures using an n-photon absorption process is n. This results from the reduction of the FWHM of a Gaussian by a factor of $\sqrt{n}$ when it is raised to the n-th power. In the first part of the paper, we investigate two-photon absorption of a photoresist and compare exposures with single pulses and multiple pulse trains. In the second part we demonstrate the resolution enhancement by comparing the minimum achievable separation of two lines written with single and two-photon exposure.

Two-photon exposure of isolated lines

We assume that a certain number density of photoactive compound (PAC), N_{th}, has to be excited in order to expose the photoresist, independent of whether one- or two-photon absorption occurs. This can be expressed as

$$N_{th} = N_1 = \frac{\alpha}{h\nu_1}\int I_1(t)dt = N_2 = \frac{\beta}{2h\nu_2}\int {I_2}^2(t)dt \qquad (1)$$

where α (β) is the one- (two-) photon absorption coefficient in units of cm^{-1} (cm/W), I is the laser intensity, ν is the laser frequency, and the index 1 (2) describes the one (two-) photon excitation. Note that $\nu_1 = 2\nu_2$ since both absorption processes excite the same transition. Typical values for the one-photon absorption coefficient for conventional photore-

a. Also Department of Physics and Astronomy
b. Email: schwarz@chtm.unm.edu
c. Current address: Washington State University, Pullman, WA
d. Current address: Lucent Tech., Bell Laboratories, Murray Hill, NJ
e. Also Department of Electrical and Computer Engineering

sist are engineered to be on the order of $\alpha \approx 104$ cm-1, while for the two-photon absorption coefficient b ≈10-9 cm/W6,7. Thus an intensity on the order of $I \sim \alpha/\beta \sim 10^{13}\,\mathrm{W/cm^2}$ is required for two-photon lithography. In order to keep the total energy realistic and to avoid thermal damage to the resist and the underlying materials, this intensity must be deivered using a short-pulse (femtosecond) source to take advantage of the very high peak powers. It is emphasized that it is the peak intensity, not the exposure dose that is increased by this large factor. As is discussed below, with fs pulse illumination, the exposure energy remains comparable to that for a one-photon case.

A schematic diagram of the experimental setup is shown in Fig.1. A Kerr-lens modelocked Ti:sapphire laser with a repetition rate of 100 MHz provides pulses with a duration of 10 fs at a center wavelength of 800 nm. The pulses are focused with a microscope objective onto the photoresist. Assuming Gaussian beam characteristics the focal spot diameter was ~ 1 mm. In order to avoid temporal broadening in the objective, the pulses were dispersion-precompensated with a prism pair and a third-order diffraction mirror4. The autocorrelation of the pulses obtained in the focal plane of the objective is shown in inset a) of the Fig.1 and indicates a pulse width of 12 fs. We used a Shipley 510 I-line photoresist in a diluted form (1:1), which was measured to have a film thickness of about 250 nm after spinning. The absorption coefficient is α ≈0.5x104cm-1 at 400 nm according to the manufacturer5. There is essentially no single photon absorption at the fundamental of our laser (800 nm). Consequently, no exposure of the resist was observed with our laser operating in continuous wave (cw) mode to total fluences well above those required for two-photon exposure.

To study the two-photon absorption quantitatively we proceeded as follows. In order to write a line, the sample was moved perpendicular to the beam using a piezo driven translation stage. A small longitudinal translation component characterized by the angle φ (see Fig.1, inset b)) was introduced to assure that the sample passed through the focal plane during the scan. Consequently the intensity at the sample varies along each line. Several lines were written with different scan speeds, varying the total dose (see Fig.2).

Each line was exposed to a different dose, that is to a different number, m, of pulses per line element. As expected longer lines were observed with higher doses. The length, l, of the developed line as a function of m is shown in the lower part of Fig.2. This data was used to determine the two-photon absorption coefficient. The number density of ex-

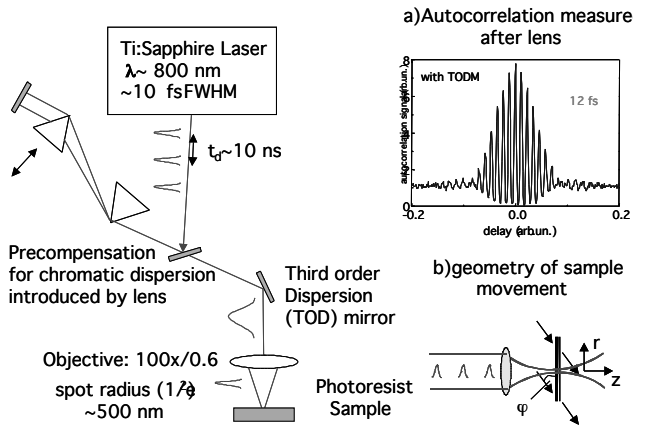


Fig. 1: Experimental arrangement for fs-pulse train, two-photon exposure of photoresist. Inset a) shows the pulse autocorrelation measurement of a 12-fs pulse at the focal point of the lens. Inset b) shows the geometry of sample motion during the experimental scans.

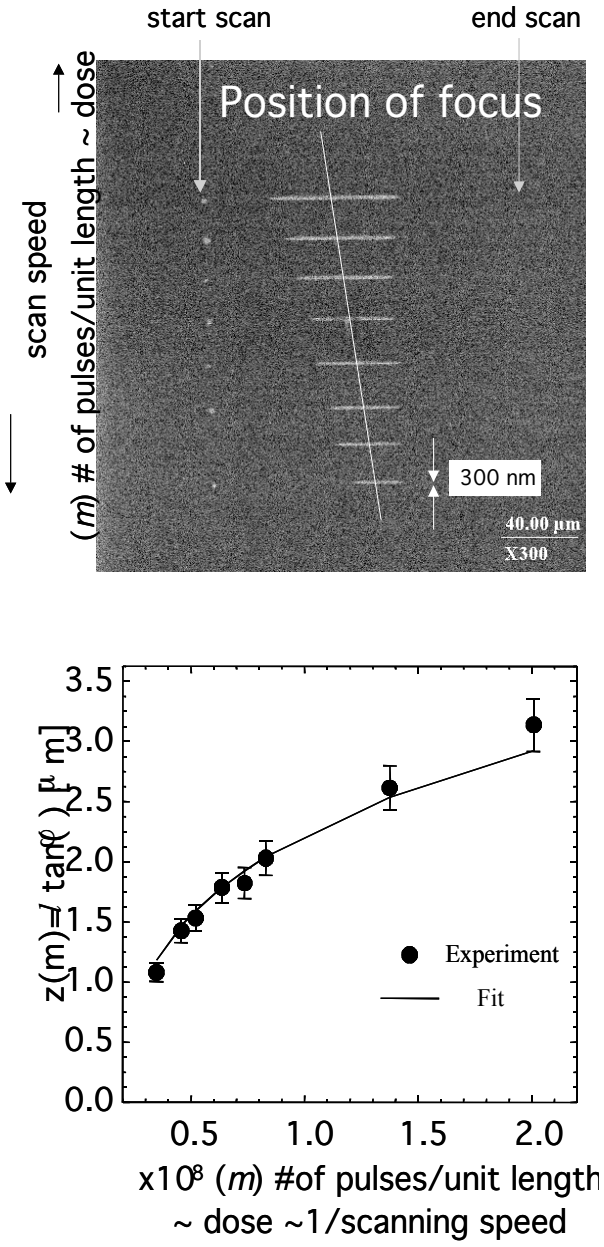


Fig. 2: Top panel shows the developed sample. The vertically separated traces correspond to different developed lengths (*l*) due to different scan speeds. The bottom panel shows the fit of these exposed length results to a simple model for extracting the two-photon absorption coefficient.

cited molecules is related to the intensities according to Eq. (1). A permanent structure can be expected in a region (r,z) characterized by

$$N_{th} \le \frac{\beta}{2h\nu_2} m \int I^2(z,r,t)dt = N_2(r,z) \qquad (2)$$

The length of the observed lines, *l*, can be obtained from the requirement $N_2(z, r=0) \ge N_{th}$. For a Gaussian pulse profile we find

$$l(m) = z_0 \cot(\varphi) \sqrt{\left[\frac{\beta}{2h\nu} \frac{I_0^2}{N_{th}} \frac{\sqrt{\pi}\,\tau}{\sqrt{8\ln 2}}\right]^{1/2} m^{1/2} - 1} \qquad (3)$$

where z_0 is the Rayleigh range of the beam, $I_0 = I(r = 0, t = 0)$, and τ is the pulse width. From a fit of Eq. (3) to the data we find $\beta \approx 2.6\text{x}10^{-14}$ cm/W. This deviates considerably from typical literature6 values, which are on the order of ~ $10^{-9} - 10^{-10}$cm/W.

To investigate this further the photoresist was exposed by single pulses from a Ti:sapphire amplifier. The pulse duration was ~30 fs. For a spot diameter of 330 μm, an exposure threshold energy of 25 μJ was determined. From Eq. (2) and m =1 we obtain $\beta \sim 6\text{x}10^{-9}$ cm/W assuming again a Gaussian spatial and temporal profile. This value is consistent with that found for other organic materials7. This threshold corresponds to a dose of ~ 25 mJ/cm^2 comparable to the single photon (I-line) dose-to-clear for this resist.

For a tentative explanation of the difference between single and multiple pulse exposure we look at the photochemical processes following the excitation. With single photon absorption experiments it was found that the photoactive compound undergoes a rapid transition to an intermediate conformation (IC) with a lifetime extending to as much as several hundred ms that can absorb the incident radiation8. Qualitatively, we can explain the observed discrepancy between exposures using single pulses and pulse trains if we assume that the excited IC can decay back into the ground state of the initial compound, see Fig. 3. Based on a simple rate equation analysis, we calculated N_2 as a function of m for this scenario and for the case when every absorption event leads immediately to the formation of the final compound. From the graphs in Fig.3 it is clear that parameters can be found that vary the excitation efficiency by several orders of magnitude. More detailed experiments and modeling are necessary to confirm this hypothesis.

The minimum observed linewidth was ~ 300 nm, much less than the laser spot size. Because of other nonlinearities in the resist response, similar sub-optical resolution results can be achieved with one-photon excitation. A multiple line exposure, showing an increase in the available packing density of the lines and providing clear evidence of the improved resolution associated with two-photon absorption is described below.

Resolution enhancement using two-photon lithography

In order to determine the minimum spacing possible and to demonstrate the improved resolution of two-photon, direct-write optical lithography, pairs of lines were written. We compared a two-photon exposure at 800 nm with a one-photon exposure at 400 nm. The latter wavelength was obtained by frequency doubling the output of the fs Ti:sapphire laser. Because different focusing lenses were used, we first determined the spot sizes with a knife-edge method. The results were $2\omega_0 \sim 17.6$ μm (800 nm) and ~ 13.2 μm (400 nm). It should be noted that even though the focal spot size was smaller for the 400 nm one- photon exposure, the achievable developed linewidth was smaller for the 800 nm, two-photon exposure, as shown in Fig.4. Also, in the two-photon case the line-edge slope is higher than in the one-photon case - as expected.

In the next step, two lines were written as close to each other as possible in this setup without having them merge into a single, wide line. In order to compare both exposure types near threshold we varied the scan speed (exposure dose) during one scan to assure that the threshold was reached. Pairs of lines with different separation were written. A smaller distance between the lines was achieved in the

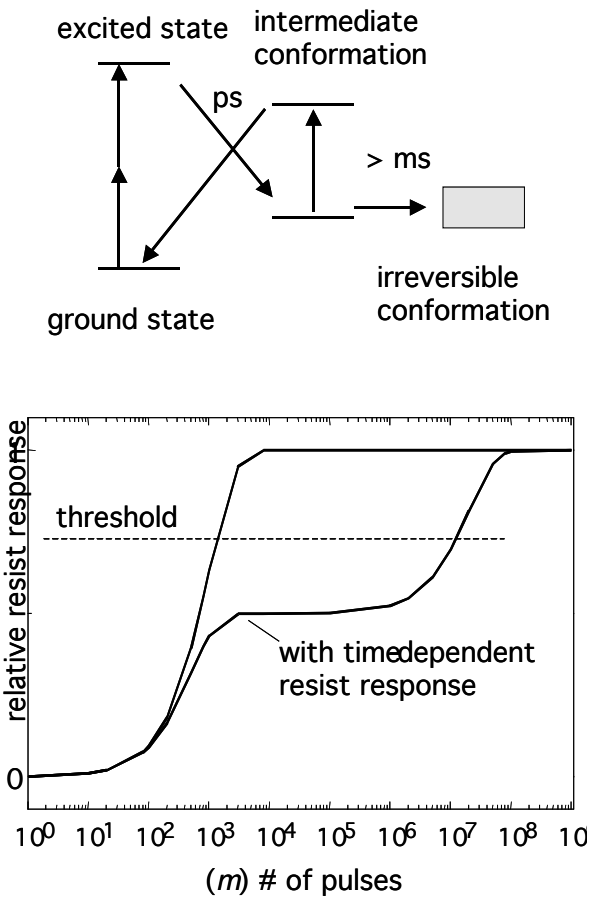


Fig. 3: Possible mechanism for the observed differences in two-photon absorption coefficient between the multiple- and single-pulse experiments. Top level diagram shows reabsorption in an intermediate conformation and relaxation back to the ground state of the PAC. The bottom plot shows the results of a simple rate equation analysis of this model that can qualitatively explain the observations.

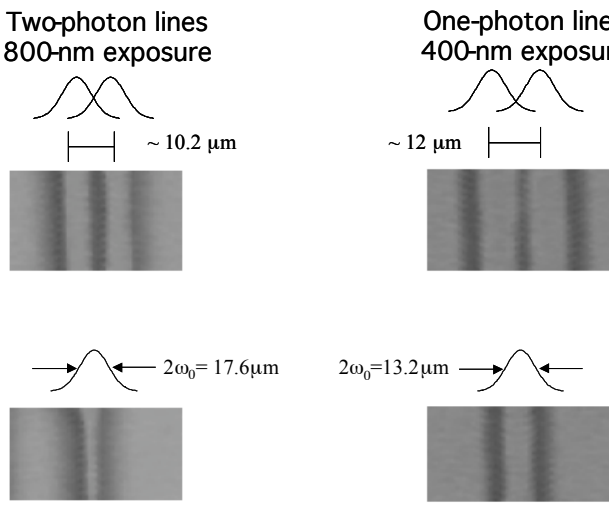


Fig. 4: Comparison of one-photon (400 nm) and two-photon (800 nm) exposures of both single and double lines. The two-photon results show: a smaller achievable linewidth (even with a larger spot size); an improved edge definition; and the expected $1/\sqrt{2}$ improvement in achievable linear packing density.

two-photon case (10.2 μm) compared to the one-photon case (12 μm). The corresponding pairs with minimum spacing are shown in Fig.4. Normalizing the minimum achievable separation of the line pairs to the corresponding spot diameters, gives a ratio of the distance between the two-photon exposed and the one-photon exposed line pairs of 0.64, which is close to the expected ratio of $1/\sqrt{2}$.

For these first experiments, we have investigated the resolution for serial writing of individual structures. While this is of intrinsic interest, extending these results to parallel lithographic processes will be necessary for application to manufacturing. There are at least two approaches to this requirement. Yablonovitch et al2 have shown, using coherently related, frequency shifted sources, that the ultimate doubled resolution of periods of λ/4 can be achieved in a single exposure, at least for grating structures. Alternatively, two mask exposure patterns can be interpolated, achieving the same two-photon-enhanced resolution for arbitrary patterns, but at the expense of a second alignment and exposure step. This interpolation is essentially equivalent to the serial process we have demonstrated.

Conclusion

The two-photon absorption coefficient of the photoresist Shipley 510 was measured to be $\beta \sim 6\text{x}10^{-9}$ cm/W. A dramatic drop in the exposure efficiency was observed with multiple pulse excitation. This is tentatively explained by the existence of a photoactive intermediate state whose excitation leads to re-population of the initial ground state.

We have demonstrated that, in direct writing using Gaussian beams, two-photon exposure enhances the resolution by a factor of $\sim \sqrt{2}$ providing a factor of two increase in the achievable writing density in two dimensions. This is possible, since the two-photon process involves higher harmonic spatial frequencies due to the intensity squared resist response. In addition, an improvement of the contrast for a given resist chemistry is reported. As lithography progresses to vacuum uv wavelengths, the advantage of improved resolution at a wavelength for which lenses are still transparent may be significant. With currently available fs laser technology, three-photon absorption at the fundamental or two-photon absorption of the second harmonic could effectively extend this technique to shorter wavelengths and higher resolution. Alternatively, photoresists optimized for nonlinear absorption could possibly adapting existing narrow-bandwidth UV laser sources and high-NA optics to nonlinear exposure schemes.

Acknowledgement: Partial support for this work was provided by ARO/MURI and by the NSF.

References

[1] E. S. Wu, J. H. Strickler, W. R. Harrell and W. Webb, "Two-photon lithography for microelectronic application", Optical/Laser Microlithography V, Proc. SPIE 1674, 776 (1992).

[2] Eli Yablonovitch, Rutger B. Vrijen, "Optical projection lithography at half the Rayleigh resolution limit by two-photon exposure", Optical EngineeringVol. 38(02), pp.334-338 (1999)

[3] George Witzgall, Rutger Vrijen, and Eli Yablonovitch, Vinh Doan and Benjamin J. Schwartz, "Single-shot two-photon exposure of commercial photoresist for the production of three-dimensional structures", Optics Letters, Vol. 23, No.22, p.1745-1747 (1998).

[4] J. Jasapara, W. Rudolph, "Characterization of sub-10-fs pulse focusing with high-numerical- aperture microscope objectives", Opt. Lett. 24 777-779 (1999).

[5] www.shipley.com

[6] I.M. Catalano, A.Cingolani, and M. Lepore, "Two-photon spectroscopy in AgCl,", Solid State Commun. 35, 833-836 (1980).

[7] Richard L. Sutherland, "Handbook of Nonlinear Optics", Marcel Dekker, Inc., ISBN: 0-8247-9426-5 p.564 (1996)

[8] James R. Sheats, "Reciprocity Failure in Novolak/Diazoquinone Photoresist with 364-nm Exposure", IEEE Transactions on Electron Devices, Vol. ED-35, NO.1, 129-131 (1988).

5 There Are No Fundamental Limits to Optical Lithography

Steven R. J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106

5.1 Introduction

Moore's law, the continued doubling of semiconductor integrated circuit capability on an ~18-month time scale, has become ingrained not only in technology but also in popular culture. Not only electronics, but also communications (in bandwidth capability of a fiber and in total installed bandwidth) and storage technologies (in magnetic storage bits/cm^2) have their own "Moore's law" describing exponential increases in capability. Together, these technologies define the "information age" and are dramatically impacting society.

Advances in optical lithography have been responsible for much of this improvement, by some estimates as much as 50% of the advance in scaling of silicon circuits. Descriptions of the evolution of optical lithography from the beginning of silicon integrated circuits [wavelength of 432 nm (G-line of Hg); proximity printing] to today's advanced scanners [wavelength of 193 nm (ArF laser); 4x reduction optics; NA ~ 0.75, phase shift masks with optical proximity correction] have been presented elsewhere by the pioneers of the field.[1]

Many corollaries to Moore's law have been advanced, aimed at pointing out impending limits or attendant consequences. One familiar one is that the cost of a fabrication facility is also exponentially increasing, is today at several billion dollars, and simple extrapolation will tell you when it will exceed the gross world product. Another, more pertinent to this discussion, is that the "end" of optical lithography is always two IC generations away. The purpose of this article is to examine the fundamental issues that determine the limits of optical lithography. While some comments will be made concerning practical limitations, our purpose is to outline the realm of the possible in contrast to the probable. There are many practical issues that will ultimately limit the ability to continue the downward scaling of lithography. And new technologies such as nanoimprint and self-assembly are being advanced that may make the limits of optical lithography a moot point. My own bias is that combinations of these techniques will enable inexpensive access to the nanoscale without requiring radically new approaches. Ultimately, the manufacturing decision will be made by economics, not by theo-

International Trends in Applied Optics, A. H. Guenther, ed., SPIE Press, Bellingham, WA (2002), pp. 85-109

retical discussions of possibilities. Nevertheless, there is some utility in asking what physics and chemistry allow, and how possibly we might get there.

A useful starting point is the International Technology Roadmap for Semiconductors (ITRS).[2] This document is continually updated by the semiconductor industry; the latest version, published in December 2001, provides the best understanding of the semiconductor industry both as to what is needed and how they are going to get it. The table relevant to lithography is reproduced here to set the stage on the scales that will be required by this industry.

This table is divided into near-term, with one-year increments, and long-term (shaded) sections, with three-year increments. Memory requirements are still driving lithography, although the distinction between memory and microprocessors is lost as the roadmap progresses.

While much of the present discussion is framed in the context of silicon integrated circuits, it is worth noting that lithography has a much broader impact. The National Nanotechnology Initiative in the United States[3], and similar initiatives elsewhere in the world, is recognition of the emerging importance of processes characterized by nanometer scale lengths, at which macroscopic scaling breaks down and quantum and interface effects dominate material behavior. The nanometer scale is also the scale of many biological processes. Continued advances in our abilities to manipulate material at these scales will likely lead to many new technologies. A characteristic of nature is the integration of many length scales to produce complexity and enhance functionality. Manifestly, lithographic techniques, spanning many length scales, will play a critical role in the development of nanotechnology.

5.2 Optical Lithography

The starting point for a discussion of optical lithography is typically framed in terms of the Rayleigh resolution criteria that describe the resolution and depth-of-focus of an optical system. These are:

$$CD = \kappa_1 \frac{\lambda}{NA} \tag{5.1}$$

Table 5.1 ITRS key lithography-related characteristics by product type

Year of Production	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*	*2007*	*2010*	*2013*	*2016*
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65	45	32	22
MPU ½ Pitch (nm)	150	130	107	90	80	70	65	45	32	22
MPU Printed Gate Length (nm)	90	75	65	53	45	40	35	25	18	13
MPU Physical Gate Length (nm)	65	53	45	37	32	28	25	18	13	9

and

$$\mathrm{DOF} = \kappa_2 \frac{\lambda}{2[1-\cos(\mathrm{NA})]} \approx \kappa_2 \frac{\lambda}{\mathrm{NA}^2} \tag{5.2}$$

where CD is the critical dimension or smallest printed feature (line or space), DOF is the depth-of-focus, λ is the optical wavelength, NA is the numerical aperture of the optical system and κ_1 and κ_2 are constants, typically of order unity. These equations describe the aerial image resolution and depth-of-focus of a conventional, incoherently illuminated, optical system. Innovations in both the mask and the illumination and imaging systems allow some improvement, usually denoted as a smaller κ_1 in the context of Eq. (5.1). These will be described below. As we will see, this is only a part of the story. Equally important is the interaction of the actinic radiation with the photoresist. Nonlinearities in this interaction can be, and are being, used to further extend the accessible range of optical lithography.

Rayleigh first obtained these equations in connection with the resolving power of telescopes used for astronomical observations. For what follows, it is worth noting that the resolution limit [Eq. (5.1)] was intended to refer not to the width of a single object, but to the ability to distinguish two separate objects. "According to this (Rayleigh's) criterion, two images are just resolved when the principal maximum of one coincides with the first minimum of the other."[4] The physical basis of these equations is, of course, the diffraction of light. For a point source at infinity, the radial dependence of the intensity distribution in the focal plane of an unaberrated optical system is given by

$$I(r) = 2\left[\frac{J_1(2\pi \mathrm{NA} r/\lambda)}{2\pi \mathrm{NA} r/\lambda}\right]^2 I_o \tag{5.3}$$

where J_1 is the Bessel function of the first order. The Rayleigh resolution criterion is that the second peak should be at the first zero of $I(r)$ or $r_0 = 0.61\lambda/\mathrm{NA}$ which corresponds to Eq. (5.1) with $\kappa_1 = 0.61$. The important point is that the Rayleigh resolution equation refers to the ability to distinguish that there are two spots, not to the size of a single spot. This has critical implications for the limits of lithography as we shall see below.

Three approaches to reducing the CD are suggested by Eq. (5.1): reduce the wavelength, increase the optical system NA, and reduce κ_1. Remarkable progress has been made in each of these directions. As noted above, the wavelength has progressed from 432 nm (G-line of Hg) to 193 nm (ArF excimer laser). There are substantial efforts underway to further reduce the wavelength to 157 nm (F_2 excimer laser); as well as preliminary investigations of even shorter wavelengths such as 126 nm (Ar_2 excimer) and 121 nm (hydrogen Lyman-α). Each of these wavelength shifts must be accompanied by an extensive, and expensive, materials effort to establish both optical and photoresist materials suitable for the new wavelength. All of the above wavelengths are in spectral regions where both

transmissive and reflective materials are available, although good optical properties and material durability under actinic radiation clearly get more difficult to achieve with each generation. There is also a large, worldwide effort underway on extreme ultraviolet lithography (EUV) using a 13-nm wavelength laser-produced plasma source. At these short wavelengths, a fully reflective system must be used. The main point of this chapter is to discuss the limits of optical lithography, so EUV will not be further discussed except to note that EUV is an optical imaging technology and all of the concepts introduced herein apply in this short wavelength regime as well.

The next parameter is the numerical aperture (NA) of the optical system. Again, remarkable progress has been made with NA's increasing from ~0.1 in the first optical steppers to about 0.7 today, and probably over 0.8 in the near future. The usual approximation that the DOF is inversely proportional to NA^2 breaks down in this high NA regime and the DOF saturates at about $\lambda/2$. Eq. 5.2 is somewhat oversimplified as the NA is image dependent, and Eq. 5.2 is a worst case. Each new wavelength has to be introduced at an NA comparable to the last generation if any resolution enhancement is to be achieved.

Finally, there is κ_1. To some extent, κ_1 is a process control parameter. Figure 1 shows the progression of the intensity profile across two spots as κ_1 (or equivalently the distance between the spots) is varied. For $\kappa_1 = 1$, the two spots are clearly resolved with the intensity dip between the spots reaching almost to zero, for the Rayleigh limit, $\kappa_1 = 0.6$, the intensity dip is just visible. If the nonlinearities in the resist allow the region between the spots to be unexposed and at the peaks to be exposed, two lines can be printed. This is clearly a process control issue as well as a strictly imaging issue. The intensity varies from spot to spot and from exposure to exposure; tighter control is required to reliably print two features for the lower κ_1. As κ_1 is further decreased, the two spots merge into a single spot and there is no way to resolve them. As we will see below, it is possible to reduce κ_1 further by modifying the optical system. Several different approaches to this end have been investigated. They involve variations to the mask,

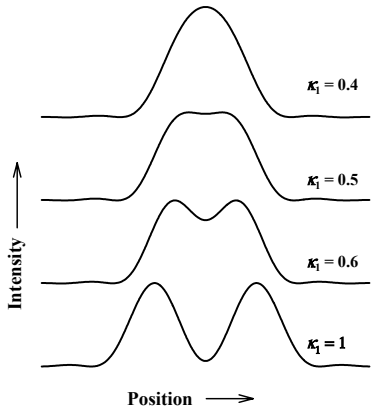


Figure 5.1 Intensity profiles of two spots in a imaging system as a function of κ_1.

to the illumination system, or to the imaging optics. Figure 5.2 shows a generic lithography system. The various "knobs" that have been explored to impact the resolution are indicated in the brackets. These will be discussed below after some additional background.

5.3 Fundamental limits of optical lithography.

Fourier optics[5] provides a good pathway to understanding the ultimate limits of lithography. The highest spatial frequency that can be supported by optics in free space corresponds to two coherent counterpropagating plane waves (or more accurately Gaussian beams), which interfere with a period of $\lambda/2$. It is possible to approach this limit very closely with a relatively simple interferometric lithography (IL) set-up where the two coherent optical beams are incident on a wafer at angles of $\pm\theta$ equally disposed about the surface normal giving a period of $\lambda/2\sin\theta$.[6-9] As an example, Fig. 5.3 shows a photoresist grating at a period of 108 nm written using a 213-nm coherent laser source (fifth harmonic of a 1.06-μm YAG laser). This corresponds to $\theta \sim 80°$ or NA = $\sin\theta \sim 0.985$. For IL the aerial

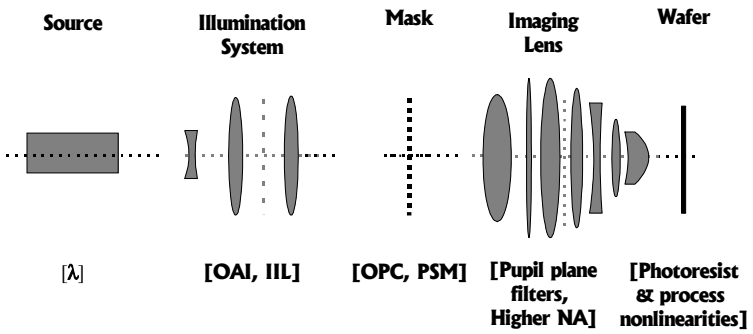


Figure 5.2 Generic lithography system consisting of a source, an illumination system, a mask, an imaging optical system and a photoresist coated wafer. Techniques for enhancing the resolution are indicated in the brackets.

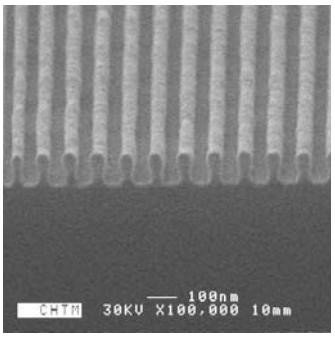


Figure 5.3 108-nm pitch photoresist grating written at $\lambda = 213$ nm and NA ~ 0.985.

image is simply

$$I(x) = I_0 \left| e^{+i\frac{2\pi \mathrm{NA}}{\lambda}x} + e^{-i\frac{2\pi \mathrm{NA}}{\lambda}x} \right|^2 = 2I_0 \left[1 + \cos\left(\frac{4\pi \mathrm{NA}}{\lambda} x \right) \right]. \qquad (5.4)$$

The corresponding frequency space picture is shown in Fig. 4. The aerial image contains a dc term and the positive and negative frequencies of the cosine. The very high NA puts these frequencies very close to the limiting frequencies of $\pm 2/\lambda$ represented by the darker (inner) circle. The outer circle represents the increased spatial frequency space that is available for optical propagation in a medium of refractive index n. For IL exposures such as shown in Fig. 5.3, immersion is a relatively simple optics issue and indices of ~ 1.5 are feasible, even at the very shortest current DUV wavelength of 157 nm.[10,11] For an imaging system, immersion will be much more problematic.

Also shown in Fig. 5.4 outside of both optics bandpass circles are two additional frequency components at $\pm 4\mathrm{NA}/\lambda$ that arise because of nonlinearities in the exposure and develop process. As is seen in Fig. 5.3, the final resist profiles exhibit nearly vertical sidewalls, corresponding to this higher frequency content. Thus the result of the total photoresist process – application, prebake, exposure, postbake, develop – is to add substantial higher harmonic content into the final image compared with the strictly band-limited aerial image. These nonlinearities allow extension of optics beyond the single-exposure linear systems limits as will be discussed in Sect. 5.8.

Resist processing is a complex, 3D chemical and physical process with nonlocal effects due to diffusion and dissolution kinetics. A simplified model[12], that neglects these nonlocal effects, is sufficient for the purposes of this presenta-

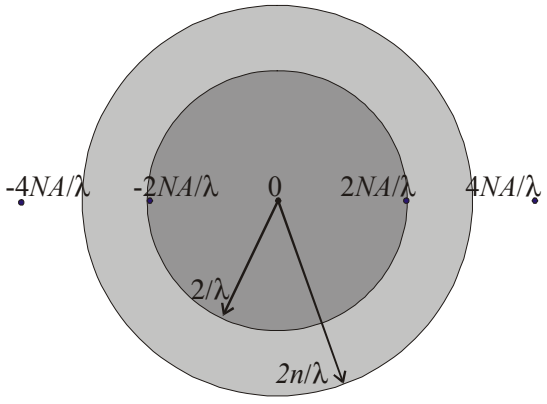


Figure 5.4 Frequency space representation of the grating in Fig. 5.3. There are aerial image components at dc (0) and at $\pm 2\mathrm{NA}/\lambda$. The maximum frequency permitted by the optical system is $2/\lambda$ or $2n/\lambda$ for a transmission medium of index n. Process nonlinearities add higher harmonics.

tion. In this model the final resist thickness, $T(x)$, is given by

$$T(x) = \max\left\{ T_0 \left[1 - \left(\frac{1 - e^{F(x)/F_0}}{1 - e^{-1}} \right)^n \right], 0 \right\} \tag{5.5}$$

where $F(x)$ is the local optical fluence (J/cm^2); F_0 is the fluence to clear the resist, and n is a coordination number that describes the resist nonlinearity. Figure 5 shows this function for various values of n. Novolak-based[13] I-line (365 nm) resists typically have n's of ~ 5 to 6. Today's high-speed chemically-amplified resists, as a consequence of the large thermal amplification inherent in their operation, have much higher effective n's. To a first approximation, it is sufficient to model their response as the sigmoidal threshold function shown for $n \to \infty$.

The result of applying this sigmoid response function to the sine-wave aerial image is the square wave response observed in Fig. 5.3 with the frequency content show in Eq. (5.5) where $\mathcal{R}$ is the nonlinear operator associated with the processing. Note that this frequency content extends well beyond the second-harmonic shown in Fig. 5.4, decaying only as $1/l$ where l is the harmonic number.

$$T(x) = \mathcal{R}[I(x)] = T_0 \frac{2\mathrm{NA}a}{\lambda} \sum_l \left[\frac{\sin\left(\frac{2\pi l \mathrm{NA}a}{\lambda} \right)}{\frac{2\pi l \mathrm{NA}a}{\lambda}} \right] e^{i\frac{4\pi l \mathrm{NA}}{\lambda}x} \tag{5.6}$$

This expression clearly highlights the higher harmonics added by the processing. Other post-develop processing also adds higher harmonic content to the final pattern. Examples are an isotropic oxygen plasma ash to thin the resist lines, or pattern transfer into the underlying materials. These and other fabrication steps are routinely used in the semiconductor industry as standard manufacturing processes. All of them modify the frequency content of the pattern, most often adding higher harmonics.

These higher harmonics offer the possibility of extending optical lithography beyond the linear systems limits imposed by diffraction. The difficulty is that all

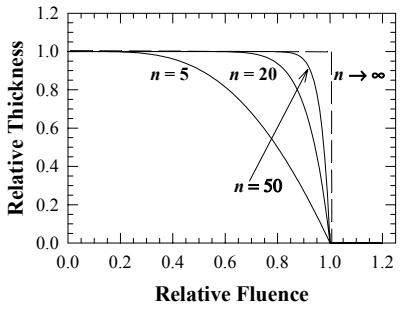


Figure 5.5: Resist response for different coordination numbers.

of the harmonics are in phase with the fundamental frequency component and are generally smaller in amplitude, so the impact of the nonlinearities is to sharpen the image, but not to increase the pattern density beyond the Rayleigh limit. In the next section we will describe a strategy for increasing the pattern density, still with IL and periodic patterns as the example. The application to arbitrary patterns will be discussed later in the chapter.

5.4 Nonlinear Lithographic Processes to Extend the Pattern Density

In the previous section, we showed that diffraction limits the one-dimensional pattern density in a single exposure to ~ λ/NA, while the nonlinearities in the processing results in many higher harmonics extending in frequency well beyond the diffraction limits of Eq. (5.1). These nonlinearities can be used to extend the available pattern density,[14] leading to the title of this chapter. The key is to apply the nonlinearities *between* exposure steps and combine images in such as way as to reduce or eliminate the lower frequency components, leaving the higher density of structures.

A one-dimensional example of a spatial period division is shown in Fig. 5.6 to illustrate the general concept. The top panel Fig. 5.6(a) shows a photoresist pattern atop a sacrificial layer on a substrate. The period of the pattern is limited by the optics to λ/2NA; the CD, in contrast, is not limited by the optics and is chosen to be less than λ/8NA. As discussed above, this does not violate any optical limits. In Fig. 5.6(b), the photoresist pattern has been transferred into the sacrificial layer, in the experimental case shown below the sacrificial layer was Si_3N_4 and lift-off and etching was used to transfer the pattern into this layer. In Fig. 5.6(c), a second photoresist layer has been spun, exposed and developed. The period and CD is the same as in the first case, but the pattern has been shifted by λ/4NA to interpolate the second pattern with the first. Finally this pattern is also transferred into the sacrificial layer [Fig. 5.6(d)], that layer is then used as a mask to transfer the total pattern into the substrate layers, and finally the sacrificial layer is removed leaving a final pattern at a period of λ/4NA, twice as dense as the single exposure optical limit.

The mathematics follows simply,

$$\begin{aligned} T(x) &= \mathcal{H}[I_1(x)] \oplus \mathcal{H}[I_2(x)] \\ &= \frac{a}{2}\left\{\sum_l \frac{\sin\left(\frac{2\pi l \mathrm{NA}a}{\lambda}\right)}{\frac{2\pi l \mathrm{NA}a}{\lambda}}\left[1+(-1)^l\right]\right\} e^{i\frac{4\pi l \mathrm{NA}}{\lambda}x} \\ &= a\sum_l \frac{\sin\left(\frac{4\pi l \mathrm{NA}a}{\lambda}\right)}{\frac{4\pi l \mathrm{NA}a}{\lambda}} e^{i\frac{8\pi l \mathrm{NA}}{\lambda}x}. \end{aligned} \tag{5.7}$$

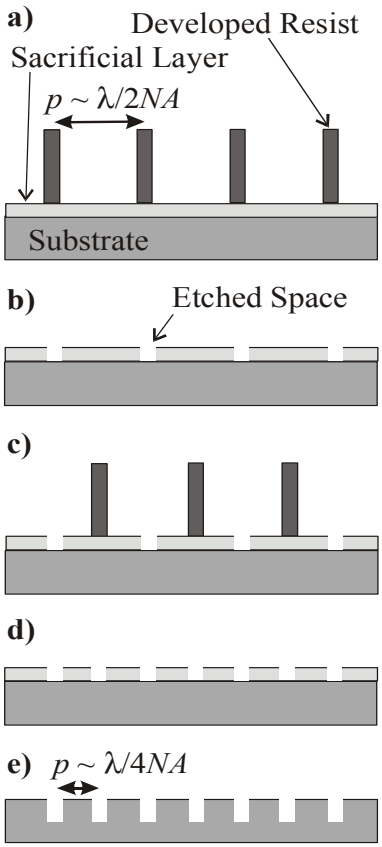
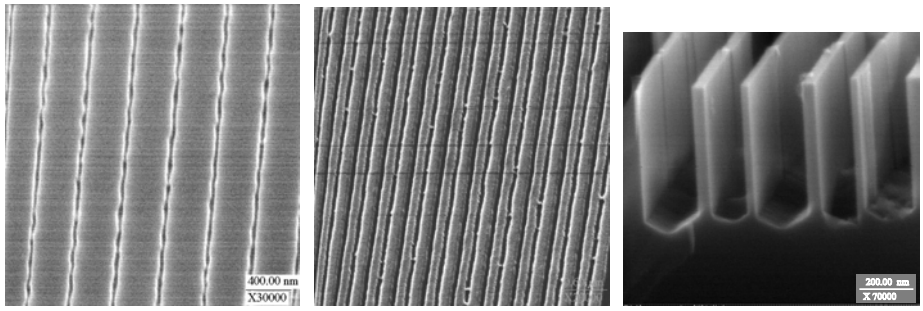


Figure 5.6 Sequence of steps to double the 1D pattern density by spatial period division.

The result is the combination of two exposures with the nonlinearities applied to each exposure independently. The symbol used to represent the combination is left ambiguous; in this case the combination is simple addition. In another case, for example the use of two resist layers sensitive at different wavelengths, it might be multiplication. The $p/4$ phase shift in the second exposure results in a factor of $(-1)^l$ in the Fourier coefficients, with the net result that all of the odd terms cancel leaving a $\sin(x)/x$ pattern at twice the original period or half the pitch.

A demonstration of this process is shown in Fig. 5.7. The top panel, Fig. 5.7(a), shows the etched grooves in a sacrificial nitride layer with a period p of 365 nm and a CD of ~ 50 nm. These grooves were made by the process outlined in Fig. 5.6: using positive photoresist, inverting the pattern with a Cr liftoff step, using the Cr as a hard mask to etch the groove into the nitride layer, and finally stripping the Cr. The line-edge roughness is a result of the Cr grains and is not



(a) (b) (c)

Figure 5.7 Demonstration of spatial period division using multiple exposures and nonlinearities. a) 50-nm wide, 360-nm pitch grooves etched into nitride mask layer; b) 180-nm pitch nitride grooves after interpolating second 360-nm pitch grating; c) final ~ 180-nm pitch grating etched into Si.

inherent in the lithography process. Figure 5.7(b) shows the nitride film after the second pattern transfer, the period has been halved from 380 nm to 180 nm. Finally the bottom panel, Fig. 5.7(c), shows the result of a pattern transfer into the <110> Si using anisotropic KOH etching and stripping of the sacrificial nitride layer. The alternations in both spacing between lines and linewidths are indications of imperfections in the *p*/4 shift and the CD of the second exposure, nonetheless, this result proves the concept.

5.5 Arbitrary patterns – Single-Exposure Limits of Optics

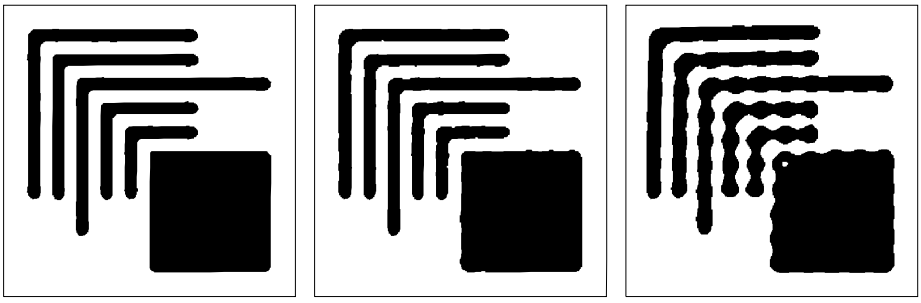
The discussion so far has been presented for IL, using periodic patterns whose Fourier content is restricted to a small number of points in spatial frequency space. Realistic integrated circuit patterns are much more complex, containing regions of isolated features, regions of dense patterns, shapes such as ell's and tee's, etc. The consequence in spatial frequency space is that there are a great many spatial frequencies in a typical pattern, theoretically as many as there are pixels in the image or ~$4x10^{10}$ for a 100-nm CD and a 2x2 cm^2 field. Realistically, there are fewer frequencies in typical integrated circuit patterns, which are more regular than a totally random Rorschach pattern, but still many more than could addressed by any conceivable summation of individual IL exposures.

Imaging addresses this issue by capturing a large number of spatial frequencies in a single exposure. The lens is a low-pass filter that restricts the spatial frequency components collected to a smaller range than that available from the transmission medium. The photoresist is a square law detector, responding to the intensity not the electric field, and adds additional nonlinearities as discussed above. For coherent imaging the low-pass constraint is NA/λ, for incoherent imaging the limit is 2NA/λ. Numerous techniques have been introduced to capture

more of the spatial frequencies, and thus produce a better image at higher resolution; these are discussed in Sections 5.5 and 5.6 for conventional imaging and resolution enhancement techniques, respectively. In this section, the ultimate limits of optics are presented, i. e., the arbitrary image corollary to the λ/4 resolution limit of IL for dense patterns.

Figure 5.8 shows an example of a test pattern, consisting of a series of nested "ells" with isolated lines along with a large, low-frequency box that is often used to evaluate lithographic system limits. The calculation was carried out simply by Fourier transforming the pattern, applying a low-pass filter at the optical limit of 2/λ, transforming back into real space, and applying a sigmoid nonlinearity to mimic the resist response. The main impact of this processing for the largest pattern, CD = λ/3, is a slight rounding of the corners of the "ell" patterns. For the CD = λ/3.5 pattern, the rounding is more pronounced, the corners of the "ell's" are less well defined, and some edge roughness is apparent. For the CD = λ/4 pattern, the edge roughness is pronounced and a hole has appeared in the large box

The reason for this progressive deterioration of the image is evident from Fig. 5.9 which shows the amplitudes of the Fourier coefficients for this pattern, scaled to the optical frequency space for the CD = λ/3 case. There is spectral intensity at low frequency, corresponding to the overall shape of the test structure, and peaks in the f_x- and f_y-directions corresponding to the repetitive pattern of the nested 'ells.' For this Manhattan geometry integrated circuit pattern, which is laid out on a rectangular grid, the majority of the spectral intensity is in the vertical and horizontal directions. The spectral content in each of these directions is not a simple δ-function as it would be for a perfectly repetitive structure, but rather shows some breadth associated with the information that there are a limited number of repeating lines with a limited extent. The spatial frequency scale of this structure is on the order of 1/(nCDL), where n is the number of nested lines in the radial direction and L is the length of the lines in the perpendicular direction. For CD = λ/3, all of this information is contained within the bandpass limits of op-



CD = λ/3 CD = λ/3.4 CD=λ/4

Figure 5.8 Calculated real space images for a dense (half-pitch) lithography test pattern with equal lines and spaces for three values of the CD.

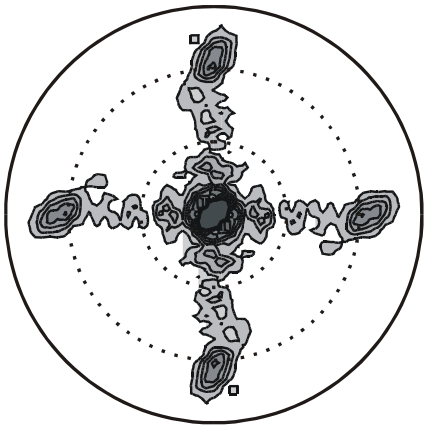


Figure 5.9 Spatial frequency space plot of the Fourier coefficient amplitudes for the pattern of Fig. 5.8, scaled to the optical wavelength for the λ/3 case. The solid outer circle is the optical limit of 2/λ the inner dotted circles are at 2NA/λ and NA/λ, respectively for a NA of 0.7.

tics, and all of it is retained in the final image. It is important to note that the aerial image for this case is sinusoidal, only the fundamental frequency defining the period of the nested 'ells' is transmitted by the optics. The nonlinearity of the resist, as described above, restores the higher harmonics and allows a developed. photoresist pattern with vertical sidewalls. For smaller CDs, the Fourier space plot expands accordingly and the higher frequencies approach the optical cutoff. For CD = λ/4, the center frequency of each of the high frequency spectral areas is just at the cutoff, and the higher frequencies are not included in the image. The loss of these high frequency terms is reflected in the deterioration of the image.

From this discussion, the ultimate *single-exposure* resolution of optics is limited to a pattern dependent dense (half-pitch) CD of ~ λ/3. The techniques to approach this limit are discussed below. Given the enormous effort being applied and the strong record of advances made to date, it seems reasonable to postulate that this limit will be closely approached as the field progresses. To provide some definiteness, this limit is a dense CD of 64 nm for a wavelength of 193 nm (ArF) and of 52 nm for a wavelength of 157 nm (F_2). Further reductions are possible with the use of immersion and of multiple exposures with nonlinear mixing, covering essentially all of the range of feature sizes demanded by the present roadmap.

5.6 Conventional Imaging

The illumination scheme has a strong impact on conventional imaging.[5] The limiting cases are coherent and incoherent illumination. In coherent illumination, a single plane wave (or more generally a Gaussian beam with a radius of curvature

matched to the optical system) is incident on the mask, most often at normal incidence. The optical system then acts as a low pass filter transmitting all of the Fourier components scattered by the mask up to a frequency of NA/λ at unity amplitude, and completely eliminating all higher frequencies. The square of the sum of these fields then describes the aerial image at the wafer plane. Finally the resist nonlinearity again sharpens the profile.

For incoherent illumination, a large number of incoherently related waves illuminate the mask with a uniform distribution of incident angles, the field-squared image is obtained for each of these plane waves and the photoresist responds to the sum of these images. Higher frequencies are included in the image, up to 2NA/λ, as a result of the off-axis illumination. However, the amplitude of these high frequency terms is reduced since they are present in only a small subset of the images. This gives rise to the well-known modulation transfer function, MTF, of an optical system that is ~ 0.5 at NA/λ and goes to zero at 2NA/λ. Because each of the individual images must be positive everywhere, there is a higher baseline exposure over the entire image. Figure 10 shows the resulting images for κ_1= 0.6 at an NA of 0.7.

The coherent image contains all frequencies up to NA/λ with the same intensity ratios as in the Fourier transform of the mask as a result of the interference of the zero-order transmission of the mask with the scattered fields transmitted through the lens. In addition, the square-law intensity detection results in additional frequency components ranging from zero to 2NA/λ. These components arise from the autocorrelation of the imaging fields and do not necessarily correspond to image spatial frequencies. An advantage of coherent imaging is that all of the Fourier components interfere coherently, the requirement is only that, at any point on the image, the total intensity must be positive. Coherent imaging tends therefore to have the lowest baselines, or unwanted intensity across the unexposed regions of the image. However, the hard cutoff of the lens tends to result in ringing of the intensity leading to line-edge roughness, as can be seen in Fig. 5.10. For incoherent illumination, each image spatial frequency is multiplied by a modulation transfer function (MTF) that is unity at zero frequency, is ~0.4 at

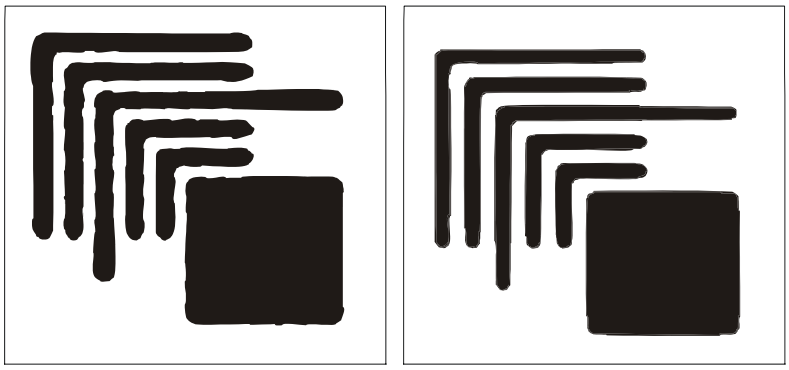


Figure 5.10 Coherent (left) and incoherent images at κ_1 = 0.6 and NA = 0.7 (160 nm at λ = 193 nm).

modulation transfer function (MTF) that is unity at zero frequency, is ~0.4 at NA/λ and goes to zero at 2NA/λ. The net result is that incoherent illumination contains higher image frequencies, but with a lower transfer function at mid-spatial frequencies than coherent illumination. The ringing associated with the frequency cutoff of coherent illumination is eliminated by the gradual reduction of the frequency response for incoherent illumination. Partial coherence provides a different balance of these frequency responses. In any case, conventional imaging does not provide any spatial frequencies above 2NA/λ in the aerial image. For the cases of coherent and partially incoherent illumination, it is not possible to simply define an image intensity MTF, rather the result depends on the details of the image because of the autocorrelation contributions discussed above. More detailed discussion of these issues is provided in Refs. 1, 4, and 5. For simplicity, all of the results presented in this paper were calculated with a simple, scalar imaging theory. Particularly at high NAs it will be important to extend these results to a full vector imaging code.

5.7 Resolution Enhancement Techniques to Reduce κ_1.

A number of techniques have been introduced to reduce κ_1 below the Rayleigh limit of ~ 0.6. These are optical proximity correction (OPC), off-axis illumination (OAI), phase-shift masks (PSM), and imaging interferometric lithography (IIL). The basic idea of each of these techniques is to provide a higher frequency content in the final image. A more complete discussion of the relationship between these techniques, and of their Fourier optic interpretation, has been presented elsewhere.[15,16] Only a brief overview is included here.

5.7.1 Optical Proximity Correction (OPC)

With OPC, sub-resolution features (corresponding to high spatial frequencies) are added to the mask, significantly increasing its complexity, to boost the magnitudes of the high spatial frequencies in the Fourier transform beyond those of the image. For example these can be sub-resolution lines that do not print but make a isolated line appear more like a dense array of lines to address dense-isolated feature biases. These magnitudes are attenuated by the optical system but are still larger than they would have been without the OPC, resulting in an image closer to the desired pattern, but no longer simply related to the mask pattern. This approach is appropriate for partially coherent illumination. For coherent illumination, the electric field transfer function is already unity up to NA/λ and there is nothing to be gained. It is intuitively clear that this approach can deal with frequencies near NA/λ where the incoherent MTF is ~ 0.5 but is inadequate to add substantial frequency content out towards 2NA/λ where the MTF is very low. Because of these restrictions in frequency space, OPC extends κ_1 to ~ 0.5 but not below. As a result of diffraction, these changes are nonlocal; e.g. the features impact an area of ~ λ^2 that may encompass more than the single desired feature. As

a result, OPC modeling is complex and numerically intensive. The pattern is no longer the "what you see is what you get" paradigm that was sufficient for lower resolutions. The complexity of OPC masks can be as much as 10x greater (more features) than conventional masks, increasing the burden and cost of fabricating the masks.

5.7.2 Off-Axis Illumination (OAI)

In OAI, the (coherent or partially coherent) illumination beam is tilted away from the optical system normal to allow higher spatial frequencies to pass through the optical system. These tilts are restricted to a fraction $\eta < 1$ of NA/λ in order to transmit the zero-order beam through the optical system and the maximum spatial frequency is given by $(1+\eta)NA/\lambda < 2NA/\lambda$. Both annular and quadrupole illumination have been explored. As is the case with any partial coherence, the baseline is always higher than for coherent illumination since each image can only have positive intensities. Figure 5.11 shows the basic concept of OAI for illumination of a simple grating. The grating pitch is greater than NA/λ, but less than $(1+\eta)NA/\lambda$. At normal incidence both first order diffracted beams are outside of the lens pupil, but for off-axis illumination, one of the diffracted orders as well as the zero-order transmitted beam are collected by the lens and an image of the grating is formed by their interference in the photoresist.

An issue with OAI is that the multiple exposures result in placing greater weight on the lower frequencies as illustrated in Fig. 5.12 which shows the frequency space coverage for a quadrupole illumination with an offset of $\eta = 0.75$. The axis of the quadrupole is usually shifted 45° relative to the Manhattan geometry axes since this gives a higher response to the dominant frequency components. The numbers and the shading indicate the relative weighting of the spatial frequencies within each area. These spectral overlaps restrict κ_1 to ~ 0.43. Pupil plane filters provide an additional degree of freedom for OAI.[15-18] If these overlaps are eliminated with pupil plane filters, OAI can be extended to $\kappa_1 \sim 0.3$.

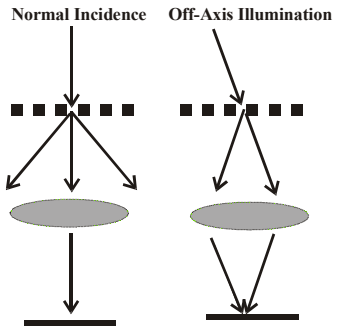


Figure 5.11 Illustration of the higher spatial frequencies possible with off axis illumination.

5.7.3 Phase-Shift Masks (PSM)

For PSM, the mask is drastically modified (to a 3D object) to enhance the amplitudes of the fundamental frequency components (out to NA/λ) while reducing the amplitude of the dc-component of the electric-field. Because of the square-law resist response, this dramatically boosts the quadratic image intensity components between NA/λ and 2NA/λ allowing κ_1's of ~ 0.35. As was the case for OPC, the simple relationship between the mask and image pattern is lost requiring extensive modeling, design and mask fabrication efforts. Figure 5.13 shows the general concept of phase shift for a chromeless, strong phase-shift case. The lefthand column shows a conventional coherent imaging of a grating of period *p* that is not captured by the imaging system. The right column shows a chromeless phase-shift mask with a period of 2*p*, the depth of the etching is (*n*-1)λ/2 so that there is a λ/2 phase shift for the transmitted fields. The result is that the zero-order transmission is identically zero and the diffracted orders interfere as in IL to produce a grating at *p*. The maximum frequency that can be printed is 2NA/λ. An issue with PSM is that the mask is transformed dramatically from the conventional imaging paradigm requiring significant modeling efforts. Strong phase

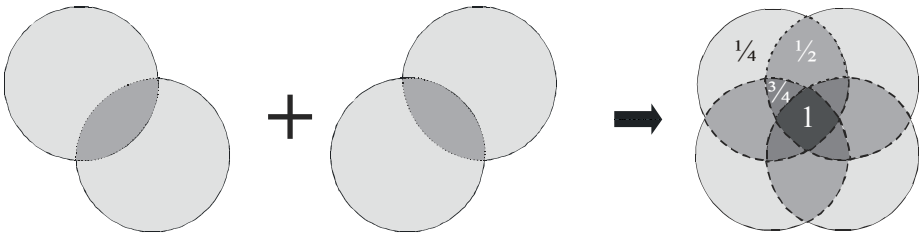


Figure 5.12 Frequency space coverage for quadrupole OAI with η = 0.75. The overlaps give greater weight to the low frequency components.

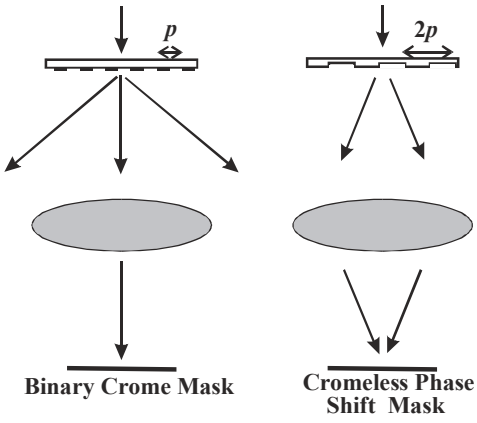


Figure 5.13 Principle of phase shift masks. The maximum spatial frequency is 2NA/λ.

shift masks produce extra lines that need to be compensated for with a second mask and a second pass through the lithography tool.

5.7.4 Imaging Interferometric Lithography (IIL)

IIL is a newly introduced[19,20] imaging concept that is closely related to OAI, with the major difference that, for relatively low NA optics, the offsets are chosen beyond NA/λ requiring a separate optical system to arrange for the zero-order reference beam at the wafer. For continuous coverage along the principal (f_x, f_y) frequency axes, the maximum frequency at the wafer is as high as 3NA/λ allowing κ_1's of as low as 0.2. With the advent of very high NA lenses, IIL is essentially OAI with the addition of pupil plane filters to eliminate the multiple coverage of frequency space that favors the low frequencies as discussed above. As with OAI, the mask and image patterns are identical (except for magnification scaling) and traditional binary chrome-on-glass masks are used.

Significant benefits of IIL, especially for high-NA optics, include the ability to optimize the polarization for the highest spatial frequencies, and to vary the relative intensities of the multiple exposures to optimize the image contrast. In addition, the ability to vary the pupil aperture for different exposures gives some control over the unwanted autocorrelation (dark-field imaging) contributions to the image. Figure 5.14 shows the limiting results for an NA of 0.9 and a CD of λ/3 (κ_1 = 0.3) for incoherent illumination, coherent illumination, and imaging interferometric lithography. The top panels show the calculated image results for the test pattern, the bottom panels show the corresponding frequency space coverages. Incoherent illumination resolves the individual nested 'ells' but clearly suffers from a serious loss of linewidth and foreshortening of the isolated lines. The outermost 'ell' is also significantly narrower than the others. The frequency space coverage is indicated by the shaded circle with an NA of 0.9. For coherent illumination, the pattern is totally unresolved. The spatial frequency corresponding to the line/space pattern is 1.5/λ > NA/λ and is not imaged. Finally, IIL with pupil plane filters does a very good job of imaging the pattern. The set of filters used for the calculation is shown in the frequency space diagram. The central exposure was at an NA of 0.45 while the offset NAs were at 0.9 with η = 1 and with additional filters to eliminate overlaps. In this case, all of the beams are transmitted through the lens pupil, there is no need for the interferometers around the outside of the lens.

5.7.5 Combinations of RETs

Application of these techniques is not mutually exclusive. Clearly, OPC will be combined with the other techniques to address remaining image-specific issues. OAI and IIL are closely related. In the context of lower NA lenses, IIL is OAI with a sufficiently large offset that the zero-order beam is not captured by the lens but rather is reintroduced after the lens. With the advent of very high NA

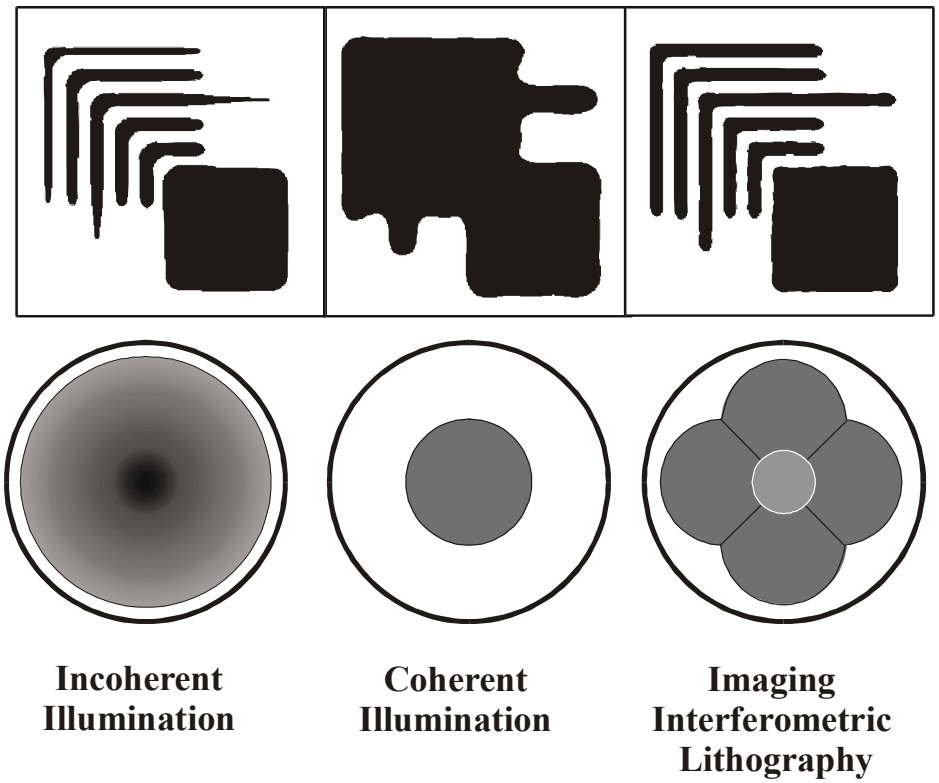


Figure 5.14 Comparison of incoherent illumination, coherent illumination and imaging interferometric lithography for a CD of λ/3 and an NA of 0.9 ($\kappa_1 = 0.3$).

lens, IIL effectively reduces to OAI with the use of pupil plane filters to eliminate overlaps in the spatial frequency coverage between different exposures and assure a uniform electric field transfer function across the entire frequency space. All of the beams now pass through the lens so the need for interferometric length control is reduced to that for the imaging system.

5.8 Nonlinear Processes Applied to Arbitrary Patterns

The nonlinear concepts introduced in Sect. 5.2 apply equally well to the case of arbitrary patterns. The general concept is to print as much as diffraction allows in a single exposure (Sect. 5.7), transfer the information to a temporary storage layer (often a sacrificial layer that will be removed later), and interpolate subsequent exposures that are also transferred into the storage layer. Finally use the storage layer combining the information from all of the exposures to transfer the complete pattern into the underlying layers. An example of this process for the test pattern introduced above for a dense CD of 32 nm (λ/6) with an exposure wavelength of 193 nm is shown in Fig. 5.15. As discussed in Sect. 5.2, it is criti-

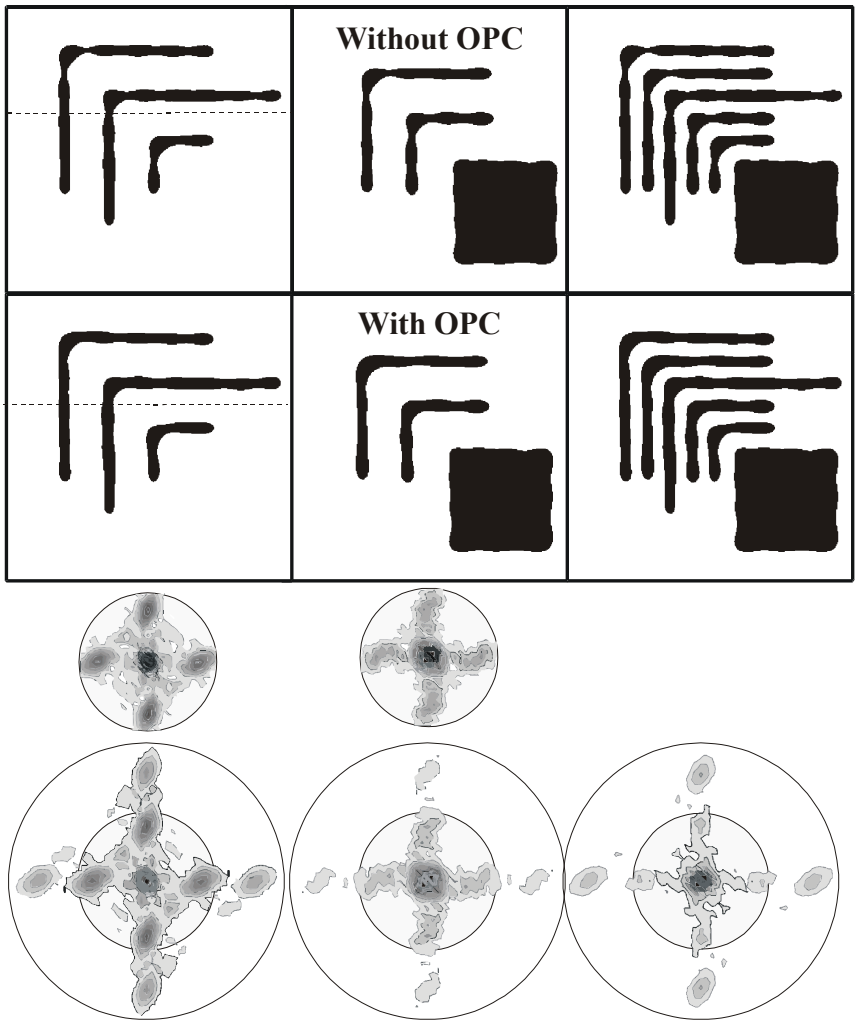


Figure 5.15 Model calculation for interpolating two images, each near the single-exposure limit of optics to provide a final pattern at twice the usual optics limits. The calculation is for a 32-nm dense CD at a wavelength of 193 nm. The top panels show the results of the two exposures and their combination for a simple mask pattern. The second row of panels shows similar results with a small amount of OPC to counteract some of the edge ripple in the top images. The third row shows the Fourier content of the aerial image, bandpass limited at $2/\lambda$, and the bottom panel shows the Fourier content of the image out to $4/\lambda$ after the resist nonlinearity is applied.

cal to apply the nonlinearities separately to each exposure. If the multiple exposures are accumulated in a single resist layer, the patterns merge and enhanced resolution is not realized.

The calculation for Fig. 5.15 was based on the full transmission bandpass of optics and is similar to the calculations of Fig. 5.8. The top panels show the re-

sults of the two exposures and their combination using a straightforward mask without any attempt at OPC. Line edge variations are clearly evident for this level of exposure. The left-most panels show the result of the first exposure. Alternate lines of the nested "ell" are printed, so that the pitch is the same as in each of the single exposure examples, but the exposure dose is adjusted (increased) to get the required CD. This is similar to industry practice for logic chips that have a lower density than memory devices. The left-hand panel in the third row shows the aerial image frequency content, limited by the $2/\lambda$ by the transmission limit of optics. The bottom left-hand panel shows shows the Fourier content of the final image after the resist nonlinearity has been imposd. There is now evident frequency content at both the fundamental (passed by the optics) and the second harmonic (beyond the optics cutoff). There are higher frequency components as well that are not shown in the figure. The inner shaded circle denotes the $2/\lambda$ optics frequency limit.

The middle panels show similar results for the second exposure in which two of the "ells" and the large box are printed. Because of the low frequency content of the large box, the lower frequencies are emphasized more in the aerial image. Finally, the right-most panels show the result of combining the two images in a sacrificial layer. The final image is the desired dense CD image. The fundamental frequency components partially cancell, resulting in a relative emphasis of the second harmonic frequency content. It is instructive to compare the frequency content of Fig. 5.15 with that of Fig. 5.9.

As the linewidth is decreased, e.g. as the resist threshold is shifted to lower intensities, the variations in exposure at different positions on different lines--the linewidth ripple--becomes more severe. This is evident by comparing Fig. 5.15 for CD = $\lambda/6$, as compared with Fig. 5.8 for CD = $\lambda/3$. These linewidth variations

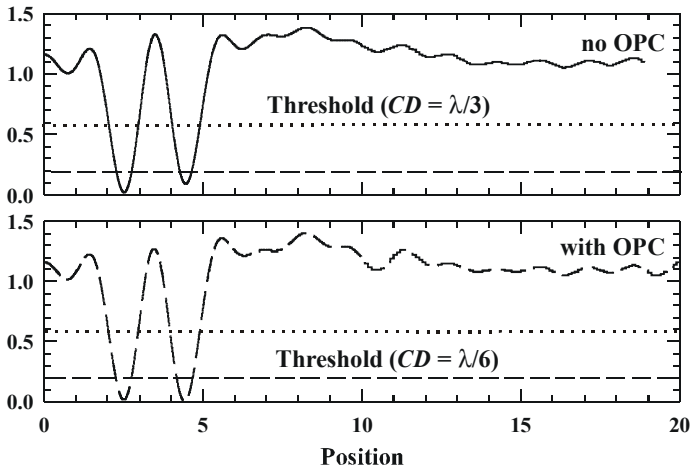


Figure 5.16 Aerial image optical intensity along the cuts indicated by the dotted lines in Fig. 5.15. The bottom trace is in the absence of any OPC correction. The top trace is with the OPC illustrated in Fig. 5.17. The resist threshold intensity (dashed lines) is also shown for imaging at $\lambda/3$ and $\lambda/6$.

are related to intrinsic variations in the aerial image rather than to any process variability in this idealized calculation. This result also implies that process control issues will be more significant at these high exposure levels as the linewidth variation for a small change in intensity becomes more severe.

This is illustrated more clearly in Fig. 5.16 which shows the aerial image intensity profiles along the cuts indicated by dotted lines in the left-hand images of Fig. 5.15, passing through two vertical bars of the "ells." Because the amplitudes of the two lines are not precisely equal, there is some narrowing of the second line in the final image. This can be addressed by widening of the mask pattern at this point as shown in the OPC mask (Fig. 5.17). The frequency of the intensity oscillation is set by the bandpass limits of optics and is the same for both lines, so the intensity (depth of the null) is the only possible adjustment; each line is the same width. If the threshold is moved up to the mid-intensity levels, as it was for the results of Fig. 5.8 (e. g. for a *CD* of $\lambda/3$) and as shown by the dash-dot lines labeled "threshold (CD ~ $\lambda/3$)" the impact of these variations on the final printed image is much less pronounced. For this calculation, the definition of this OPC was an ad hoc process of adding and subtracting bars in the areas that showed the most ripple. An extensive investigation was not carried out and better results are certainly available; these results are sufficient to define the concept. The message is that mask correction, mask *CD* control, and process

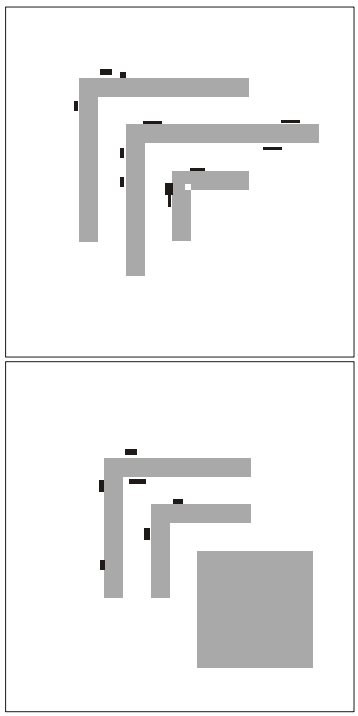


Figure 5.17 OPC additions (black)/subtractions (white) to the simple mask (shown in *grey)*

is that mask correction, mask *CD* control, and process control all will become more demanding as increasing orders of nonlinear interpolations are attempted.

The division of the pattern into two (or more) exposures is essentially a map coloring problem, where the rule is never have two objects with the same color closer than the Rayleigh resolution limit for a single exposure. Another example is shown in Fig. 5.18 for a typical SRAM pattern level. In principle, the technique can be extended to multiple exposures, and therefore to CDs much smaller than the Rayleigh criterion. In practice, nonuniformities and process control issues will be more significant the larger the number of exposures, and hence the thinner the lines required in each exposure in comparison with the limiting pitch of ~ $\lambda/3$ ($\lambda/4$) for arbitrary (periodic) patterns.

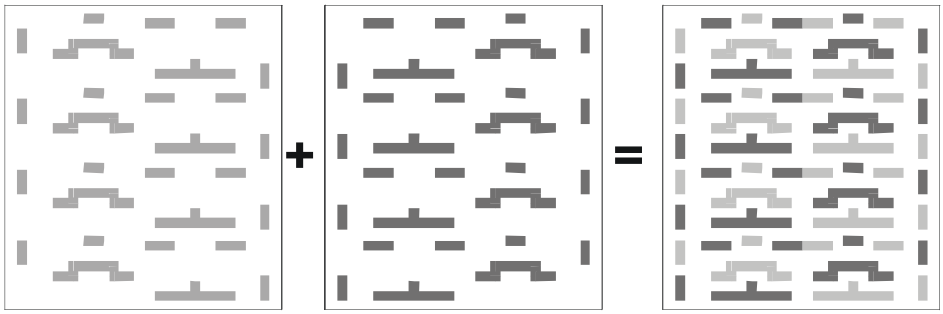


Figure 5.18 Example of interpolation of two exposures to increase the final pattern density for an arbitrary pattern (a typical SRAM level).

5.9 Summary and Conclusions.

The fundamental limits of optical imaging, assuming aberration-free lenses have been discussed. For a conventional illumination single exposure, diffraction limits the ability to resolve two objects to a κ_1 of ~0.6 giving a dense (half-pitch) CD of 0.9λ for an NA of 0.7. However, in the absence of an optical system, optics can support much higher spatial frequencies of up to $2/\lambda$ providing dense (half-pitch) lines and spaces at a CD of $\lambda/4$ for a fully periodic pattern. For maskless, interferometric lithography this limit is support by experimental results. For non-repetitive patterns the smallest CD supported by optics has to be relaxed somewhat to ~ $\lambda/3$ (pattern dependent) to allow access to all of the necessary spatial frequency information. Numerous techniques have been introduced to extend the results of practical imaging systems to these limits. Both phase shift masks and imaging interferometric lithography with very high NA lenses are promising alternatives that extend the spatial frequency coverage of conventional imaging and closely approach this single-exposure optics resolution limit.

All of these results scale as the refractive index of the transmission medium. This allows an additional reduction of linewidth of ~ 1.5, but at the expense of a much more complex optical system.

Imaging, exposure, development, and pattern transfer are all highly nonlinear processes. The optics limit is on the closest resolvable distance between two objects, not on the linewidth of an individual object.

Table 5.2 shows the limits of optical lithography, assuming a nominal 200-nm photon. ArF lasers at 193 nm are coming into widespread use in the silicon manufacturing industry. Coherent sources at 213 nm (fifth harmonic of YAG) are commercially available. The photoresist infrastructure at these wavelengths is already largely in place. There are ongoing efforts to further reduce the linewidth. All of these results scale accordingly. As mentioned in the introduction, EUV is an electromagnetic imaging technology and the same arguments apply, with the difference that very high-NA lenses are precluded by geometric considerations. Table 5.2a presents the single exposure limits for both interferometric (periodic patterns) and optical imaging (arbitrary patterns) lithographics. Table 5.2b shows the corresponding limits with a single interpolation step. Larger numbers of interpolation steps are possible at the price of increased complexity and reduced process latitude.

Nonlinearities typically introduce higher harmonics of the spatial frequencies of the image. Combining images after applying these nonlinearities independently to each image is shown to provide sufficient flexibility to extend optical resolutions; a factor of two seems readily accessible. In principle higher multipliers are possible, but process control and alignment issues will become increasingly difficult. Combining nonlinearities requires transfer into a sacrificial layer or multiple layers of resist. Multiple exposures in the same level of resist are not able to achieve the higher resolution.

It is useful to compare these tables with Table 5.1 that sets out the ITRS lithography requirements. The ultimate decisions as to lithographic technique will be made on economic grounds. While optics can be extended well beyond today's requirements based on the present analysis, there are many other issues that have not been discussed in this chapter. These have to do with depth-of-field, and in general with process latitude issues. Impacts of imperfection in the mask, the optical system, the focal plane, etc. have been ignored and deserve detailed analysis.

Table 5.2a: Single Exposure Limits of Optical Lithography (λ = 200 nm and n = 1.5)

Technique	Half-Pitch CD	Pitch
Interferometric lithography in air (or vacuum)	$\lambda/4$ (50 nm)	$\lambda/2$ (100 nm)
Immersion interferometric lithography (n = 1.5)	$\lambda/4n$ (33 nm)	$\lambda/2n$ (66nm)

Nonrepetitive patterns (imaging) in air (or vacuum)	$\lambda/3$ (67 nm)	$\lambda/1.5$ (133 nm)
Single-exposure imaging with immersion	$\lambda/3n$ (44 nm)	$\lambda/1.5n$ (88 nm)

Table 5.2b Double Exposure Limits of Optical Lithography (λ = 200 nm and n = 1.5)

Technique	Half-Pitch CD	Pitch
Multiple exposure IL in air*	$\lambda/8$ (25 nm)	$\lambda/4$ (50 nm)
Multiple exposure IL with immersion*	$\lambda/8n$ (17 nm)	$\lambda/6$ (34 nm)
Multiple exposure imaging in air*	$\lambda/6$ (33 nm)	$\lambda/3$ (67 nm)
Multiple exposure imaging with immersion*	$\lambda/6n$ (22 nm)	$\lambda/3n$ (44 nm)

*In principle, the frequency multiplication can be repeated multiple times, a single doubling of the spatial frequency in both the x- and y- directions is assumed, as illustrated in Figs. 5.15 and 5.16.

Nanotechnology is emerging as a major research field. One of the requirements for many nanotechnology directions is a facile and inexpensive patterning technique to reach the nanoscale. Parallel writing is an important advantage of optical lithography, macroscopic areas are printed in a single exposure with high uniformity and placement accuracy. Based on the analysis and results presented herein, optical lithography, and especially interferometric lithography, will play a significant role in the development of this new field.

Acknowledgements

Funding for this work was provided by the Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency and the Army Research Office, as part of the Multidisciplinary University Research Initiative, and by International SEMATECH.

Many students and colleagues have contributed to the development of these ideas, special thanks are due to Saleem H. Zaidi, Xiaolan Chen, Janusz Wilczynski (WILC Instruments), and Scott Mackay (SEMATECH). I have gained much from working with them. The 108-nm pitch structure was produced by Alex Raub. Frank Schellenberg provided useful comments on the manuscript.

References

1 H. J. Levenson, Principles of Lithography, SPIE Press, vol. PM97, Bellingham, WA, 2001.
2 ITRS available at http://public.itrs.net
3 The web site for information on the NNI is http://www.nano.gov.
4 Max Born and Emil Wolf, *Principles of Optics*, 6th ed., Cambridge University Press, Cambridge (1997)
5 J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed., McGraw-Hill, New York (1996).

6 L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll, "Generation of Periodic Surface Corrugations," *Appl. Opt.,* vol. 17, p. 1165 (1978).
7 E. H. Anderson, C. M. Horowitz, and H. I. Smith, "Holographic Lithography with Thick Photoresist," *Appl. Phys. Lett*., vol. 43, pp. 874-875 (1993).
8 S. H. Zaidi and S. R. J. Brueck, "Multiple-Exposure Interferometric Lithography," *Jour. Vac. Sci. Technol*., vol. B11, pp. 658-666 (1993).
9 X. Chen, Z. Zhang, S. R. J. Brueck, R. A. Carpio and J. S. Petersen, "Process Development for 180-nm Structures using Interferometric Lithography and I-Line Photoresist*"Proc. SPIE,* vol. 3048, pp. 309-318 (1997).
10 J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, "Liquid Immersion Deep-Ultraviolet Interferometric Lithography," *J. Vac. Sci. Technol.,* vol. B17, pp. 3306-3309 (1999).
11 M. Switkes and M. Rothschild, "Immersion Lithography at 157 nm," *J. Vac. Sci. Technol.,* vol. B19, pp. 2353-2356 (2001).
12 R. Zeiger and C. A. Mack, "Generalized Approach Toward Modeling Resist Performance," *AIChE Jour.,* vol. 37, pp. 1863-1874 (1991).
13 G. M. Wallraff and W. D. Hinsberg, "Lithographic Imaging Techniques for the Formation of Nanoscopic Features," *Chem. Rev.,* vol. 99, pp. 1801-1821 (1999).
14 S. H. Zaidi and S. R. J. Brueck, "Nonlinear Processes to Extend Interferometric Lithography," *Proc. SPIE,* vol. 3676, pp. 371-378 (1999).
15 S. R. J. Brueck and X. Chen, "Spatial Frequency Analysis of Optical Lithography Resolution Enhancement Techniques*," Jour. Vac. Sci. Technol*., vol. B17, pp. 908-921 (1999).
16 X. Chen and S. R. J. Brueck, "Experimental Comparison of Off-Axis Illumination and Imaging Interferometric Lithography," *Jour. Vac. Sci. Technol.,* vol. B17, pp. 921-929 (1999).
17 H. Fukuda, T. Terasawa and S. Okazaki, "Spatial Filtering for Depth of Focus and Resolution Enhancement in Optical Lithography," *J. Vac. Sci. Technol.,* vol. B9, pp. 3113-3116 (1991).
18 S. Orii, T. Sekino, and M. Ohta, "Quarter Micron Lithography System with Oblique Illumination and Pupil Filter," *Proc. SPIE,* vol. 2197, pp. 854-868 (1994).
19 X. Chen and S. R. J. Brueck, "Imaging interferometric lithography—A wavelength division multiplex approach to extending optics." *J. Vac. Sci. Technol*., vol. B16, pp. 3392-3397 (1998).
20 X. Chen and S. R. J. Brueck, "Imaging interferometric lithography—approaching the resolution limits of optics," *Opt. Lett*., vol. 24, pp. 124-126 (1999).

Twofold spatial resolution enhancement by two-photon exposure of photographic film

Dmitriy V. Korobkin, Eli Yablonovitch

Publication: Optical Engineering 41(07), p. 1729-1732, Donald C. O'Shea; Ed.

Publication Date: Jul 2002

Abstract
Two-photon absorption of photosensitive media can produce interference fringes with double spatial frequency. This requires the employment of multiple-frequency beams, which interfere with one another to produce a stationary image with double spatial resolution. The required beams were produced by frequency filtering of broadband radiation from a cw mode-locked femtosecond Ti:sapphire laser (lambda = 790 nm) in a dispersion-free pulse shaper. Then the two multifrequency rays converged from opposite edges of a lens, focusing on Kodak commercial film. The laser intensity was high enough to produce a two-photon exposure. The doubling of the spatial frequency of the interference pattern has been observed, but the contrast ratio of the pattern was limited by competition from the more usual one-photon absorption. Laser pulse parameters for a single-pulse two-photon exposure have been estimated. (Copyright)2002 Society of Photo-Optical Instrumentation Engineers.

©2005 SPIE--The International Society for Optical Engineering.

Deep UV immersion interferometric lithography

Alex K. Raub & S.R.J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106

ABSTRACT

The limit of optical lithography is the minimum pitch between features. This pitch limit is given by $\Lambda = \lambda/2nNA$, where λ is the optical wavelength, n is the refractive index of the final medium of the optical system which is typically air (n = 1), and NA is the numerical aperture of the exposure device. A great deal of work has been done to decrease exposure wavelengths and increase the NA of exposure tools, however, until recently very little effort has been applied towards an immersion medium with $n > 1$. This paper examines extending minimum pitches through the use of such media. Exposures are at a wavelength of 213 nm, close to the current state-of-the-art 193-nm lithography node. The possible limits of lithography are examined using 193-nm resists exposed in air and comparing these limits to those possible when implementing liquid immersion lithography (LIL) exposures. Two immersion liquids were examined: deionized water, and Krytox® a Perfluoropolyether (PFPE) oil. These liquids were compatible with 193-nm resist. A resolution enhancement factor of 28% for Krytox® and of 41% for DI water was demonstrated. Images of good dense lines with a half pitch of 54 nm are presented.

Keywords: liquid immersion lithography, interferometric lithography, DI H2O, 193 nm photoresist, Krytox, PFPE, sub 100-nm half-pitch

1. INTRODUCTION

As the microelectronic industry pushes towards smaller feature sizes, greater demands are placed on lithography. Modern lithography now commonly produces features smaller than the optical wavelength. Through the use of optical enhancement techniques such as off-axis illumination (OAI), optical-proximity correction (OPC), phase-shift masks (PSM) and imaging interferometric lithography (IIL), industry can extend the current limits of state of the art lithography tools. These techniques are quickly being pushed to their limits while next generation lithography technologies are still under development. Industry needs to find new ways to extend their lithography capabilities toward the theoretical limits. The ultimate minimum feature optical lithography can produce is a line and space whose pitch is determined by the well-known formula $\Lambda = \lambda/2nNA$, where λ is the wavelength of light, n is the refractive index of the exposure medium, and $NA = \sin(\theta)$ is the sine of the half angle of the light at the exit face of the lens.

Historically, the exposure wavelength λ has been steadily decreased and the numerical aperture NA has been steadily increased to achieve to higher resolution. Current leading edge exposure tools operate at a wavelength of 193 nm with $NA \sim 0.7$ with projections to $NA \sim 0.85$ in the near future. Currently there is no optical node beyond 157 nm but research is underway for even shorter usable wavelengths sources. There is still a limit on traditional refractive-optics-based lithography as there are no known transparent lens materials below ~ 105 nm.[1] Interest has recently grown dramatically in exploring the use of immersion media, between the exit face of the lens and the photoresist surface, to extend optical lithography.[2, 3,4,5] Currently air is the only exposure medium being used for commercial exposure tools, with n = 1. The use of immersion media to extend optical resolution is well known from optical microscopy. Through the use of other media, such as de-ionized water with a value of $n \sim 1.44$ at 193 nm, much smaller minimum feature sizes can be achieved. From the pitch equation this can be thought of either as an effective wavelength $\lambda_{eff} = \lambda/n \sim 134$ nm; or alternatively as an increase to an effective $NA_{eff} = n\sin\theta \leq 1.44$. Thus, the use of water immersion lithography could potentially extend 193-nm wavelength optical lithography by 44% or almost two generations, twice the resolution enhancement offered by the transition to 157 nm. Application of immersion techniques at 157 nm will have a similar impact; however a suitable liquid has yet to be identified.

In order to provide a perspective on semiconductor industry lithographic requirements, Table 1 reproduces material from the 2001 International Technology Roadmap for Semiconductor (ITRS).[6] A technology node is defined by ITRS as “the

half-pitch of the first-level interconnect dense lines is most representative of the DRAM technology level required for the smallest economical chip size. For logic such as microprocessors (MPUs), physical bottom gate length is most representative of the leading-edge technology level requires for maximum performance", "The smallest half-pitch is typically found in the memory cell area of the chip. Each technology node step represents the creation of significant technology progress - approximately 70% of the preceding note, 50% of two preceding nodes."[6]

Year Of Production	*2001*	*2002*	*2003*	*2004*	*2005*	*2006*	*2007*	*2010*	*2013*	*2016*
DRAM										
DRAM ½ Pitch (nm)	130	115	100	90	80	70	65	45	32	22
Contact in resist (nm)	165	140	130	110	100	90	80	55	40	30
Contact after etch (nm)	150	130	115	100	90	80	70	50	35	25
Overlay	46	40	35	32	28	25	23	18	13	9
CD control (3 sigma) (nm)	15.9	14.1	12.2	11.0	9.8	8.6	8.0	5.5	3.9	2.7
MPU										
MPU ½ Pitch (nm)	150	130	107	90	80	70	65	45	32	22
MPU gate in resist (nm)	90	70	65	53	45	40	35	25	18	13
MPU gate length after etch (nm)	65	53	45	37	32	28	25	18	13	9
Contact in resist (nm)	165	140	122	100	90	80	75	50	37	27
Contact after etch (nm)	150	130	107	90	80	70	65	45	32	22
Gate CD control (3 sigma) (nm)	5.3	4.3	3.7	3.0	2.6	2.4	2.0	1.5	1.1	0.7

White - Manufacturable Solution Exist, And Are Being Optimized
Light Gray - Manufacturable Solution are Known
Dark Gray - Manufacturable Solutions are NOT Known

Table 1 – Lithography technology requirements.[6]

This paper presents initial exposure results for immersion lithography at a wavelength of 213 nm, close to the current 193-nm exposure wavelength used in industry. A maskless interferometric lithography approach that gives essentially ideal contrast fringes was used.[7] Exposures were carried out at a number of angles corresponding to different *NA*'s. The smallest half-pitch achieved was 54 nm, limited by the image resolution of the resist systems investigated rather than by any optical effects; the optical half pitch limit ($\lambda/4n$) in these experiments was ~ 34 nm.

2. EXPERIMENTAL SETUP

The light source was a Coherent Infinity 40-100 Nd:YAG pulsed (~ 2 ns) laser operating at the fifth-harmonic wavelength of 213-nm. This source has much higher coherence length (~ 0.6 m) than a 193 nm excimer laser that simplifies the experimental arrangement for interferometric lithography. 193-nm resists work well at this wavelength. A simple diverging lens expands the ¼-inch beam to about 2.5-inches at the immersion cell 2-inch entrance window.

A Fresnel corner cube interferometer immersion cell was designed and constructed for this work. Figure 1 illustrates how the corner cube mirror folds half of the incoming UV beam onto the wafer to form a sinusoidal interference pattern on the surface of the wafer. The period of the grating exposed in the resist is given by $\Lambda = \frac{\lambda}{2 \cdot n \cdot \sin(\theta)}$ (n = 1 for exposure in air). Therefore rotating the corner cube interferometer with respect to the incoming UV beam sets the period. With this arrangement the two beams are always symmetrically incident on the wafer for any pitch. For immersion lithography the entire Fresnel corner cube mirror interferometer needs to be immersed in the liquid. To achieve this a cell is made to completely enclose a Fresnel corner cube. Advantages of this arrangement are rapid changes of pitch without requiring any realignment other than rotation of the corner cube, and direct measurements of the refractive index by exposing a wafer half immersed in liquid and half in the air space above the liquid. The ratio of the periods for these two gratings is a good measure of the liquid refractive index. A disadvantage is the relatively long liquid path length ~ 1 cm, which precludes the use of highly absorbing liquids. Fortunately, as is discussed below, both of the liquids explored in this work had sufficiently low absorption at 213 nm that this was not an issue.

A window on the cell allows for the external exposure source to be coupled into the cell. A rendering and an actual image of the immersion cell used in this work are shown in Figure 2 with a covering lid removed for clarity.

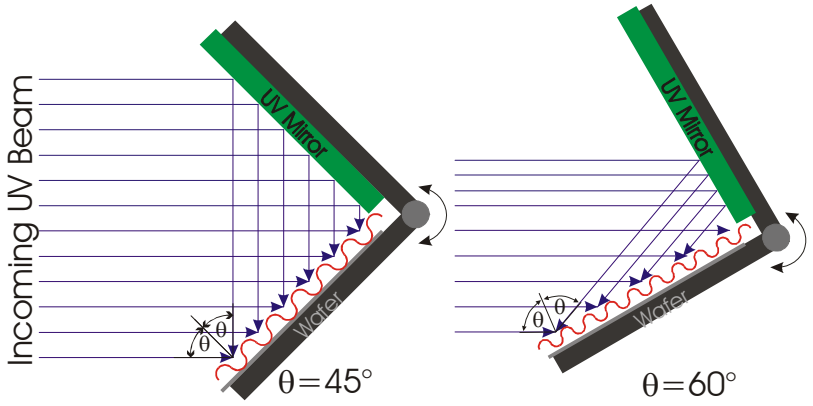


Figure 1 - Fresnel corner cube mirror interferometer diagram.

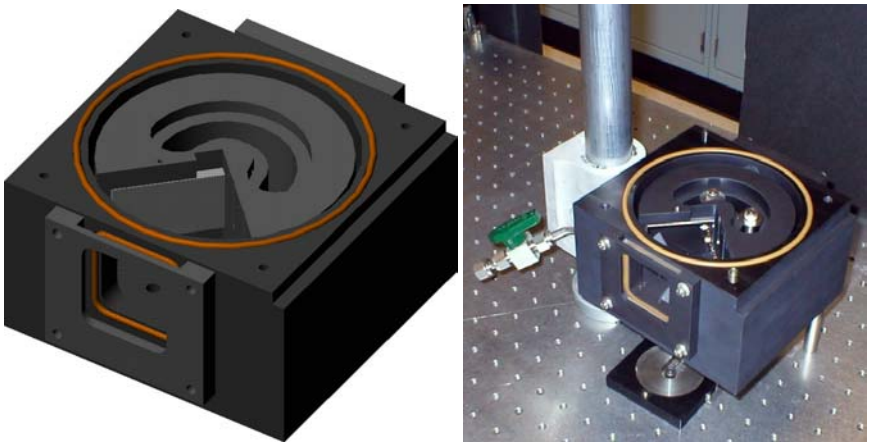


Figure 2 – Rendering and photo of immersion cell.

The only other optical components in the exposure setup are a pair of 213 nm 45° high power dielectric mirrors, used to steer the beam around the optical table. Figure 3 shows a diagram of the exposure setup.

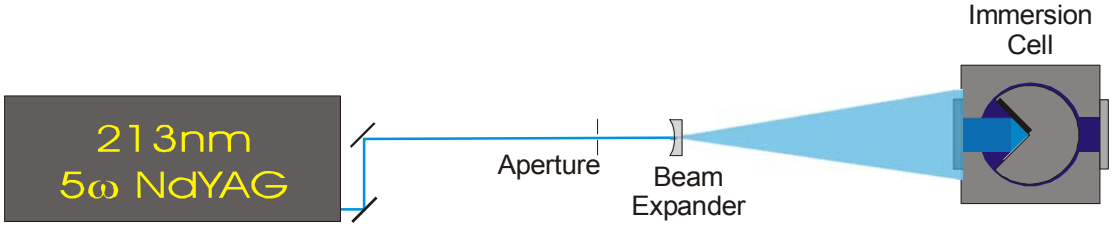


Figure 3 – Diagram of exposure setup.

3. EXPERIMENTAL RESULTS

3.1 Immersion medium index and absorption measurements

Measurements were made on two immersion liquids, deionized (DI) water and Krytox® a perfluoropolyether (PFPE) oil, to determine the absorption coefficient and index of refraction at the 213-nm exposure wavelength. Figure 4 shows the absorption measurements for both DI water and Krytox®. The absorption coefficient for DI water is $k = 2.28 \cdot 10^{-2}$ cm^{-1} (base e) at 213 nm; the absorption coefficient for Krytox® GPL100 is $k = 5.45 \cdot 10^{-2}$ cm^{-1} (base e) at 213 nm. For both liquids the 1/e lengths are much greater than the immersion distance for the imaging experiments.

In order to measure the index of refraction, it is necessary to measure the pitches of the gratings exposed by both the liquid immersed part and the air exposed part on the same resist sample. The pitch was measured by using the ¼ inch diameter 213-nm beam at low power on a developed resist sample mounted on a stepper rotation stage. The pitch is simply obtained by measuring the angle θ between the 0-order reflected beam and the 1st-order diffracted beam. Because the exposure angle determined by the Fresnel corner cube interferometer was identical for both gratings, the difference between the two pitches is due to the ratio of their respective refractive indices. This measurement was repeated at a number of exposure pitches; the number reported here is the average of about 5 measurements for each liquid. The measured index of refraction for DI water is $n = 1.42$, for Krytox it is $n = 1.29$.

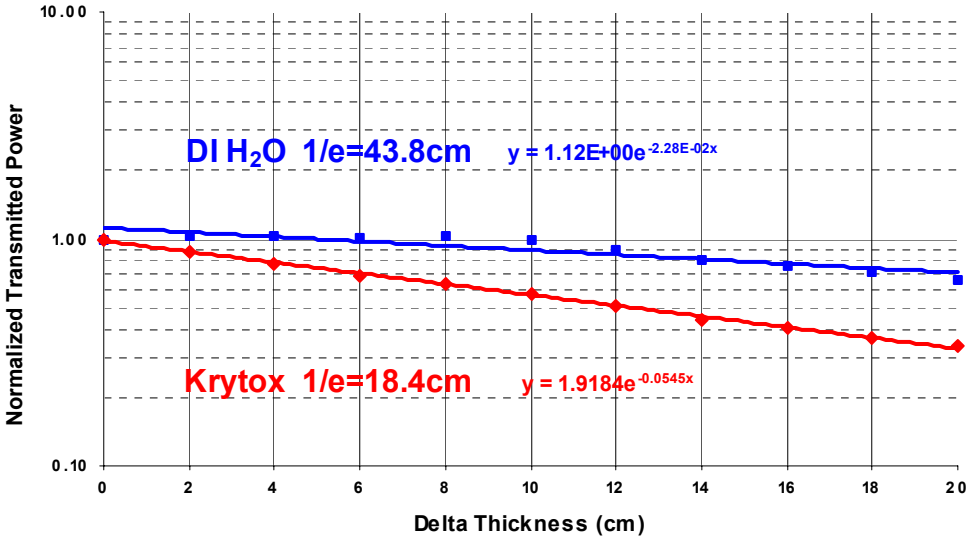


Figure 4 – Measured absorption curves for DI H2O & Krytox®

3.2 Exposures in Air

This work used two 193-nm resists, both represent commercially available resist for the year 2000. Each was provided by a major resist manufacturer. The first sets of resist samples were completely exposed in air. It was necessary to characterize the resist in air to confirm the BARC and resist coats were good, the development program was adequate, and to validate that there were no contamination issues. Once the 193-nm resist processing issues were worked, a baseline was established.

The next few sections will present the SEM images of the baseline process, for three sets of exposure for the two 193-nm resists. In all cases resist A is 100-nm thick on 45-nm thick DUV32-6 BARC, and resist B is 150-nm thick on 45-nm thick DUV32-6 BARC.

3.2.1 Exposures at *NA* = 0.50

The first set of exposures in air were done at a *NA* = 0.50, which is equivalent to an interference angle of 30°. Most photolithography steppers used today in industry are capable of exposing images at *NA* = 0.50. The pitch made by the corner cube interferometer at 30° in air is 213 nm. At this particular *NA* the feature pitch is equal to the exposing wavelength ($\Lambda = \lambda = 213$ nm).

At a *NA* = 0.50 excellent lines were achieved as seen in Figure 5, with a measured pitch of 222 nm. The dense lines pitch of 222 nm is slightly larger than the target of 213nm (28.7° instead of 30°). It is of no surprise that excellent lines were patterned, as this represents the technology node for which these 193-nm resist were designed.

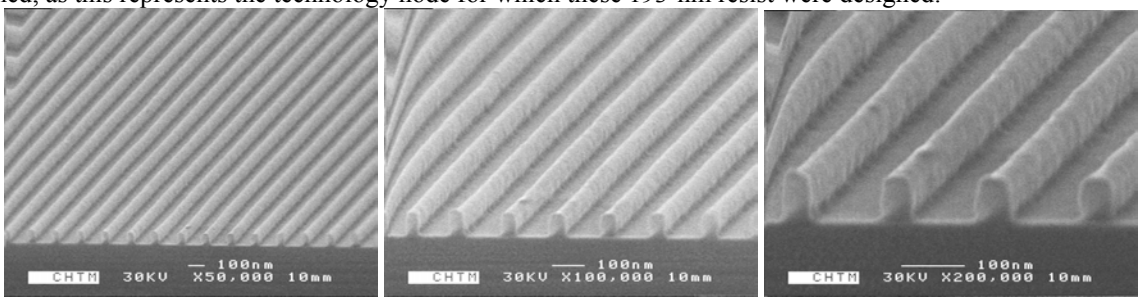


Figure 5 – 222 nm pitch with ~71 nm lines in resist A (*n* = 1 (air), *NA* = 0.50).

As expected at a *NA* = 0.50 excellent lines were achieved in resist B as seen in Figure 6, with a measured pitch of 213 nm. There are some standing wave patterns on the resist sidewalls however. The thinner resist A has no visible standing wave, most likely due to the resist thickness being to small for more than one standing wave to exist $\left(\left(t = 100nm\right) \le \left(2 \cdot \frac{\lambda}{4} = 106.5nm\right)\right)$. Resist B is thick enough to have two standing wave half-periods $\left(\left(t = 150nm\right) \le \left(3 \cdot \frac{\lambda}{4} = 159.75nm\right)\right)$.

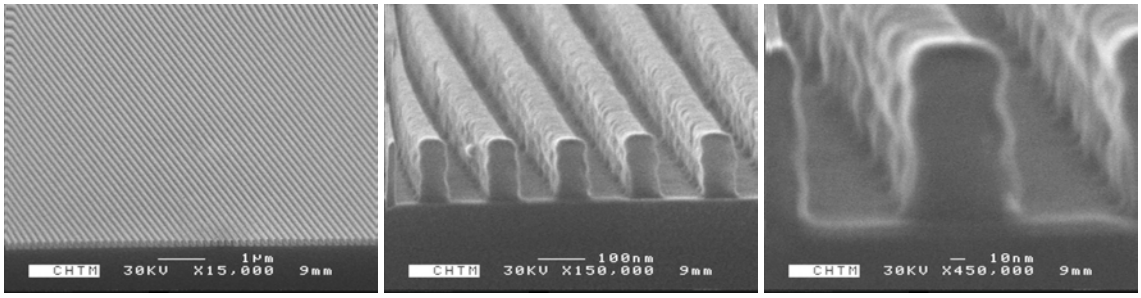


Figure 6 – 213 nm pitch with ~102 nm lines in resist B (*n* = 1 (air), *NA* = 0.50).

3.2.2 Exposures at *NA* = 0.71

The second set of exposures in air were done at a *NA* = 0.71, which is equivalent to an interference angle of 45°. The pitch made by the corner cube interferometer at 45° in air is 150 nm.

As the pitched is reduced to 151 nm with a *NA* = 0.71 it was still possible to get good lines in resist A. The measured pitch in Figure 7 is 151 nm. The sidewall profiles of the lines are more rounded than the lines patterned at *NA* = 0.50.

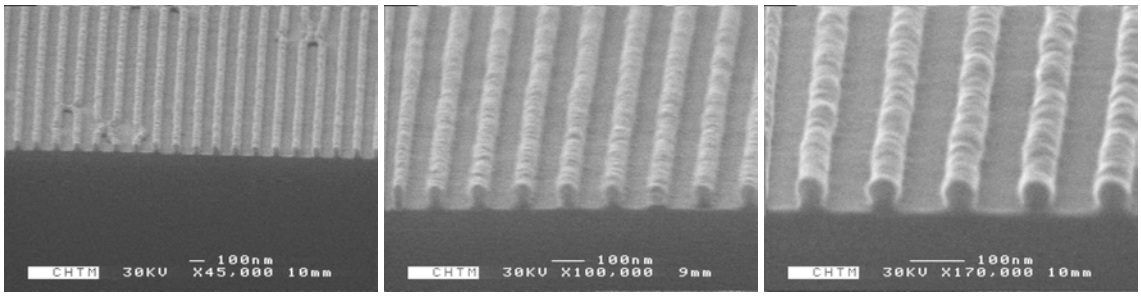


Figure 7 – 151 nm pitch with ~64 nm lines in resist A (n = 1 (air), NA = 0.71).

It was possible to get good lines as well in resist B as seen Figure 8, with a measured pitch of 151 nm. The sidewall profiles of the lines have standing waves, and are tapered towards the bottom.

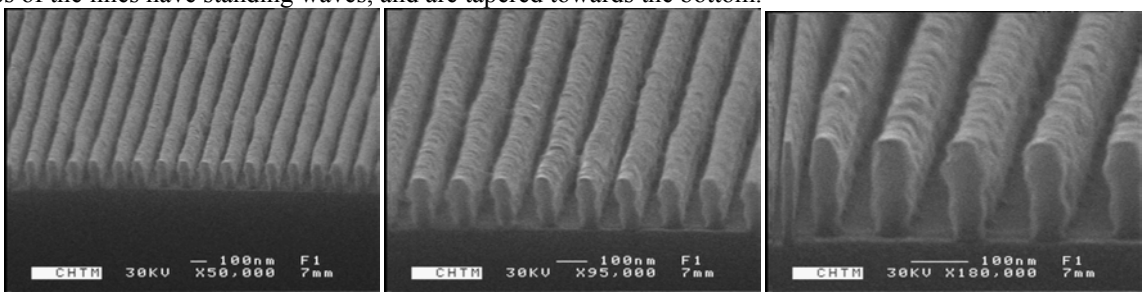


Figure 8 – 151 nm pitch with ~85 nm lines in resist B (n = 1 (air), NA = 0.71).

3.2.3 Exposures at NA = 0.98

The third and last set of exposures in air were done at NA = 0.98, which is equivalent to an interference angle of 80°. This NA represents the practical limit of how small a pitch can be made from at 213-nm light source in air. The pitch made by the corner cube interferometer at 80° in air is 108 nm, whereas the theoretical limit is 106.5 nm in air (λ/2).

As the pitch is reduced to 108 nm at NA = 0.98, near the theoretical limit of 106.5 nm at NA = 1 in air, we were still able to get lines in resist A. Figure 9 has a measured pitch of 108 nm. Many of the lines started to collapse due to the rounded profiles of the resist.

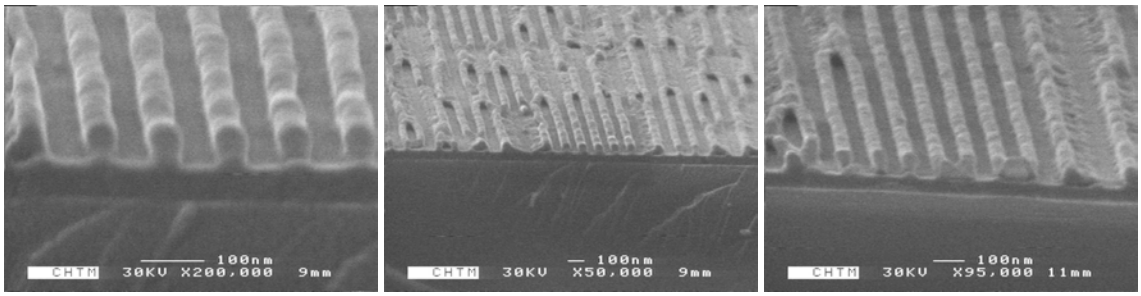


Figure 9 – 108 nm pitch with ~60 nm lines in resist A (n = 1 (air), NA = 0.98).

For resist B as the pitched is reduced to 108 nm, near the theoretical limit of 106.5 nm in air, all the lines collapsed, see Figure 10. Resist B was too thick to pattern lines this small, with a measured pitch of 108 nm. Due to the large aspect ratio (height/width) the lines collapsed from capillary forces exerted on the lines during the drying of the wafer after develop and rinse. As the water dries unevenly between the lines, capillary forces acting on the walls of the lines pull them over.

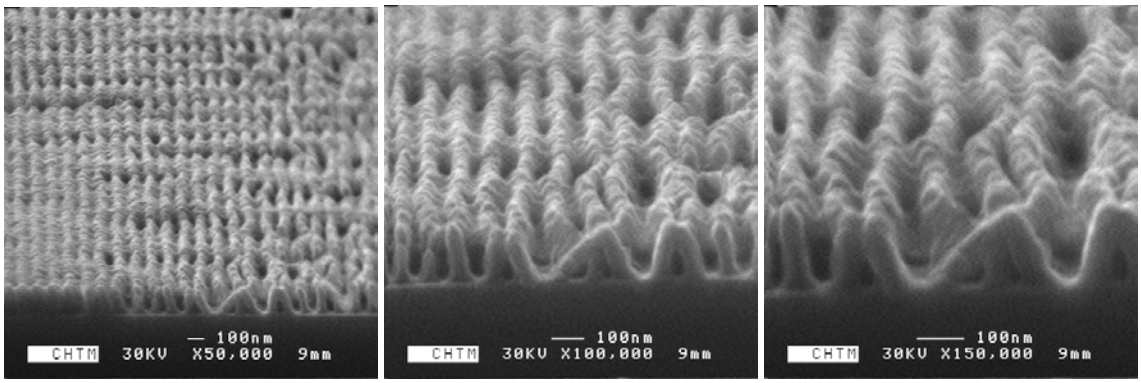


Figure 10 – 108 nm pitch with ~66 nm lines in resist B (n = 1 (air), NA = 0.98).

3.3 Exposure in DI H_2O

After the resist samples were baselined in air, the immersion cell was filled with deionized water and immersion-exposed resist samples were made. In order to ensure that the DI water was not contaminated the cell was drained, rinsed, and filled with fresh DI water between runs. There were no obvious problems with contamination during these DI water immersion experiments.

The next few sections will present the SEM images of the DI water immersion lithography, for three same three sets of exposure as done in air for the two 193-nm resists. Again in all cases, resist A is 100-nm thick on 45-nm thick DUV32-6 BARC, and resist B is 150-nm thick on 45-nm thick DUV32-6 BARC.

3.3.1 Exposures at NA = 0.50

The first set of immersion exposures in DI water were done at NA = 0.50, which is equivalent to an interference angle of 30°. The pitch made by using the corner cube interferometer at 30° in DI water is 151 nm instead of the 213nm for air. The difference is that the pitch is smaller by a factor of n_{H_2O}. This DI water immersion exposure is also equivalent to exposing at $NA \approx 0.71$ in air.

Resist A performed very well when immersed in DI water. At NA = 0.50 excellent lines were achieved as seen in Figure 11, with a measured pitch of 151 nm. These lines even look better than the 151 nm pitched lines exposed in air. This may be due to the higher index of refraction of the DI water, which provides a partial top antireflection coating and reduces the standing wave intensity within the resist.

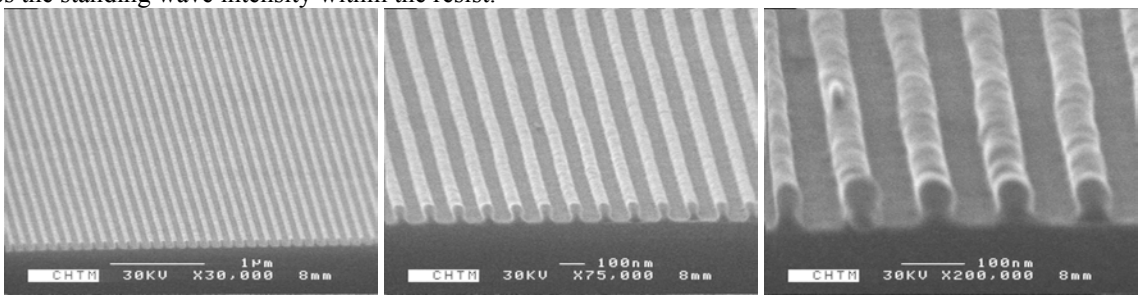


Figure 11 – 151 nm pitch with ~75 nm lines in resist A (n = 1.42 (H_2O), NA = 0.50).

Resist B also performed very well when immersed in DI water. At NA = 0.50 good lines were achieved with a measured pitch of 148 nm. Similarly, these lines even look better than the 151 nm pitched lines exposed in air. There still are standing waves on Resist B lines.

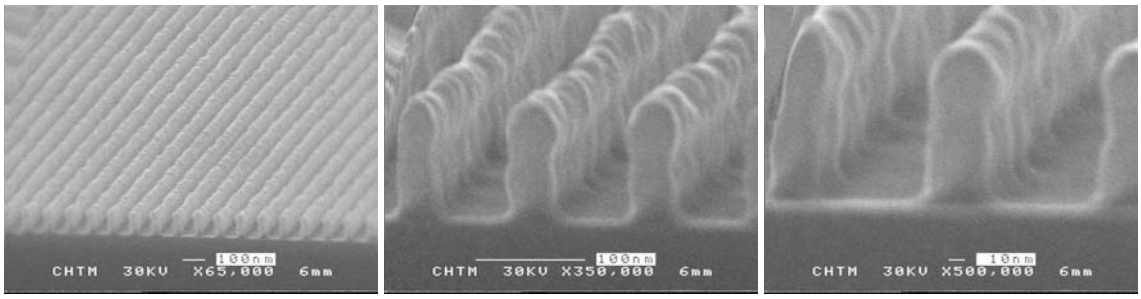


Figure 12 – 148 nm pitch with ~56 nm lines in resist B (n = 1.42 (H_2O), NA = 0.50).

3.3.2 Exposures at NA = 0.71

The second set of immersion exposures in DI water were done at NA = 0.71, or an interference angle of 45°. The pitch made by the corner cube interferometer at 45° in DI water is 107 nm. This is equivalent to exposing at $NA \approx 0.995$ in air.

As the pitched is reduced to 108 nm in DI water immersion at NA = 0.50 good resist lines were resolved in resist A. The measured pitch in Figure 13 is 108 nm. Unlike the exposure of 108-nm pitch lines in air, none of the lines collapsed in DI water immersion lithography, as seen in Figure 13; again possibly because of the top-ARC effect of the liquid.

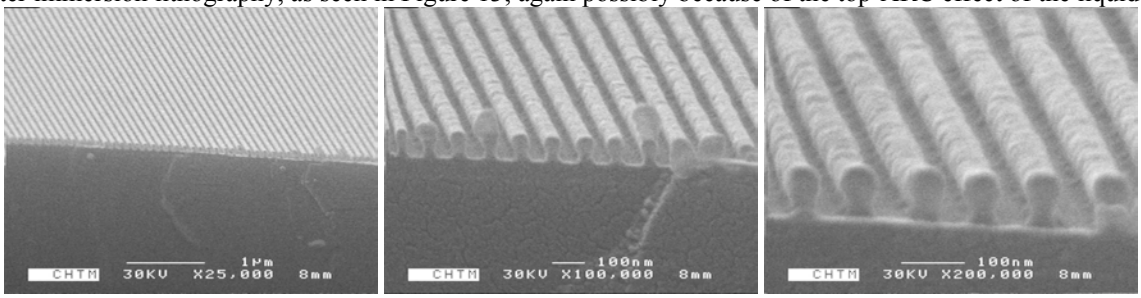


Figure 13 – 108 nm pitch with ~60 nm lines in resist A (n = 1.42 (H_2O), NA = 0.71).

For resist B as the pitched is reduced to 108 nm in DI water immersion all the lines collapsed. Just as the case of exposing 108-nm pitch lines in air, Resist B was too thick to pattern lines this small.

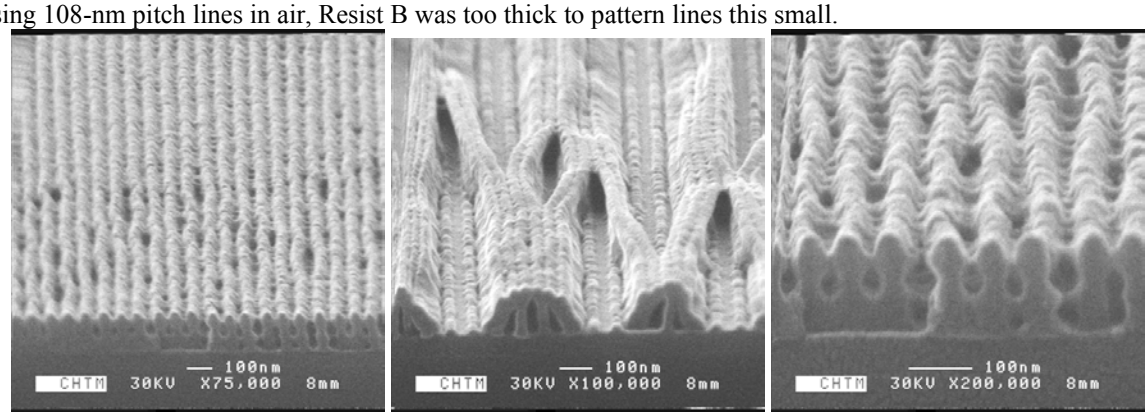


Figure 14 – ~108 nm pitch with ~59 nm lines in resist B (n = 1.42 (H_2O), NA = 0.71).

3.3.3 Exposures at NA = 0.98

The last set of immersion exposures in DI water were done at NA = 0.98, which is equivalent to an interference angle of 80° or an air-referenced NA of 1.41. The pitch made by the corner cube interferometer at 80° in DI water is 77 nm. Only through immersion lithography can you image this small at this wavelength. Unfortunately, neither of the 193-nm resists

was capable of imaging lines on a 77-nm pitch. The surface of the resist was roughened, and some resemblance of a grating pattern could be discerned. No pictures were recorded.

3.4 Exposure in Krytox®

The last sets of immersion resist samples were done in Krytox®. The immersion cell was filled with filtered Krytox® GPL 100 PFPE oil. There were no observed problems with contamination when running these Krytox® immersion runs. The Krytox® immersion samples were rinsed in Vertrel® XF after exposure to remove any residual Krytox® film on the surface of the resist.

The next few sections will present the SEM images of the Krytox® immersion lithography, for the same three sets of exposure as done in air for the two 193-nm resists. In all cases resist A is 100-nm thick on 45-nm thick DUV32-6 BARC, and resist B is 150-nm thick on 45-nm thick DUV32-6 BARC.

3.4.1 Exposures at *NA* = 0.50

This set of immersion exposures in Krytox® were done at *NA* = 0.50, which is equivalent to an interference angle of 30°. The pitch made by the corner cube interferometer at 30° in Krytox® is 165 nm. This is equivalent to exposing at a *NA* ≈ 0.65 in air.

Resist A also performed very well when immersion in Krytox®. At a *NA* = 0.50 excellent lines were achieved as seen in Figure 11, with a measured pitch of 161 nm. There still is a bit of rounding of the resist pattern. The index of reflection measurement for Krytox® on this sample is 1.289 ± 0.0018.

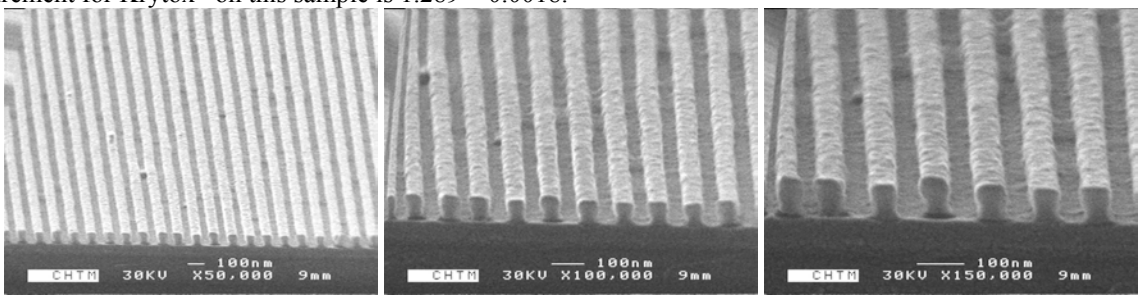


Figure 15 – 161 nm pitch with ~72 nm lines in resist A (*n* = 1.29 (Krytox®), *NA* = 0.50).

Resist B gave good-looking lines, but most of them collapsed because of their high aspect ratio (height/width) when exposed in Krytox® immersion. The measured pitch is 160 nm. The index of measurement for Krytox® on this sample is 1.296 ± 0.0018.

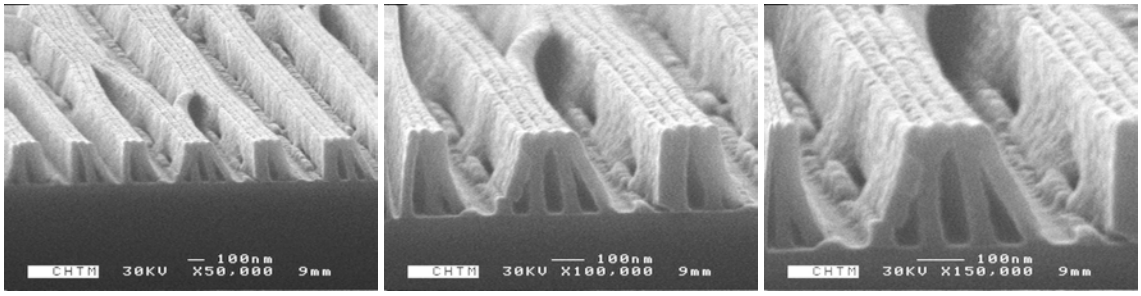


Figure 16 – 160 nm pitch with ~65 nm lines in resist B (*n* = 1.29 (Krytox®), *NA* = 0.50).

3.4.2 Exposures at *NA* = 0.71

Attempts were made to get a set of immersion exposures in Krytox® with a *NA* = 0.71, which is equivalent to an interference angle of 45°. Pitch made by the corner cube interferometer at 45° in Krytox® is 117 nm. This is the equivalence of exposing at *NA* ≈ 0.91 in air.

Resist A also performed well when immersed in Krytox®, at *NA* = 0.71. Lines were achieved as seen in Figure 17, with a measured pitch of 116 nm. There is resist scumming and footing between the lines in the resist pattern. The index of measurement for Krytox® on this sample is 1.302 ± 0.0014.

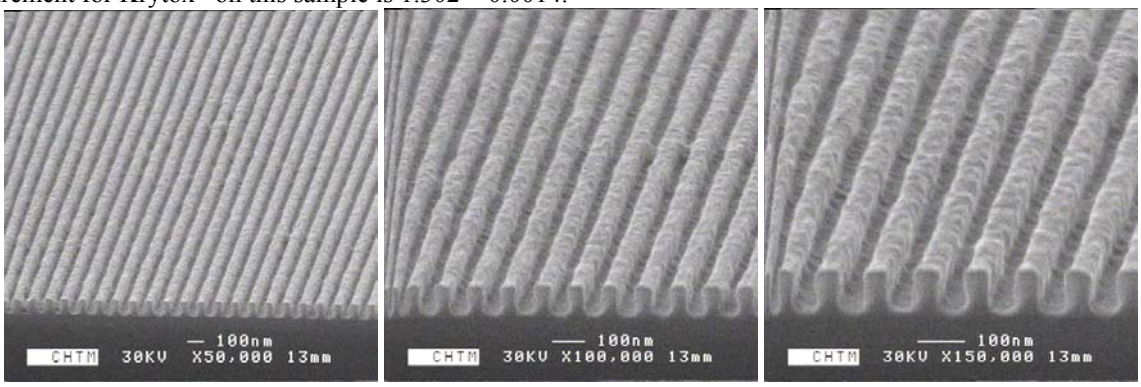


Figure 17 – 116 nm pitch with ~48 nm lines in resist A (*n* = 1.29 (Krytox®), *NA* = 0.71).

No images were recorded of resist B in Krytox® with a *NA* = 0.71.

3.4.3 Exposures at *NA* = 0.98

Attempts were made to get a set of immersion exposures in Krytox® with a *NA* = 0.98, which is equivalent to an interference angle of 80°. Pitch made by the corner cube interferometer at 80° in Krytox® is 83 nm. This is no equivalence of exposing in air, only through immersion lithography can you image this small. However as no surprise, no good samples were made in either resist A or resist B.

No images were recorded of resist A in Krytox® with a *NA* = 0.98.

Figure 18 show an attempt to make a grating in Krytox® at *NA* = 0.98. As can be seen the resist was too thick and imaging resolution of the resist is questionable. The measured pitch is 85 nm. The index of measurement for Krytox® on this sample is 1.272 ± 0.0010.

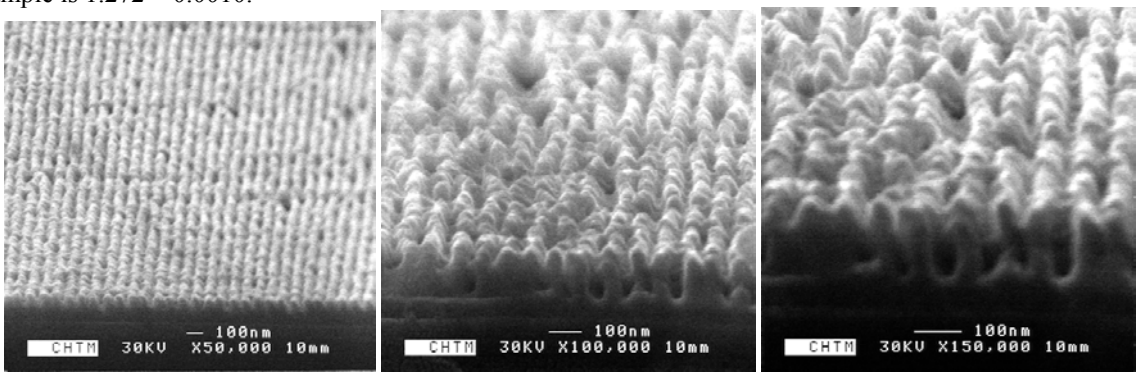


Figure 18 – 85 nm pitch with ~37 nm lines in resist B (*n* = 1.29 (Krytox®), *NA* = 0.98).

4. SUMMARY

4.1 Minimum Pitch Comparisons For Various Exposure Media

Immersion lithography reduces the minimum pitch that can be written by a factor of the index of refraction *n* of the immersion medium. The following table gives the theoretical minimum pitch for several exposure media at different wavelengths.

Exposure Medium / Exposure Wavelength (nm)	Air Pitch [Half-Pitch] (nm)	DI Water Pitch [Half-Pitch] (nm)	Krytox® Pitch [Half-Pitch] (nm)
257	128 [64]	94 [47]	Unknown
213	107 [54]	76 [38]	83 [42]
193	97 [49]	66 [33]	73 [37] [2]
157	79 [40][1]	-------	58 [29] [3]

1.) Minimum pitch in Argon gas, or vacuum.
2.) Interpolated index of refraction *n* for Krytox® at 193 nm.
3.) M. Switkes, M Rothschild, "Immersion lithography at 157 nm" [5]

Table 2 – Minimum theoretical exposure pitch for various exposure mediums.

4.2 Resist Results

Both of the 193-nm resists investigated showed comparable imaging properties and were capable of imaging dense lines with a 151-nm pitch in air. If resist B had been thinner it is believed that both resist would have been able to image dense lines down to 108-nm pitch, which would represent the practical limit of photolithography at a 213-nm exposure wavelength. Both 193-nm resists appeared to be compatible with the DI water and Krytox® immersion media. The immersion exposure gave better line profiles then the equivalent air-exposure at the same pitch. Neither of the 193-nm resists was capable of imaging dense lines at the highest resolution attempted (77-nm pitch in H_2O; 85-nm pitch in Krytox®). This is likely due a combination of increasing aspect-ratio and diffusion-limited resolution of the resist at these small scales.

4.3 Conclusions

Optical photolithography is constantly evolving to image ever-decreasing feature sizes. Liquid Immersion lithography (LIL) is a good candidate for a new image enhancing technique to further push the limits of optical lithography. With immersion lithography it is possible to print smaller dense lines on a wafer than currently possible in air. Liquid immersion lithography can work in conjunction with other image enhancing techniques such as off-axis illumination, optical-proximity correction, phase-shift masks, and imaging-interferometric lithography. Polarization issues will become more pronounced at the higher angles possible in the resist with immersion lithography.

It has been demonstrated in this work that it possible to printing dense lines using liquid immersion lithography at an *NA* of 0.71; this is equivalent to printing dense lines in air with an *NA* of 0.98. Dense lines with a half-pitch of 54 nm in a 193-nm resist were demonstrated, with exposures both in air (*NA* = 0.98) and in liquid immersion media (*NA* ≈ 0.7).

Two immersion liquids have been identified that have more than adequate transparency at the 193-nm lithography node: deionized water and Krytox® PFPE oil. At the 157-nm lithography node PFPE oils still have some transmission, although path lengths need to be extended.

It is foreseeable that with resist improvements, liquid immersion lithography technology can be extended to the 32-nm half-pitch technology node. Based on this, optical lithography could be extended until 2016, according to the initial timetable of the 2002 International Technology Roadmap for Semiconductor (ITRS). This would ease the timescales of switching over to a next generation lithography (NGL) technology.

4.4 Future studies

For the continuation of this work, a thinner version of the 193-nm resist used should by evaluated, in attempt to make dense lines with an 80-nm pitch. Improvements are continually being made to resists; additional 193-nm resists should be evaluated for immersion lithography.

Two immersion fluids have been identified for 193-nm lithography, and just one for 157-nm lithography. Other liquids that are transparent and should be evaluated at these DUV wavelengths are cryogenic liquids, such as liquid Argon.

Research is needed to find other immersion fluids, for 193 nm, 157 nm, and any future optical lithography wavelengths. Another approach is to modify PFPE oils to give better transmittance or higher index of reflection to be used at 193 nm and 157 nm.

The immersion cell designed for this work was made to be compatible with many types of illumination sources. Future evaluation can be done at 157 nm or other emerging DUV light sources.

Immersion lithography will need to be implemented and demonstrated in a lens system using masks. Wafer handling systems will need to be demonstrated for putting wafers into and out of immersion lithography systems. Evaluations of process latitude; imaging parameters - depth of focus (DOF), resolution, polarization effects, applicability of RETs; throughput; and cost-of-ownership (COO) models for LIL need to be developed.

Liquid immersion lithography has great potential to extend optical lithography as a new optical enhancement technique. The liquid immersion lithography work done in this work has shown that it can extend current 193-nm resist technology to its limits. Much more work is needed to fully assess the manufacturing viability of this new resolution enhancement technique.

REFERENCES

1. C. Tian, J. Li, P. Wang, T Sun, “Windowless, geometry-independent, phase-matched method for the measurement of the oscillator strength of Xe, Kr, and Ar”, Physical Review A, Vol. 50 No, 2, p.1133, Aug. 1994.

2. C. V. Shank, R. V. Schmidt, “Optical Technique for Producing 0.1-m Periodic Surface Structures”, Applied Physics Letters, Vol. 23 No. 3, Aug. 1973, p154.

3. W. T. Tsang, S. Wang, “Simultaneous Exposure and Development Technique for Making Gratings on Positive Photoresist”, Applied Physics Letters, Vol. 24 No.4, Feb. 1974, p196.

4. J.A. Honffnagle, W. D. Hinsberg, M. Sanches, F. A. Houle, “Liquid Immersion Deep-Ultraviolet Interferometric Lithography”, J. Vac. Sci. Technol. B, Vol 17 No.6, Nov/Dec 1999, p3306-3309

5. M. Switkes, and M. Rothschild, “Immersion Lithography at 157nm”, J. Vac. Sci. Technol. B, Vol 19 No. 6, p2353-2356, Nov/Dec 2001.

6. International Technology Roadmap for Semiconductors at www.public.itrs.net.

7. S. Zaidi and S. R. J. Brueck, “High Aspect-Ratio Holographic Photoresist Gratings,” Appl. Optics 27, 2999 (1988)

Grating Analysis of Frequency Parsing Strategies for Imaging Interferometric Lithography

Eric S. Wu[*a], Balu Santhanam[a] and S. R. J. Brueck[a,b,c]
[a]Department of Electrical and Computer Engineering, Univ. of New Mexico, Albuquerque, NM 87131, USA
[b]Center for High Technology Materials, Univ. of New Mexico, Albuquerque, NM 87106
[c]Department of Physics and Astronomy, Univ. of New Mexico, Albuquerque, NM 87131

ABSTRACT

The limitations of a conventional optical lithography system are consequences of the limited spatial frequency coverage (~ NA/λ) of the optical system. To improve the resolution of printed patterns, *Imaging Interferometric Lithography* (IIL) exploits interference phenomena to produce sub-wavelength structures on the wafer and provides a simple approach to attainment of the ultimate spatial frequency coverage of 2/λ, independent of the optical system NA. In the first promising experiments, different divisons of frequency space between multiple exposures were investigated empirically. The use of multiple exposures includes a better coverage than available from a single exposure, automated software routines or strategies need to be developed to find the optimum setup that makes the best compromise against all of the desirable lithography specifications. From a comprehensive grating analysis, we derive a set of optimal parameters for exposure energy ratio among exposures and search a proper set to improve the aerial image quality. Additionally, comparing with partial coherent imaging schemes, we examine the strengths and weaknesses of IIL through different analyses and these studies provide additional support for IIL as a promising alternative RET for deep sub-wavelength optical lithography.

Keywords: Imaging Interferometric Lithography, gratings, Fourier analysis, MEEF, OPE

1. INTRODUCTION

Imaging interferometric lithography (IIL) exploits interference phenomena to overcome the spatial frequency resolution limitation (~ NA/λ) imposed by a conventional optical lithography system, to improve the resolution of aerial images and produce sub-wavelength structures, and in the process provides a simple approach towards attainment of the ultimate optical spatial frequency coverage of 2/λ [1–5]. Since IIL involves the use of multiple exposures and off-axis illumination, automated software strategies for IIL simulations should be developed to find the optimum solution for meeting all of the specifications needed for the lithography process and to produce a robust, manufacturable process. For the simple example in Ref. 5, where alternative three-exposure IIL strategies containing the same spatial frequency coverage were empirically investigated, the final image results were significantly different due to different frequency parsings between the on-axis and off-axis exposures, and one was clearly more robust to alignment and intensity variations. As a first step towards avoiding such undesirable solutions, we examine pertinent equations in IIL theory to understand the factors that play major roles in the aerial image enhancement process. A detailed Fourier series analysis of the IIL on-axis and off-axis exposures in one-dimensional gratings is presented in Section 3 to fully understand the essentials of constructing a fine-resolution aerial intensity. Common performance measures are evaluated via simulations in MATLAB®. These measures will then be used to study the relative strengths and weaknesses of IIL, in comparison with more traditional partial-coherent illumination imaging. In Section 4 with block simulations, we extend the grating analysis into two dimensions and an optimization procedure for illumination energy in each exposure is described. Finally, we analyze the results and conclude with future study in Section 5.

[*] ericwu@unm.edu; phone 1 505 277-0108

2. FORMULATE IIL SCHEME WITH ENERGY RATIO FACTORS

For Manhattan type of mask where patterns are assembled by different sizes of rectangles, IIL needs three different types of exposures: on-axis (I_{on}), off-axis along x (I_{off_x}) and off-axis along y (I_{off_y}). And the intensity of the aerial image from an IIL scheme can be expressed as:

$$I_{IIL}(x,y) = \alpha \cdot I_{on}(x,y) + I_{off_x}(x,y) + I_{off_y}(x,y) \quad (1)$$

The first parameter, α, represents the energy/intensity ratio between IIL on-axis and off-axes exposures. Although the square-law intensity from an asymmetric, single off-axis exposure covers both positive and negative frequencies, we require two exposures for each type of off-axis exposure along x and y to ensure image symmetry with respect to focus (e. g. to restore telecentricity.). With Figure 1 showing the Fourier spatial frequency coverage of each exposure, a parsing scheme between I_{on} and I_{off_x} can be expressed with three parameters: NA_{on}, NA_{off} are the numerical aperture sizes of the optical system for the IIL on-axis and off-axis exposures separately and NA_{oDC} indicates the tilt angle in the off-axis illumination. In order to generalize the problem for any illumination light source, we use λ as the light wavelength to normalize the spatial frequency domain shown in Figure 1.

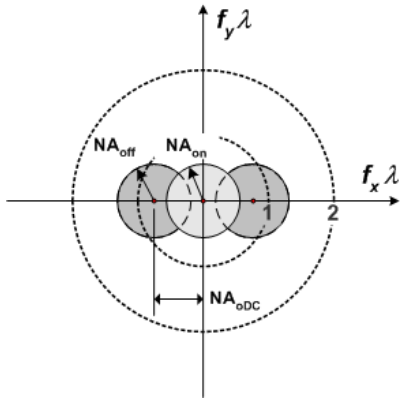


Figure 1: One-dimensional NA frequency parsings for IIL scheme

Formulating the IIL scheme, the intensity from one IIL off-axis exposure along x is obtained by using tilted coherent illumination with a plane wave as $e^{-j2\pi f_{oDC}x}$ where angle $\theta = \sin^{-1}(NA_{oDC}) = \sin^{-1}(f_{oDC}\cdot\lambda)$. This off-axis illumination causes the spatial frequency components of the electrical field after the mask $E(f)$ to be downshifted to $E(f+f_{oDC})$. These components, after passing through the lens pupil with size of NA_{off_x}, which is denoted as $E_{off_x}(f+f_{oDC})$ are interfered with an extra reference beam, a plane wave $\gamma\cdot e^{-j2\pi f_{oDC}x}$ which is the zero-order component of the mask transmission. The reference plane wave is also tilted by the same angle θ but with an intensity ratio of γ^2 with respect to the off-axis illumination intensity. For an intensity ratio $\gamma \le 1$, this can be produced by using an intensity attenuator in the system setting for IIL in Figure 2. In summary, the image reconstruction for this off-axis exposure can be treated as being similar to the recording process of holography[6] and expressed as:

$$\begin{aligned} I'_{off_x}(x) &= \left|\mathbf{F}^{-1}\left\{E_{off_x}(f_x+f_{oDC})\right\} + \gamma\cdot e^{-j2\pi f_{oDC}x}\right|^2 \\ &= \gamma^2 + \left(\mathbf{F}^{-1}\left\{E_{off_x}(f_x+f_{oDC})\right\}\right)\cdot\left(\mathbf{F}^{-1}\left\{E_{off_x}(f_x+f_{oDC})\right\}\right)^* + \\ &\quad \gamma\cdot\left(\mathbf{F}^{-1}\left\{E_{off_x}(f_x+f_{oDC})\right\}\cdot e^{j2\pi f_{oDC}x}\right) + \gamma\cdot\left(\left(\mathbf{F}^{-1}\left\{E_{off_x}(f_x+f_{oDC})\right\}\right)^*\cdot e^{-j2\pi f_{oDC}x}\right) \end{aligned} \quad (2)$$

We denote the information in the Fourier domain as $I'_{off_x}(f_x) = \mathbf{F}\left\{I'_{off_x}(x)\right\}$ and $\mathbf{F}^{-1}\{\cdot\}$ is the inverse Fourier transform. Also $E_{off}(f_x)$ is defined as the high spatial frequency portion of the mask's transmittance function $T(f_x)$ that passes through the pupil filter in the optical system as $H_{off}(f_x)$ with size of NA_{off}. It can be expressed as

$$E_{off}(f_x) = T(f_x)\cdot H_{off}(f_x - f_{oDC}) \quad (3)$$

If $T(f_x)$ is calculated from a binary image mask and $H_{off}(f_x)$ is a simple low-pass cutoff pupil-place filter, we perform off-axis exposures twice at θ and $-\theta$ along the x-axis for telecentricity reasons and then transfer the result to the Fourier domain. This symmetry becomes critical as the image through focus is investigated. Exactly at focus there is no difference between twice the intensity at one offset angle or using both plus and minus offset angles. The final equation for one (along x) off-axis exposure is :

$$\begin{aligned} I_{off_x}(f_x) &= \mathbf{F}\left\{I'_{off_x}(x) \text{ with } \theta\right\} + \mathbf{F}\left\{I''_{off_x}(x) \text{ with } -\theta\right\} \\ &= 2\cdot\left\{\gamma^2\cdot\delta(f_x)\right\} + 2\cdot\left\{E_{off_x}(f_x+f_{oDC}) \otimes E^*_{off_x}(-f_x+f_{oDC})\right\} + 2\cdot\gamma\cdot\left\{E_{off_x}(f_x) + E^*_{off_x}(-f_x)\right\} \end{aligned} \quad (4)$$

In our previous paper[5], the second term in Eq. (4) was referred to as the *quadratic* term and the third term was called the *linear* term since it is proportional to the transmittance function of the mask.

Now the IIL scheme in Eq. (1) with all three types of exposures (on-axis, off-axis along x and off-axis along y) can be formulated as

$$\begin{aligned}&\mathbf{F}\left\{I_{IIL}(x,y)\right\}=\\&\alpha\cdot\mathbf{F}\left\{I_{on}(x,y)\right\}\\&+2\cdot\left\{\gamma_x^{\,2}\cdot\delta(f_x,f_y)\right\}+2\cdot\left\{E_{off_x}(f_x+f_{oDC},f_y)\otimes E^{*}_{off_x}(-f_x+f_{oDC},f_y)\right\}+2\cdot\gamma_x\cdot\left\{E_{off_x}(f_x,f_y)+E^{*}_{off_x}(-f_x,f_y)\right\}\\&+2\cdot\left\{\gamma_y^{\,2}\cdot\delta(f_x,f_y)\right\}+2\cdot\left\{E_{off_y}(f_x,f_y+f_{oDC})\otimes E^{*}_{off_y}(f_x,-f_y+f_{oDC})\right\}+2\cdot\gamma_y\cdot\left\{E_{off_y}(f_x,f_y)+E^{*}_{off_y}(-f_x,f_y)\right\}\end{aligned}\tag{5}$$

where three energy ratio parameters need to be optimized: α,γ_x and γ_y.

2. ONE DIMENSIONAL GRATING ANALYSIS OF IIL

Considering the transmittance function from an idealized grating mask along x-axis with period p as:

$$T(x)=\begin{cases}1 & 0\le|x|<s/2\\ 1/2 & |x|=s/2\\ 0 & s/2<|x|\le p/2\end{cases}\tag{6}$$

where grating space is centered at the origin and space width is denoted as S. In this 1D case IIL only needs two different types of exposure: on-axis (I_{on}) and off-axis along x (I_{off_x}), since there is no image information in the off-axis exposure along y (I_{off_y}) which only adds an intensity baseline or a DC-term to the aerial image.

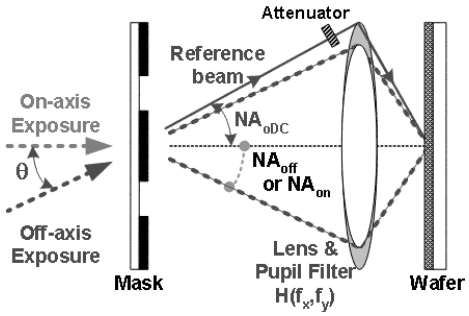


Figure 2: One-dimensional IIL setup

To generalize the Fourier series in Eq. (6), we set the fundamental frequency $f_0=1/p=1$ which is equivalent to normalizing the mask period equal to 1, we have the real-value coefficients c_k in

$$T(x)=\sum_{k=-\infty}^{+\infty}c_k\cdot e^{j2\pi kx}=\sum_{k=-\infty}^{+\infty}\left(s\cdot sinc(ks)\right)\cdot e^{j2\pi kx}\tag{7}$$

Then $E_{off_x}(x)$ from Eq. (3) that through a simple cutoff pupil filter can be expressed as

$$E_{off_x}(x)=\mathbf{F}^{-1}\left\{E_{off_x}(f_x)\right\}=\sum_{k_1}^{k_2}c_k\cdot e^{j2\pi kx}=\mathrm{Re}\left\{E_{off_x}(x)\right\}+j\,\mathrm{Im}\left\{E_{off_x}(x)\right\}\tag{8}$$

where k_1 and k_2 are specified by the limit of the pupil $H_{off}(f_x)$. By transforming Eq. (5) back to the spatial domain, the IIL intensity result for 1D grating is shown as

$$I_{IIL_grating}(x)=\alpha\cdot I_{on}(x)+2\cdot\gamma^2+2\cdot\left\{\mathrm{Re}\left\{E_{off_x}(x)\right\}^2+\mathrm{Im}\left\{E_{off_x}(x)\right\}^2\right\}+4\cdot\gamma\cdot\mathrm{Re}\left\{E_{off_x}(x)\right\}\tag{9}$$

Since the advantage of IIL is in its capability to print much smaller and dense patterns than the conventional lithography by retaining more of the high frequency image components from the mask in $E_{off_x}(x)$ that are missed by $I_{on}(x)$, the limiting-case scenario in IIL is when only the first order frequency component is captured, $k_1 = k_2 = 1$ in Eq. (8), resulting in

$$E_{off_x}(x)=\frac{\sin(\pi s)}{\pi}\cdot\left(\cos(2\pi x)+j\sin(2\pi x)\right)\tag{10}$$

Then in this case, IIL 1D grating intensity with a coherent exposure on $I_{on}(x)$ is

$$I_{IIL_grating}(x) = \left\{\alpha \cdot s^2 + 2\cdot\gamma^2 + 2\cdot\left\{\frac{\sin^2(\pi s)}{\pi^2}\right\}\right\} + \left\{4\cdot\gamma\cdot\frac{\sin(\pi s)}{\pi}\right\}\cdot\cos(2\pi x) \quad \text{[while } k_1=k_2=1\text{]} \tag{11}$$

Clearly, the first term is the DC term and we can obtain the equation for contrast/visibility as

$$\mathcal{V}_{(\alpha,\gamma)} = \left\{\frac{4\cdot\gamma\cdot\frac{\sin(\pi s)}{\pi}}{\alpha\cdot s^2 + 2\cdot\gamma^2 + 2\cdot\left\{\frac{\sin^2(\pi s)}{\pi^2}\right\}}\right\} \quad \text{[while } k_1=k_2=1\text{]} \tag{12}$$

By setting $s = 0.5$ for dense grating as duty cycle = 50%, we are able to generate a visibility contour plot shown in Figure 3. The plot indicates an important conclusion: for lithography purpose, we would like to adjust the energy ratio parameters (α,γ) to gain the best image contrast as possible. Furthermore, while searching the optimum solutions of energy ratio for a 2D mask, a proper probing range for α is (0 … 4) and γ is (0 … 1.5).

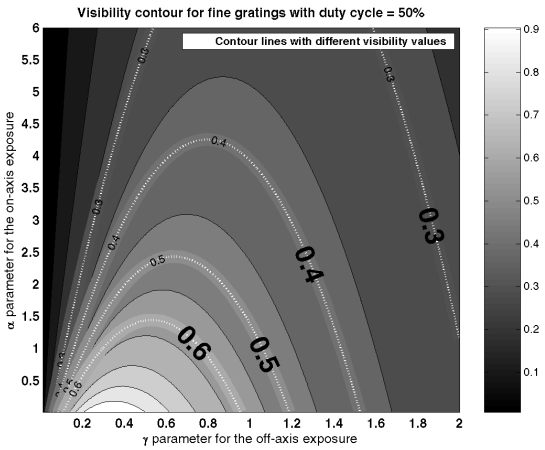


Figure 3: Visibility contour plot for 1D grating

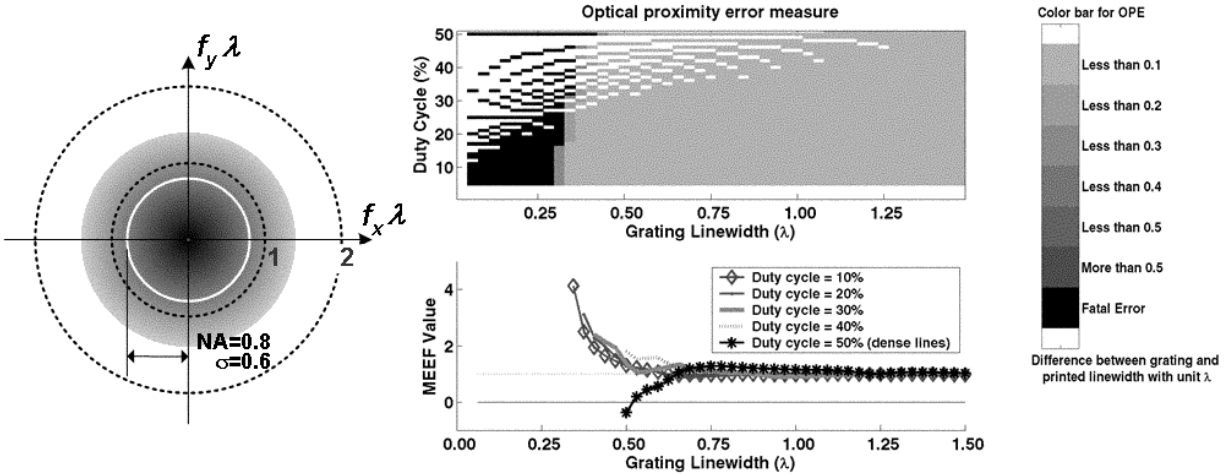
Comparing the aerial image analysis of one-dimensional gratings from both IIL and partial-coherent illumination schemes can help us to understand the essential differences between these two imaging approaches. For simulation convenience and as a setting for the rest of this paper, we set $NA_{on} = 0.8$ and $NA_{off} = 0.8$ for IIL optical system and set the illumination off-axis angle $NA_{oDC} = 0.84$ in order to have higher frequency coverage of the mask. The illumination light source has wavelength λ in general. As to the partial coherent illumination scheme, $NA = 0.8$ and the partial coherence factor σ is set to 0.6. [**is this a scalar calculation or do you treat polarization?]** Furthermore, to quickly review printed pattern results directly from aerial images, a constant threshold model[8] (CTM) for photoresist with a threshold value Th_0 was applied for all simulations. The Th_0 values for critical dimension (CD) measurements in Figure 4 are different for the partial coherent (Th_0 = 0.3) and IIL (Th_0 = 0.6 or 0.7), since IIL has additional intensity baseline $2\gamma^2$ in Eq. (9) that requires a high contrast resist. For analysis on frequency parsing strategies, we do IIL simulations with two different coverage configurations: (i) allowing overlapped between on-axis and off-axis coverage without any pupil filtering, (ii) no overlapping coverage in between while applying moon-shaped pupil filters (circular filters with a second overlapped circle missing, see Fig. 4) during the off-axis exposures.

The first study of interest is the difference between printed patterns and the grating masks as optical proximity error (OPE) measurement at different duty cycles. The OPE plots in Figure 4 show that IIL in both configurations obviously can print more smaller sub-wavelength features than the conventional partial coherent scheme. To study the sensitivity of the printed critical CD on the wafer as we gradually change the reticle CD in gratings, we can use the MEEF[9] concept and measure the results when the gratings' dark lines approach the sub-wavelength limit. Shown in Figure 4, the MEEF values from the partial coherent illumination are gradually diverging from ideal value one; but IIL with help from the off-axis exposures can have the deviation happen later at a smaller sub-wavelength feature, which means more fine features can be printed with IIL technique.

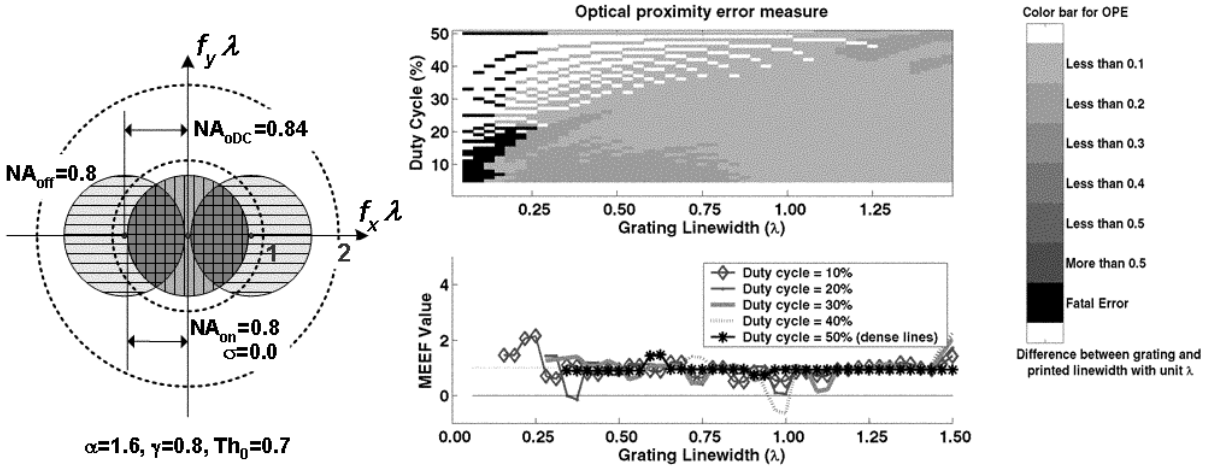
In Figure 4(c), IIL with moon-shaped pupil filters has the cutoff frequency between on-axis and off-axis exposures at $0.8/\lambda$. This abrupt frequency transition cause MEEF lines to go up and down irregularly around one, not as smooth as MEEF lines for the partial coherent scheme in Figure 4(a). To reduce the abrupt transition between on-axis and off-axis exposure coverage, a certain degree of overlapping between on-axis and off-axis coverages should be considered as the case in Figure 4(b) without pupil filters, which MEEF lines shows less roughness than in Figure 4(c). Furthermore by using apodized pupil filter designs such as Gaussian pupil filters, a smoother transition can be expected in the MEEF and OPE plots.

As to the depth of focus (DOF) analysis, since the reference beam in IIL coming from a large incident angle upon the wafer, IIL theoretically have less DOF than the partial coherent illumination for larger patterns. [**not correct, if only the off axis beam comes in at high angle and the information beams are at a lower angle, the depth-of-field corresponds to the information beams, the best situation is symmetrized with the reference beam from one side**

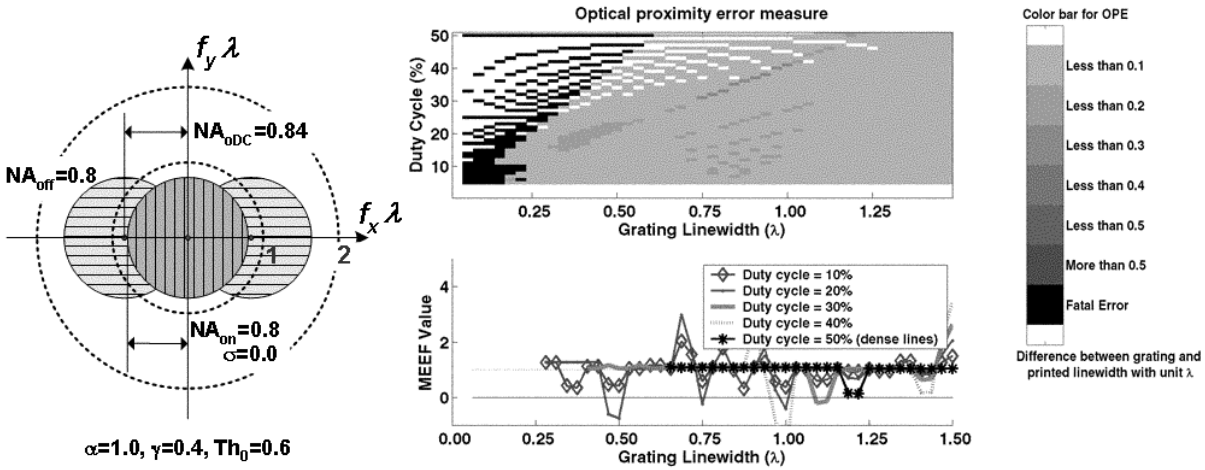
and the information beams centered at the negative angle, you haven't shown, or probably done, any calc. on DOF!] But overall on sub-wavelength gratings, IIL has better exposure latitude performance than the partial coherent illumination method due to aerial image enhancement (higher visibility) in IIL. However in optical lithography, the deficiency of DOF is becoming less critical, since it can easily overcome by reducing the thickness of the photoresist layer and also can be overcome by advance pupil filter designs[10]. Therefore, this IIL drawback is more than compensated by the benefits offered by IIL for sub-wavelength masks.



(a) Partial coherent illumination scheme



(b) IIL scheme without pupil filters



(c) IIL scheme with moon-shape pupil filters on off-axis illuminations

Figure 4: OPE and MEEF measurements with different illumination schemes

3. SEARCHING OPTIMUM ENERGY RATIOS BY 2D GRATING ANALYSIS

To find suitable energy ratios γ_x and γ_y in Eq. (5) for individual I_{off_x} and I_{off_y}, 1D grating analysis is not sufficient for processing 2D masks, because γ_x and γ_y are correlated to each other and not always the same due to the orientation of patterns. Furthermore, running the whole mask simulations according to different $(\alpha, \gamma_x, \gamma_y)$ energy ratios would be very time-consuming and would rapidly become impractical as the mask becomes more complicated. Therefore, we need a routine to cut down the searching time for optimum energy ratios first by doing simpler simulations on simple 2D structures that are related to the mask.

Giving a Manhattan type of 2D mask in Figure 5(a), the mask is assembled by different size of rectangles. For this reason, doing analysis on each different 2D block (rectangle) extracting from the mask, as shown in Figure 5(b), can help us to first find some suitable sets of $(\alpha, \gamma_x, \gamma_y)$ from numerous possible solutions and later individually apply and analyze their performances on the actual 2D mask. In order to consider the difference between dense and isolated features, we can set different Period X and Y in Figure 5(b) for different duty cycles along X and Y axes. The performance measure on a printed block pattern in Figure 5(c) is done by an error measure function of $(\alpha, \gamma_x, \gamma_y)$. It includes three common analyses used in lithography: (i) overlay area between the printed and desired pattern, (ii) line-end shortening (LES) effect, and (iii) corner rounding (CR) effect. Since a resist CTM with a parameter Th_0 is used, by using different Th_0 value for CTM can be related to different exposure energy we actually apply. Hence, the optimization procedure should output four numerical numbers $(\alpha, \gamma_x, \gamma_y, Th_0)$ for actual exposure energy. After obtaining block simulation results, we can quickly evaluate the performance based on one overlay area measurement, two LES along x and y, and one CR measurement. Then depending on results from the error measure function, we can rule out bad energy settings that cause fatal errors and sort the suitable results by performance. At the final stage we pick several best-performing settings from block simulations and run them again with the whole 2D mask. By evaluating the error measure based on the de-

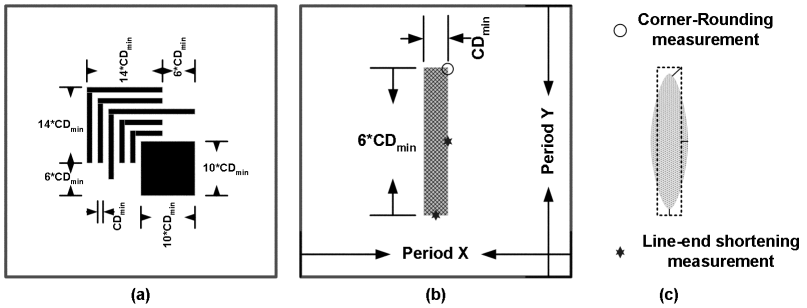


Figure 5: 2D grating analysis from a 2D Mask

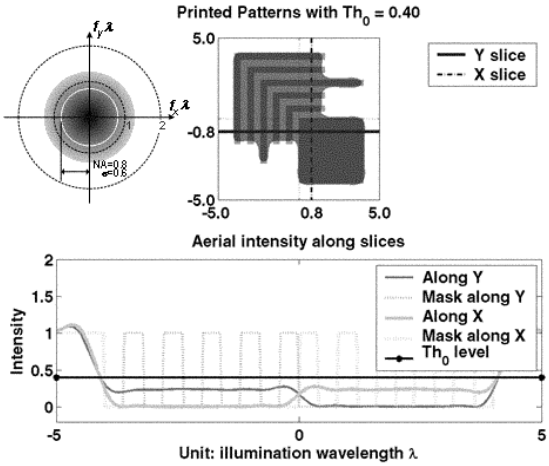
sired/target 2D patterns, we can pick the optimum energy ratio set for IIL and the Th_0 value for exposure energy. [**I like this idea, you will need to flesh it out more in another place.]**

A simulation example is given by setting $CD_{min} = 0.4\lambda$ with the mask in Figure 5(a), where we have set IIL optical system with NA_{on}=0.8, NA_{off_x}=NA_{off_y}=0.8 and NA_{oDC_x}= NA_{oDC_y}= 0.84 for simulations. After the optimization procedure, we obtained optimum values for energy ratios as: $(\alpha = 1.0, \gamma_x = 0.8, \gamma_y = 0.8, Th_0 = 0.7)$ without pupil filters and $(\alpha = 1.0, \gamma_x = 0.5, \gamma_y = 0.5, Th_0 = 0.7)$ with moon-shaped filters and the results of comparing IIL with partial coherent illumination are shown in Figure 6. Clearly the partial coherent illumination scheme is not able to resolve a sub-wavelength pattern but both IIL schemes can. Comparing two IIL schemes, even though Figure 6(c) IIL result with the hard cutoff pupil filters is better than Figure 6(b), the aerial intensity profile on both X and Y slices contain more ripples that could result in small holes in a large printed pattern due to the energy variations. To increase the energy latitude or contrast in IIL, depending on the actual mask in use, careful frequency parsing strategy should be made. **[again need more details on model, vector?, switch polarization x and y?….]**

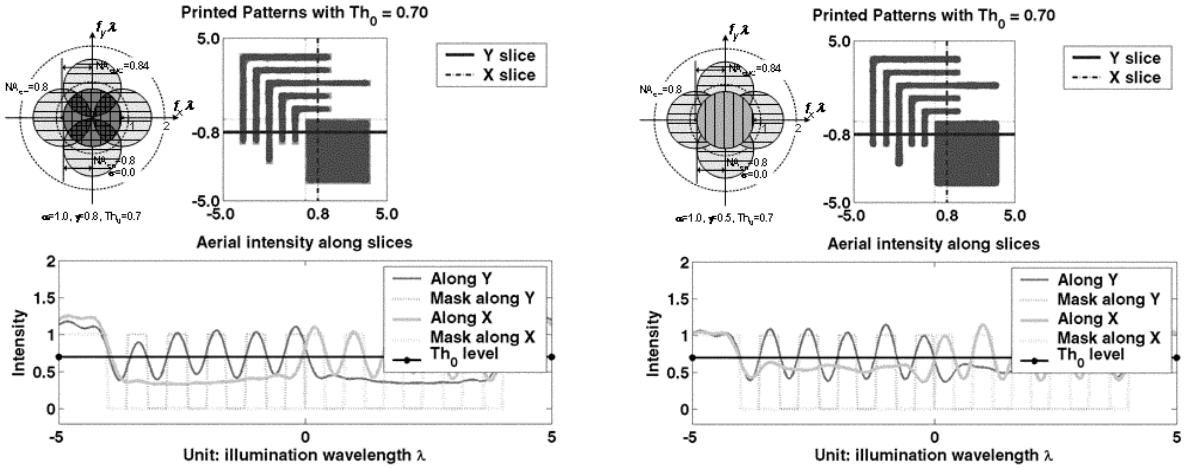
4. CONCLUSIONS AND FUTURE STUDY

The 1D and 2D grating analysis based on the IIL model with energy ratio parameters have been very useful to understand the behavior of IIL on-axis and off-axis exposures and shows the potential advantage that IIL can offer for nanotechnology. Besides illustrating that IIL can be treated as a useful resolution enhancement technique (RET) for sub-wavelength patterns, at this point we have learned how to keep a safe threshold margin with new parameter restrictions and observed the abrupt changes when one spatial frequency component moves from one exposure coverage to the other. To smooth those jumps in MEEF and OPE plots, we need to take further steps in the frequency parsing technique by finding the optimum pupil filters for IIL on-axis and off-axis exposures.

Beyond 1D CD measure analysis, the optimization strategy for IIL parameters in 2D mask case has included quantification of image quality based on overlay, LES and CR. By doing 2D block simulations first from the mask, the strategy showed reliable results with a given example. Currently we are focusing on more accurate IIL rigorous two-dimensional aerial image modeling and possible hybrid schemes such as advance pupil filter designs and the IIL-OPC technique.



(a) Partial coherent illumination scheme



(b) IIL scheme without pupil filters (c) IIL scheme with moon-shape pupil filters

Figure 6: Simulation results under different illumination schemes

ACKNOWLEDGEMENTS

This work is supported by ARO/MURI (grant number DAAD19-99-1-0196): Deep Subwavelength Optical Nanolithography.

REFERENCES

1. S. H. Zaidi and S. R. J. Brueck, "Multiple exposure interferometric lithography", *J. Vac. Sci. Technol.*, **B11**, pages 658-666, 1993.
2. S. R. J. Brueck, "Imaging interferometry - a novel approach to nm-scale optical lithography", *Microlithography World*, **7(1)**, pages 2-10, 1998.
3. X. Chen and S. R. J. Brueck, "Imaging interferometric lithography: a wavelength division multiplex approach to extending optical lithography", *J. Vac. Sci. Technol.*, **B16(6)**, pages 1-6, 1998 (Nov/Dec).
4. Xiaolan Chen, *A Study of interferometric lithography - approaching the linear system limits of optics,* PhD thesis, University of New Mexico, 1998 (Aug).
5. X. Chen and S. R. J. Brueck, "Experimental comparison of off-axis and imaging interferometric lithography", *J. Vac. Sci. Technol.*, **B17(3)**, pages 921-929, 1999 (May/Jun).
6. R. J. Collier, C. B. Burckhardt and L. H. Lin, *Optical Holography,* Academic Press, New York, 1971.
7. A. K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, 2001.
8. T. Brunner and R. Ferguson, "Simple models for resist processing effects", *Solid Sate Technol.*, pages 95-103, 1996 (June).
9. F. M. Schellenberg and C. Mack, "MEEF in theory and practice", *19th Annual BACUS Symposium on Photomask Technology,* Proc. SPIE 3873, Pages 189-202, 1999.
10. R. M. V. Bunau, H. Fukuda and T. Terasawa, "Effects of radially non-symmetric pupil filers and multiple-pupil exposure", *Optical Microlithography IX*, Proc. SPIE 2726, pages 375-385, 1996.

PROOF COPY 011401JMM

Simulation of the 45-nm half-pitch node with 193-nm immersion lithography—imaging interferometric lithography and dipole illumination

Abani Biswas
Steven R. J. Brueck
University of New Mexico
Center for High Technology Materials
Albuquerque, New Mexico 87106
E-mail: abiswas@chtm.unm.edu

Abstract. With immersion in water ($n=1.44$), the highest spatial frequency available with ArF-based (193-nm) lithography tools with a numerical aperture (NA) $= n \times \sin\theta$ of 1.3 (1.44×0.9) corresponds to a half-pitch of 37 nm. This suggests that the 45-nm half-pitch node should be accessible. A detailed vector simulation study is reported for two approaches to printing for this node. Both imaging interferometric lithography (IIL, with a single mask and multiple exposures incorporating pupil plane filters) and dipole illumination (with two masks separating the *x* and *y* oriented small features) are shown to be capable of printing arbitrary structures under these conditions. There is a substantial loss of contrast for TM polarization at this NA that demands that different polarizations be used to capture the high spatial frequencies in the *x* and *y* directions. Both dipole and IIL schemes offer this capability; IIL provides more robust imaging results. *© 2004 Society of Photo-Optical Instrumentation Engineers.* [DOI: 10.1117/1.1631007]

Subject terms: imaging interferometric lithography; dipole illumination; immersion lithography; 45-nm half-pitch; polarization effects in lithography.

Paper 014002 received May 27, 2003; revised manuscript received Jul. 23, 2003; accepted for publication Aug. 21, 2003.

PROOF COPY 011401JMM

1 Introduction

The relentless demands of Moore's law on lithography are well known. The current industry roadmap[1] calls for initial volume manufacturing of the 45-nm half-pitch node in the 2005 to 2007 time frame, which demands a decision on the manufacturing technology in the near future. It now appears that next-generation lithography techniques will be hard pressed to meet this time line, and attention has turned to extensions of optical lithography. Immersion lithography,[2–10] particularly using water as an immersion fluid at 193 nm, has attracted quite a bit of recent attention as a promising optical lithography extension.

The minimum half-pitch available to a lithography tool is given by the well-known equation,

$$\Lambda_{\text{half-min}} = \frac{\lambda}{4n \sin\theta}, \tag{1}$$

where λ is the vacuum exposure wavelength, and n is the refractive index of the immersion medium. The two beams propagate in the immersion medium at angles of $\pm\theta$ with respect to the wafer surface normal and interfere to create a sinusoidal intensity pattern. For 193-nm exposures with an immersion medium of water ($n=1.44$) and a maximum $\sin\theta$ of 0.9, this minimum half-pitch is 37 nm, suggesting that a system with these capabilities should be capable of printing 45-nm half-pitch patterns.

Of course, imaging is more complex than simply evaluating the minimum feature size. Much effort has gone into various resolution enhancement techniques (for a recent review of resolution enhancement techniques see Ref. 11) such as optical proximity correction (OPC), off-axis illumination (OAI), and phase-shift masks (PSMs). A purpose of this paper is to point out that at the very large propagation angles in the photoresist necessary to image patterns at the 45-nm node using a 193-nm source, the optical response becomes very polarization sensitive. This will require that the high spatial frequencies in the *x* and *y* directions be printed with a different polarization to retain the necessary contrast. This, in turn, will require some form of multiple exposure lithography to enable the necessary polarization control. Two imaging schemes that lend themselves to this constraint are imaging interferometric lithography[12,13] (IIL), which uses OAI at the limits of the pupil along with pupil-plane filters to control the image fidelity, and dipole illumination, where different masks are used for the high-spatial-frequency content in the *x* and *y* directions of a Manhattan geometry pattern. Both of these techniques offer the required independent control of the polarization for the *x* and *y* direction exposures.

Section 2 discusses the polarization effects in immersion lithography. Section 3 presents modeling results for partial coherent illumination, IIL, and dipole illumination for a simple nested L-bar test structure. Finally, conclusions and future research directions are outlined in Sec. 4.

2 Polarization Effects in Immersion Lithography

Just as evaluating the highest spatial frequency available with an exposure tool is most straightforward in the case of a periodic grating pattern, polarization effects are most readily understood in terms of a simple interferometric li-

PROOF COPY 011401JMM

PROOF COPY 011401JMM

thography (IL) exposure with two beams symmetrically incident on the wafer.[14] It is straightforward to calculate the interference intensity pattern for two-beam IL with equal intensity beams with for TE and TM polarizations. The results are

$$I_{\mathrm{TE}}(x) = 2I_0\left[1+\cos\left(\frac{4\pi n_{\mathrm{resist}}\sin\theta_{\mathrm{resist}}}{\lambda}x\right)\right],$$
$$I_{\mathrm{TM}}(x) = 2I_0\left[1+\cos(2\theta_{\mathrm{resist}})\cos\left(\frac{4\pi n_{\mathrm{resist}}\sin\theta_{\mathrm{resist}}}{\lambda}x\right)\right]. \quad (2)$$

Here, I_0 is the intensity of each beam and the subscripts emphasize that the refractive index and internal angles within the resist are used in evaluating these expressions. The TE result retains full contrast over the entire accessible half-pitch range independent of θ_{resist}. The extra factor of $\cos(2\theta_{\mathrm{resist}})$ that appears in the TM expression has dramatic consequences for lithography. For $\theta_{\mathrm{resist}}=45$ deg [numerical aperture (NA)=1.21 at a photoresist refractive index of 1.71], the contrast of the interference pattern goes to zero, and at higher angles there is a 180-deg phase shift so that lines and spaces are interchanged. This reduction in contrast and phase reversal arises because both the *x* directed and the *z* directed fields interfere independently for the TM polarization. There is a 90-deg phase shift between the two contributions, with the *x* directed fields dominating at small angles ($\theta_{\mathrm{resist}}<45$ deg) and the *z* directed fields at high angles ($\theta_{\mathrm{resist}}>45$ deg).

Even at the high NAs of current nonimmersion lithography tools, polarization effects have yet to be of great significance for imaging. Unpolarized or circularly polarized illumination, which averages the response of the two possible polarizations, has been adequate for most imaging. This is because the immersion fluid today is air ($n=1$), while the index of a typical resist is considerably higher ($n_{\mathrm{resist}}\sim 1.71$). As a consequence of Snell's law, the propagation angles in the resist are significantly smaller than in the air. For a period corresponding to the 90-nm half-pitch node, $\sin\theta\sim 0.54$ the propagation angle in air is 32 deg, while θ_{resist} is only 18 deg and the corresponding contrast for TM polarization is 84%. However, for the 45-nm half-pitch node, the propagation angle of 48 deg in the immersion fluid corresponds to $\theta_{\mathrm{resist}}\sim 39$ deg, where the contrast for TM polarization is reduced to 21%.

This is illustrated in Fig. 1(a), which shows the intensity distribution versus position in the resist for periods corresponding to the 90-, 65- and 45-nm half-pitch nodes. The top trace is for TE polarization, which retains its contrast independent of the pitch. The TM curves for the 90-, 65-, and 45-nm half-pitch nodes at 193 nm are also shown. Figure 1(b) shows the pitch dependence of the TM visibility,

$$\text{fringe visibility} \equiv (I_{\max}-I_{\min})/(I_{\max}+I_{\min}). \quad (3)$$

The dashed vertical lines correspond to the center spatial frequencies for the three upcoming half-pitch nodes. The shaded region (half-pitch <54 nm) is accessible only with immersion. Plotted against half-pitch, the curve is universal, dependent only on the optical wavelength and independent of the immersion fluid optical properties. The cusp at a half-pitch of 40 nm corresponds to $\theta_{\mathrm{resist}}=45$ deg and to the phase reversal region for smaller half-pitches. For the center frequency corresponding to the 90-nm node, the TM visibility is 84%, dropping to 62% for the 65-nm node (NA =0.74) and 21% for the 45-nm node (NA=1.07). The low TM contrast at the 45-nm node and the proximity to the phase reversal region implies that any lithography technique will be required to use TE polarization exclusively for the highest spatial frequencies in the image. The developed photoresist line-edge roughness is strongly dependent on the visibility[15] so that this loss of contrast impacts the quality of the image in terms of both resolution and line-edge quality.

The immersion fluid also impacts the reflection at the immersion medium-photoresist interface. Figure 2 shows

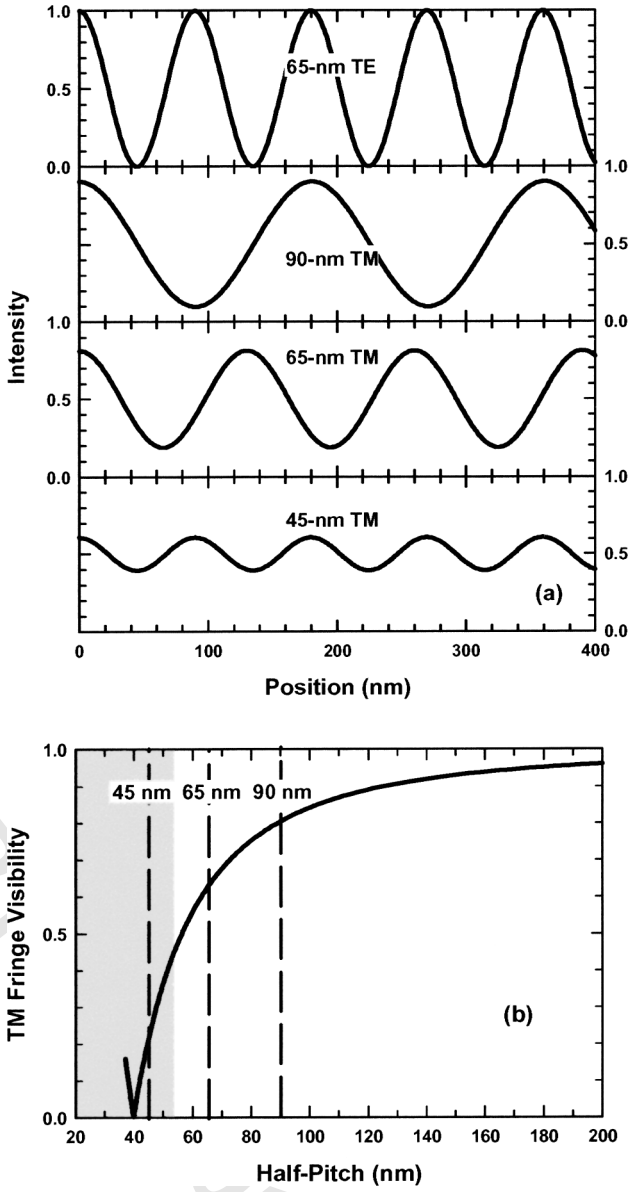


Fig. 1 (a) Intensity profiles in resist for 65-nm half-pitch TE polarization (top) with full contrast, and for 90-, 65-, and 45-nm half-pitch patterns with TM polarization. The decrease of contrast is severe. (b) Visibility of the TM polarization pattern in the resist as a function of half-pitch. The shaded region is accessible only with immersion.

PROOF COPY 011401JMM

PROOF COPY 011401JMM

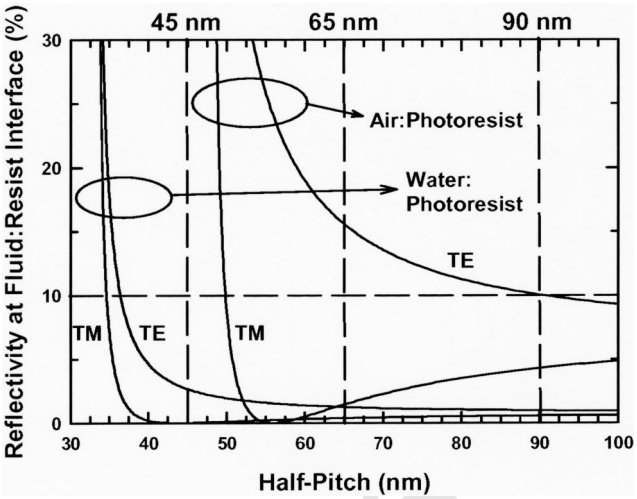


Fig. 2 Reflectivity at the fluid:photoresist interface for both TE and TM polarizations as a function of half-pitch. The dashed horizontal line corresponds to a 10% reflectivity. Water immersion both enables a smaller half-pitch and reduces the reflectivity to negligible levels over the entire range of spatial frequencies required for the 45-nm half-pitch node.

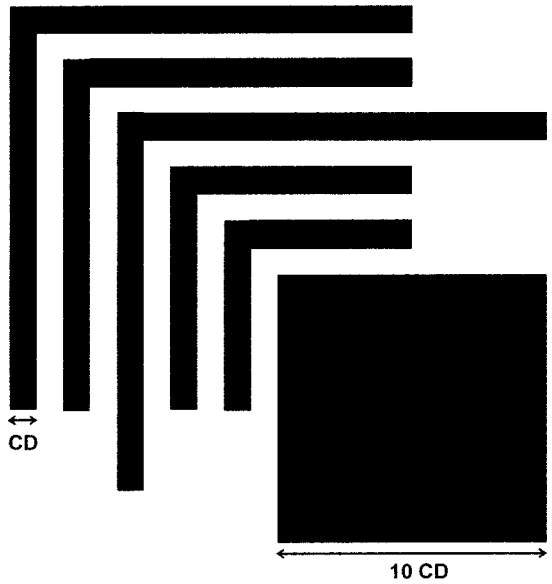


Fig. 3 Test pattern used for the simulations.

the calculated reflection for both TE and TM polarizations for the two cases of air ($n=1$) and water ($n=1.44$) as the immersion fluid, with a photoresist refractive index of 1.71. Water immersion both enables a smaller half-pitch and reduces the reflectivity to negligible levels for the entire range of spatial frequencies required for the 45-nm half-pitch node. As expected, the reflectivity for both polarizations approaches unity at grazing incidence. The unity transmission (0% reflection) into the resist at Brewster's angle for TM polarization is evident in both TM curves.

3 Immersion Lithography Simulations for Arbitrary Patterns

All of the simulations presented in the following are for the pattern shown in Fig. 3 consisting of a series of five nested L-bars at a 45-nm half-pitch CD and a large box at 10×CD. For the nested features, the middle feature extends beyond the others (a different amount in the x and y directions to further break symmetry) to give some indication of dense isolated biases. For partial coherent illumination and for imaging interferometric lithography, this mask is used without any OPC or other changes. For dipole illumination, the features are split into two separate masks, as described next. Extension of this modeling to a variety of additional structures of relevance to semiconductor manufacturing will be the subject of future work.

Simulations were carried out with Prolith™ Version 7.2.2 (KLA-Tencor) in full vector electromagnetic mode and with the full physical model for the photoresist (PEB diffusion, 3-D etching). In all cases, the simulations used the same film stack. Substrates were bare silicon with a 30-nm-thick bottom antireflection coating (BARC) ($n=1.73+i0.395$) and a 50-nm-thick, positive photoresist ($n=1.7135+i0.0169$; dill parameters $A=0$, $B=1.1\,\mu\mathrm{m}^{-1}$, $C=0.0455\,\mathrm{cm}^{-2}/\mathrm{mJ}$). Standard postexposure bake (125°C for 90 s) and develop conditions (60 s) were used for all simulations.

Two techniques were used to force the calculations to accommodate the important immersion physics. First, an adaptation to immersion was employed—the wavelength was reduced by the immersion fluid index (193/1.44=134 nm), the NA was replaced by $\mathrm{NA}/n_{\mathrm{immersion}}$, and all of the indices of the materials in the film stack (including the Si) were replaced by $n_{\mathrm{film}}/n_{\mathrm{immersion}}$. As the previous section indicated, separating the exposures of the high spatial frequencies in the x and y directions and using orthogonal polarizations for the two cases, so that all exposures are dominantly TE polarization, will be key to imaging at the limits of immersion lithography. The simulations presented next bear this out. The present version of Prolith does not allow this flexibility. The approach used was to carry out the exposures individually, with the correct polarizations, up to the aerial image, extract and combine the aerial images external to Prolith, and then reinsert the composite aerial image and allow Prolith to carry out the exposure and resist develop simulations.

In the following sections, simulation results for three approaches to imaging the 45-nm half-pitch node at 193 nm using water immersion are presented. First is traditional partial-coherent illumination. Second is IIL, essentially off-axis illumination at the edge of the pupil with a small partial coherence, along with imaging pupil-plane filters to ensure uniform weighting of all spatial frequencies. Finally, imaging with dipole illumination (DI)—again at the edges of the pupil with a small partial coherence—with two masks that separate out the high x and high y spatial frequency parts of the image is evaluated. In both the IIL and DI cases, the polarization was rotated between exposures to ensure predominately TE polarization exposures for the high spatial frequencies. OPC was not explored even though it can be used to improve the results as it is important to understand the baseline, non-OPC, case before adding the complications and multiple variations possible to improve the images. To bound the content of this paper, PSMs were not explored since both IIL and DI capture the highest spatial frequencies possible with a given optical

PROOF COPY 011401JMM

PROOF COPY 011401JMM

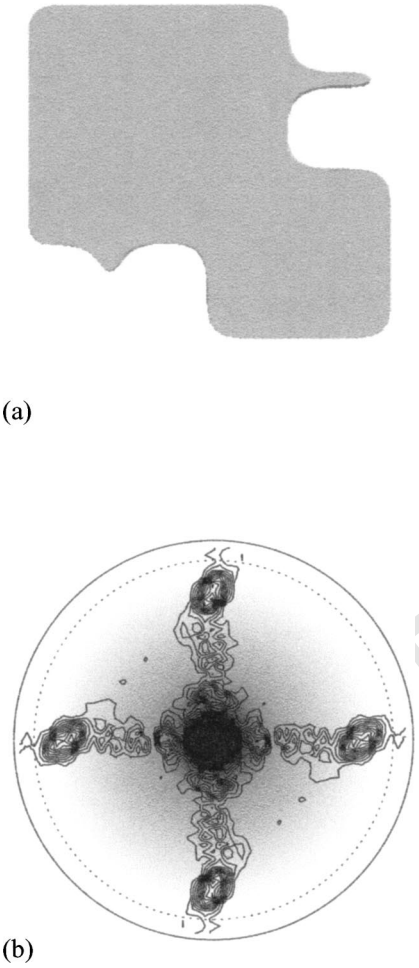


Fig. 4 Partial coherent imaging result for $\sigma=0.7$: (a) the top view of the real-space Prolith simulation and (b) the corresponding frequency space picture. The outer circle is the limiting spatial frequency with a radius of $2n/\lambda$. The inner dashed circle is the maximum spatial frequency coverage of the tool with a radius of $(2n \sin\theta/\lambda)$. The shading inside this circle is an approximate representation of the modulation transfer function, which decreases from the center of the circle to the edge. The contour plot is the intensity Fourier transform of the mask. The pattern is not resolved because of the loss of contrast at the high spatial frequencies characteristic of partial coherent imaging and the poor TM polarization response at high angles.

system and hence are sufficient to show that the 45-nm node is accessible for immersion lithography at 193 nm. PSM also enables printing of these high spatial frequencies and it will be important to investigate its applicability in future work.

3.1 *Partial Coherent Illumination*

Figure 4(a) shows the top-down view of the developed image for a partial coherence (ratio of the illumination pupil NA to the imaging pupil NA) of $\sigma=0.7$, circularly polarized illumination using a bright field mask and positive resist. Clearly the dense line/space structures of the nested L-bars are not captured in this image, which rather shows only the general outlines of the entire pattern. This is true across the entire available range of possible partial coherence $(0<\sigma<\infty)$. The reason for this is clear with reference to Fig. 4(b), showing the frequency space situation. The outer circle in Fig. 4(b) represents the frequency space limit of the immersion medium with a radius of $2n/\lambda$. This is the maximum spatial frequency that can be imaged through the immersion medium, corresponding to counterpropagating plane waves at grazing incidence to the wafer. The inner dotted circle represents the maximum frequency allowed by the optical system with $\sin\theta=0.9$ and with radius $2n \sin\theta/\lambda$. The shaded region in the center of these circles is a rough representation of the frequency space coverage associated with partial coherent imaging. Darker regions correspond to a better transfer of the frequency components from the mask to the image (effectively a higher modulation transfer function). With $\sigma=0.7$ and NA$=1.3$, the maximum frequency response extends to $(1+\sigma)\text{NA}/\lambda = 1.53n/\lambda$ compared to the half-pitch center frequency of $1/2\,\text{CD}\sim 1.49n/\lambda$. Finally, the contour plot in the figure is the magnitude of the Fourier transform of the mask, scaled to the imaging conditions. Corresponding to the Manhattan (x,y) geometry of the pattern, the spectral content is dominantly along the f_x and f_y directions. There are five regions of significant frequency content. One is around the origin, corresponding to the low-frequency content of the large feature and the overall shape of the pattern. There are intense frequency components centered about the $(\pm)$ frequencies corresponding to the pitch of the dense lines and spaces in both the f_x and f_y directions. Partial coherent illumination has an effective modulation transfer function that is highest at low frequencies and approaches zero at the highest frequencies. The maximum frequency response for $\sigma=0.7$, with a low effective transfer function, is roughly at the center frequency corresponding to the lines and spaces of the 45-nm half-pitch node. This results in an underweighting of the high-frequency image components and the lack of resolution for the individual lines and spaces. The polarization effects noted in the previous section additionally degrade the partial coherence image since half of the exposure intensity for the high frequencies is in the TM or ''wrong'' polarization, which has very poor contrast and further degrades the transfer function.

3.2 *IIL*

IIL was introduced[12,13] as a technique to cover the extremes of frequency space by tilting the illumination as far as possible off-axis. With the modest NAs of $\sim$0.5 to 0.65 that were available at the time, this required tilts that were beyond the limits of the imaging pupil and consequently reintroduction of a zero-order beam on the image side of the lens. With today's very high NA lenses of $\sin\theta\sim 0.9$, the gain for the significant added optical complexity of building interferometers around the lens is small and it is sufficient to set the offset to the largest value allowed by the lens NA. Thus, IIL is similar to quadrupole off-axis illumination with a small partial coherence for each off-axis illumination beam, and with the quadrupole oriented along the principal (x,y) axes of a Manhattan geometry pattern to maximize the spatial frequency capture. The polarization of each of the off-axis illumination beams can be independently controlled to optimize the imaging. Quadrupole off-axis illumination inherently emphasizes low frequencies at the expense of high frequencies[13,16] because each of the

PROOF COPY 011401JMM

PROOF COPY 011401JMM

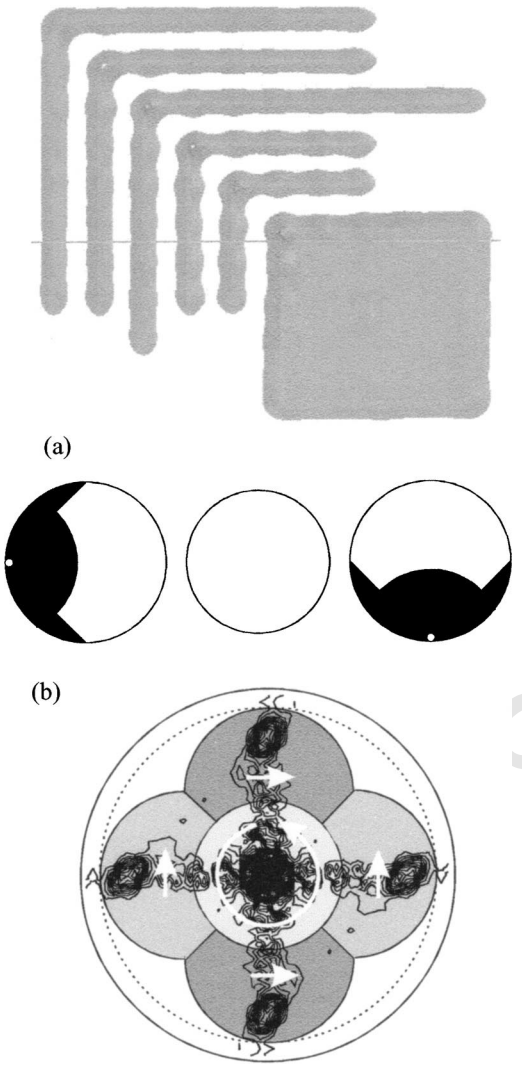


Fig. 5 IIL modeling: (a) the final real-space image, where the horizontal line denotes the position at which the intensity profile shown later in Fig. 12 is taken; (b) the pupil filters used in the off-axis exposures; and (c) the resulting frequency-space coverage. The gray shadings simply denote the frequency coverage, the effective modulation transfer function (MTF) for the linear imaging terms is uniform across the image. The polarization of the individual exposures is indicated.

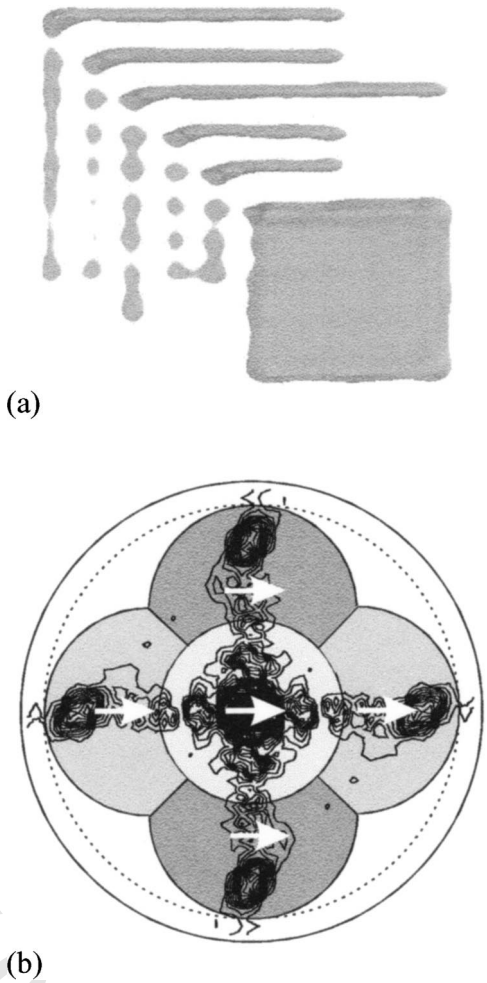


Fig. 6 Exposures identical to Fig. 5 with the exception of the polarization. The impact of using the "wrong" polarization for the high spatial frequencies is clear.

exposures covers the same low-frequency information, while the individual high frequencies are covered only in selected exposures. Pupil plane filters can be used to eliminate these multiple coverages and thereby provide an improved image, close to the ultimate frequency-space limits. Thus, IIL for NA$\sim$1.3 becomes quadrupole OAI with low σ and tilts to the edges of the imaging pupil along with pupil-plane filters to eliminate the overrepresentation of low spatial frequencies.

Figure 5 shows the result of this procedure for the 45-nm half-pitch pattern. The pattern [Fig. 5(a)] is well resolved. The impacts of the limited resolution include rounding of the line edges and some line edge ripple that results from the coherent exposures. Both of these issues could be addressed with the use of modest OPC on the mask. The horizontal line denotes the position at which the intensity profiles discussed in Sec. 3.3 (Fig. 12) are taken.

The pupil plane filters used in this simulation are shown in Fig. 5(b). The low-frequency (center filter) exposure is carried out at a lower NA of 1.15. The offset exposures (left and right filters) are at the full NA of 1.3 with blocked regions to avoid overlaps in frequency coverage between the various exposures. Coherent illumination is assumed ($\sigma\sim0.02$) for all of the exposures. Two offset exposures from opposite directions are included in the simulation for both the x and y offset directions to restore a telecentric response. This has no impact at focus, but is important for the depth-of-field evaluation discussed later. The polarization is controlled in each of the exposures, as indicated in Fig. 5(c) to ensure a dominant TE polarization, especially for the highest spatial frequencies. Figure 5(c) shows the resulting spatial frequency coverage with the intensity Fourier transform of the image superimposed. The different gray levels in this figure simply distinguish individual exposure coverages. The effective MTF for the linear imaging terms is uniform across the entire accessible frequency space.[13]

The impact of polarization control is shown in Fig. 6. This result is for the same pupil filters as for the result of Fig. 5, the only difference is that a fixed polarization was used for all exposures. The impact of having the ''wrong''

COPY UNDER

PROOF COPY 011401JMM

PROOF COPY 011401JMM

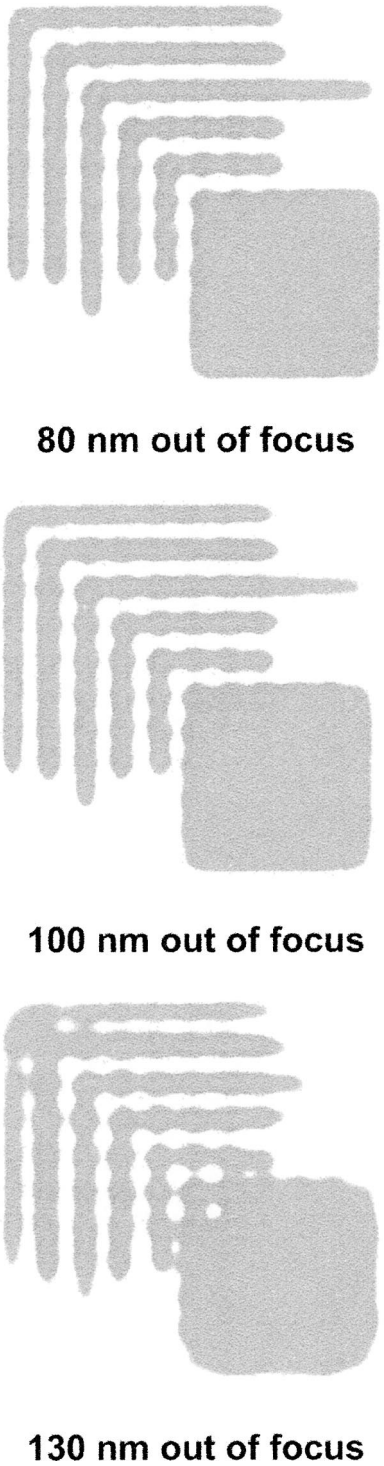


Fig. 7 Evolution of the IIL image of Fig. 5 though focus. The depth of field is of order 200 nm.

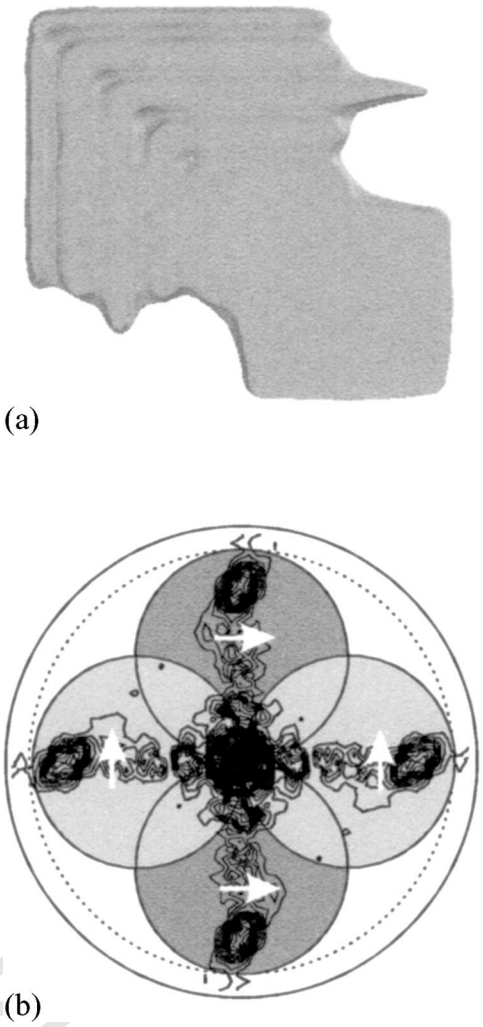


Fig. 8 (a) Result of IIL exposures without pupil plane filters, where the dense features are not resolved, and (b) the corresponding frequency space coverage. The multiple coverage of the low-frequency regions leads to the lack of resolution.

(e.g., TM) polarization for the high spatial frequencies in the f_x direction is clear. Symmetric results were obtained if y directed polarization was used for all exposures, the opposite parts of the L-bar shapes did not print. This result confirms the simple IL visibility results already given—it will be critical to independently optimize the polarization for the high spatial frequencies in orthogonal spatial directions involved in printing 45-nm node patterns by immersion with a 193-nm source.

The impact of defocus on the IIL pattern is shown in Fig. 7. At an 80-nm defocus, the pattern is maintained. At a 100-nm defocus, the dense lines and spaces are largely intact but dense/isolated bias is evident in the line-end shortening of the isolated features. Finally, at a 130-nm defocus, the lines begin to merge together and a hole has developed in the large box. The addition of OPC should extend the depth of field (DOF), which is ~200 nm from these simulations. It is important to recognize that the large offsets of IIL (or DL, later) maximize the DOF compared with conventional on-axis illumination for these high-spatial-frequency structures. In the limit of a periodic structure where the off-axis illumination is chosen to symmetrize the zero- and first-order diffracted beams on the image pupil, i.e., interferometric lithography, the DOF approaches infinity since both beams are inclined at the same azimuthal angle (θ). The frequently cited Rayleigh result for incoherent illumination, $\mathrm{DOF} \propto \kappa_2(\lambda/\mathrm{NA}^2)$, does not hold for this situation.

PROOF COPY 011401JMM

PROOF COPY 011401JMM

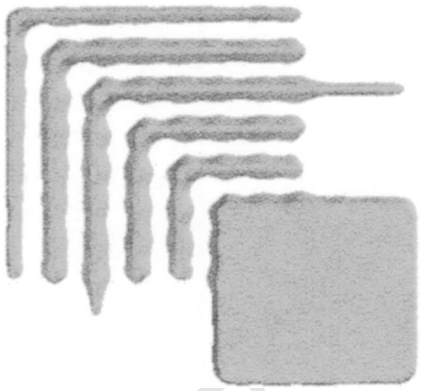


Fig. 9 IIL with a larger σ=0.04. There is an increased dense-isolated line bias and poorer definition of the L-bar corners at this higher σ.

The importance of including the pupil-plane filters is shown by the simulation result in Fig. 8, which includes only the offset exposures but without incorporating any pupil plane filters—e.g., quadrupole illumination with the quadrupole oriented along the principal pattern axes and with optimized polarizations. The dense features are incompletely resolved and there is a significant dense-isolated bias. The frequency space coverage in Fig. 8(b) shows that the low-frequency components of the pattern are counted multiple times, while the high-frequency components are counted only once. The resulting overemphasis on the low frequencies is the reason for the inability to resolve the dense features. This exposure strategy is equivalent to DL, but with a single mask with all of the features included. The next section shows that somewhat better results are available by parsing the features between two masks and employing a double-exposure strategy.

Finally, the impact of a slightly larger σ of 0.04 is shown in Fig. 9. This corresponds to the incoherent addition of coherent exposures at different offsets. The limits were set by the maximum offset at the limit of the pupil $(0.9n/\lambda)$ at one side, and by the smallest offset that can capture the highest spatial frequencies in the image, which from the Fourier plots in Fig. 8 extend out to almost $1.8n/\lambda$. Empirically, this restricted the σ to $\sim$0.04 with a center offset of $0.88n/\lambda$. The pupil plane filters, with the small hole for the zero-frequency diffraction enlarged for the larger σ, and polarization were the same as used in the exposure of Fig. 5. The image is significantly degraded with a much more significant dense-isolated line bias, a narrowing of the outermost L-bar and more exaggerated sidewall ripples, particularly in the corners. All of these imaging artifacts could be addressed with OPC. Extension to higher σ's results in poorer images as the lower offsets do not enable capture of the necessary high spatial frequencies and thereby overemphasize the larger features.

3.3 *DL*

DL, where the pattern features are split into two masks each with the high spatial frequencies in only one direction, is an alternative exposure technique that offers the necessary polarization flexibility. For these simulations, no pupil plane filters were used and all exposures were at the full system NA. The wide range of spatial frequencies in the present test pattern poses some difficulties for this approach. Figure 10 shows an example of the result of coherent ($\sigma\sim$0.02) DL at an offset of 0.89 just as was used for the IIL example of Fig. 8. Each mask is printed with the appropriate dipole illumination and polarization. The large box has been removed from the pattern and the remaining dense features split into separate masks for the x and y directions, as shown. The polarization is optimized separately for exposure of each mask. The dense-isolated bias is more pronounced in this configuration, and the corners of

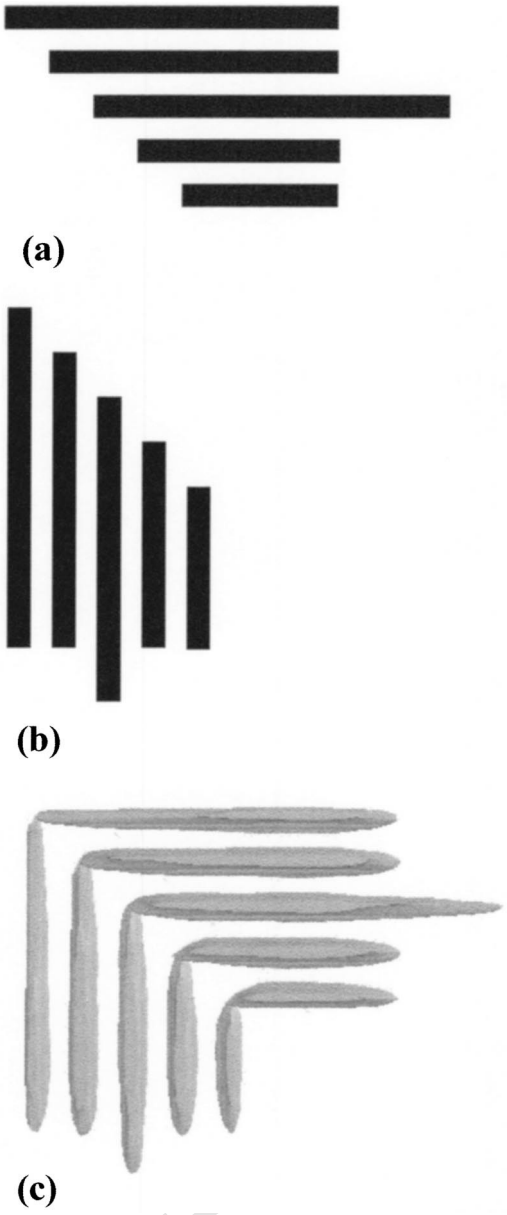


Fig. 10 DL with two masks (a) and (b) separating the high small features in the x and y directions. The large box was eliminated, since it was not possible to print the pattern when it was present. The illumination offsets were at the edge of the imaging pupil and the frequency space coverage was identical to that shown in Fig. 8. Because the large box was absent, the frequency space overlaps were tolerable.

PROOF COPY 011401JMM

PROOF COPY 011401JMM

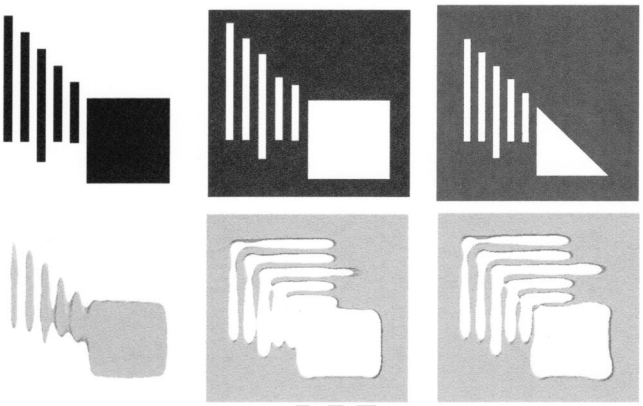


Fig. 11 Results of three different dipole mask strategies. A single bright-field mask printing only the lines parallel to the *y* axis and the large box. It was not possible to print both the lines and the box in a single exposure. Also shown are two different attempts at dark-field (inverted) images. The second mask for the first case (full large box) included only the lines parallel to the *x* axis with no box. The second mask for the second case (triangular box segment) was symmetric to the mask shown.

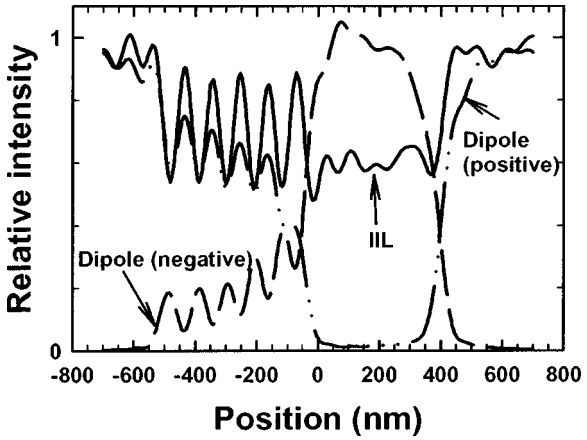


Fig. 12 Intensity of the aerial image across a cut of the pattern for IIL, a single dipole exposure with a positive tone mask (Fig. 11, left column) and a single dipole exposure with a negative tone mask (Fig. 11, middle column).

the L-bars are less robustly defined. Alignment of the two masks between exposures is an added requirement for this dipole technique.

Figure 11 shows the result of a single-mask exposure including the large box (left column) with the appropriate DL and polarization. The individual lines close to the box are not resolved. The two right columns show two different dark-field (inverted image) mask strategies. In the center column, the mask includes the entire box, while the second mask (not shown) has only the dense features parallel to the *x* axis. Again in this case, there is difficulty in separating the dense lines from the box, and there is significant distortion of the pattern. The right-hand column shows a more symmetric division of the mask. This produced the best-printed pattern, but still shows significant distortion and linewidth variation. More detailed investigations are necessary to optimize these DL strategies. In particular, the use of OPC with or without pupil-plane filters will certainly provide better imaging results.

Figure 12 shows intensity cuts across the aerial image for IIL and for single dipole exposures for both the bright-field and dark-field cases for the full pattern including the large box. Clearly IIL provides the best aerial image and hence the best developed image. In both dipole examples, because of the large low-frequency content of the aerial images, it is impossible to choose an exposure threshold level for either of the dipole exposures that captures all of the pattern features.

4 Conclusions

A simulation of arbitrary pattern lithography for the 45-nm half-pitch node using a 193-nm source with an NA=1.3 water-immersion lithography tool was presented. Of particular importance is the requirement of optimizing the polarization individually for different parts of the image. This will require a major change to illumination optics and will impact exposure strategies and tool cost-of-ownership evaluations.

To access the highest spatial frequencies—necessary for printing 45-nm half-pitch structures—off-axis illumination near the edges of the imaging pupil will be required. Two strategies for separating the exposures to allow for the necessary polarization control were presented. IIL, related to off-axis illumination with the addition of pupil plane filters to adjust the details of the image, provides relatively robust results. Different pupil plane filter strategies were presented along with details of variation of the image through focus. DL with the division of the pattern into two masks, each with patterns primarily oriented in orthogonal directions, was also investigated. Initial results are not as promising as the IIL strategy, but more investigation is required.

These simulations show that 45-nm half-pitch lithography is indeed possible by extension of 193-nm tools to water immersion. Much more investigation is required to optimize the imaging over the range of patterns required for actual circuits, and to add OPC to provide more robust results. This evaluation did not consider PSMs, which are an alternative approach to reaching the frequency space limits of optical systems. Experimental confirmation of the simulations using microstepper platforms is also an urgent need. Interferometric lithography (grating) investigations are underway and are reported elsewhere[10] and in this volume. These reports confirm the frequency capabilities and the IL polarization analysis and demonstrate that work is necessary on resists to extend their resolution to the 45-nm half-pitch node. Ultimately, decisions as to manufacturing direction are as much economic as technical. Significant further investigation is necessary to evaluate all of the possible directions; this work is but an initial contribution.

Acknowledgment

Support for this work was provided as part of the ARO/MURI in deep subwavelength optical nanolithography.

References

1. http://public.itrs.net.
2. C. V. Shank and R. V. Schmidt, "Optical technique for producing 0.1-μm periodic surface structures," *Appl. Phys. Lett.* **23**, 154 (1974).
3. W. T. Tsang and S. Wang, "Simultaneous exposure and development

PROOF COPY 011401JMM

PROOF COPY 011401JMM

technique for making gratings on positive photoresist,'' *Appl. Phys. Lett.* **24**, 196 (1974).
4. M. D. Feuer and D. E. Prober, ''Projection photolithography-liftoff techniques for production of 0.2-mm metal patterns,'' *IEEE Trans. Electron Devices* **ED-28**, 1375 (1981).
5. H. Kawata, J. M. Carter, A. Yen, and H. I. Smith, ''Optical projection lithography using lenses with numerical apertures greater than unity,'' *Microelectron. Eng.* **9**, 31–36 (1989).
6. H. Kawata, I. Matsumura, H. Yoshida, and K. Murata, ''Fabrication of 0.2 μm fine patterns using optical projection lithography with an oil immersion lens,'' *Jpn. J. Appl. Phys.* **31**, 4174–4177 (1992).
7. G. Owen, R. F. W. Pease, D. A. Markle, A. Grenville, R. L. Hsieh, R. von Bünau, and N. I. Maluf, ''1/8 μm optical lithography,'' *J. Vac. Sci. Technol. B* **10**, 3032 (1992).
8. J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, ''Liquid immersion deep ultraviolet interferometric lithography,'' *J. Vac. Sci. Technol. B* **17**, 3306 (1999).
9. M. Switkes and M. Rothschild, ''Immersion lithography at 157 nm,'' *J. Vac. Sci. Technol. B* **19**, 2353 (2001).
10. A. K. Raub and S. R. J. Brueck, ''Deep UV immersion interferometric lithography,'' in *Optical Microlithography XVI*, A. Yen, Ed., *Proc. SPIE* **5040**, 667–678 (2003).
11. A. K.-K. Wong, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press, Bellingham, WA (2001).
12. X. Chen and S. R. J. Brueck, ''Imaging interferometric lithography—approaching the resolution limits of optics,'' *Opt. Lett.* **24**, 124–126 (1999).
13. S. R. J. Brueck and X. Chen, ''Spatial frequency analysis of optical lithography resolution enhancement techniques,'' *J. Vac. Sci. Technol. B* **17**, 908–921 (1999).
14. D. G. Flagello and T. D. Milster, ''High-numerical-aperture effects in photoresist,'' *Appl. Opt.* **36**, 8944–8951 (1997).
15. W. Hinsberg, F. A. Houle, J. Hoffnagle, M. Sanchez, G. Wallraff, M. Morrison, and S. Frank, ''Deep-ultraviolet interferometric lithography as a tool for assessment of chemically amplified photoresist performance,'' *J. Vac. Sci. Technol. B* **16**, 3689 (1998).
16. Xiaolan Chen and S. R. J. Brueck, ''Experimental comparison of off-axis illumination and imaging interferometric lithography,'' *J. Vac. Sci. Technol. B* **17**, 921–929 (1999).

Abani Biswas received his PhD degree in physics from the Indian Institute of Technology, where he studied the fabrication and characterization of rare-earth-doped silicate glasses and fibers. Prior to joining the Center for High Technology Materials, he was a postdoctoral fellow with the State University of New York, Buffalo. His current research interests include nonlinear effects in glasses and optical lithography.



Steven R. J. Brueck received the BS degree from Columbia University in 1965 and his SM and PhD degrees from the Massachusetts Institute of Technology (MIT) in 1967 and 1971, all in electrical engineering. From 1971 to 1985 he was a staff member at MIT Lincoln Laboratory. In 1985 he moved to the University of New Mexico, where he is currently a professor in the EECE and Physics and Astronomy Departments and directs the Center for High Technology Materials. His current research interests include nanoscale lithography, the physics of nanostructures, the nonlinear optics of poled glasses, and semiconductor laser physics.

PROOF COPY 011401JMM

PROOF COPY 011401JMM

Optimization and apodization of aerial images at high NA for imaging interferometric lithography

Thanis M. Tridhavee[a], Balu Santhanam[a], and Steven R. J. Brueck[ba,c]

[a] Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, 87106 USA;
[b] Center of High Technology Materials, University of New Mexico, Albuquerque, 87106 USA;
[c] Department of Physics and Astronomy, University of New Mexico, Albuquerque, 87106 USA

ABSTRACT

Imaging Interferometric Lithography (IIL) offers several optimization parameters such as pupil filtering, parsing of frequency coverage, polarization control, and multiple exposure dosage ratios. We discuss the optimal frequency coverages for IIL at NA = .9 and the effects of the dark field (quadratic image autocorrelation terms) on the aerial image under pupil filtering. Next, comparisons are made of exposure latitudes for various dosage ratios and exposures for several weighted errors under different conditions. Lastly, apodization of the pupil filters is considered and shown to alleviate artifacts associated with Gibbs phenomena at hard frequency stops and improve overall image fidelity.

Keywords: optimization, pupil filtering, image quality, aerial image, imaging interferometric lithography, apodization

1. INTRODUCTION

As κ_1 decreases in optical lithography, a detailed understanding of nonlinear imaging effects becomes necessary to optimize image formation at minimum dimensions. For any lithographic technique, high contrast and exposure latitude are necessary to achieve successful printing of high quality patterns, especially in a high volume manufacturing environment. IIL at high NA [1] offers several *resolution enhancement techniques* (RET) such as *off-axis illumination* (OAI) at the largest oblique illumination angle to down-shift the highest frequency components into the pass-band of the optical system, variable pupil filtering strategies to improve image robustness, and polarization control to maximize interference and print high spatial frequencies at the optical system minimum half-pitch. This paper explores some of these strategies and demonstrates their impacts on the final images. With the IIL scalar model [2,3], the up-shifted high frequency components of a mask pattern retain the high amplitudes necessary to create high exposure latitude for high resolution patterns at low κ_1 imaging. At high NA, four exposure IIL is similar to double dipole off-axis illumination RET, however the primary differences are the use of coherent light, on-axis exposure, and pupil filters; all in combination to create high resolution images with little image degradation.

Since, numerical apertures (*NA*) will approach $\gtrsim 0.9$ in air and ~ 1.3 or beyond in immersion lithography, IIL uses off-axis and on-axis illumination, polarization control for proper interference at high NA angles for each exposure, pupil filtering to the control iso/dense bias, and multiple exposure ratio dosage parameters to control contrast/exposure latitude. These variables all allow IIL to fully exploit all parameters of multiple exposure optical lithography. IIL promises full spatial optical resolution without costly phase-shift masks and with minimal optical proximity correction for small critical dimensions.

In this paper, we compare the aerial images at high NA of four and five exposure IIL under scalar and vector models. We explore variations of dosage ratios between the reference beam and the first diffraction order using various simulation approaches. Optimized ratio dosages for the multiple exposures have a strong effect on printability, exposure latitude, and defects. In addition, use of pupil filters and apodization can improve resolution and combat artifacts yielding a maximally robust aerial image.

2. OPTIMIZING FREQUENCY COVERAGE AND PARSING

IIL provides the ability to use the parsing of different frequency space image components into separate exposures as a parameter for image quality and resolution. The tiling or parsing is not unique due to the ability to vary the NA,

SPIE 5377 p. 1544 - 1554 (2004).

polarization, and offset illumination angles in separate exposures. To better understand the impacts of these variables on the frequency coverages and parsings, numerical algorithms were created to optimize the illumination scheme for a given mask [4]. Afterwards, an understanding is gained which coverages are most likely to work well for a given NA and mask pattern. At high NA's, the vector model is necessary due to the large propagation angles. However, with TE polarization control for each exposure, the aerial images will have similar aerial image profiles when compared to scalar aerial images in simulation [5]. In general, the *trends* of the aerial image should still hold due to the Manhattan pattern assumption, use of pupil filters, and use of polarization control. The scalar model is used in this paper with care for simplicity and to understand the general trends of varying parameters of IIL. Where necessary a vector calculation is used to validate the results. The IIL intensity is described by the sum of multiple incoherent intensities [2]

$$I = \alpha I_{on} + I_{off-x} + I_{off-y}$$

where α is the ratio of the dosages between the I_{on} (on-axis intensity) and I_{off} (off-axis intensity) for the horizontal and vertical frequency components. When the off-axis illumination point source is at edge of the illumination pupil for the horizontal frequencies the off-axis intensity is described by

$$I_{off-x} = \underbrace{\gamma^2}_{DC} + \gamma \underbrace{\iint_{f_x,f_y} \left(O(f_x,f_y)P(f_x+f_{off-x},f_y)e^{i2\pi(f_x x+f_y y)} + \left(O(f_x,f_y)P((f_x+f_{off-x}),f_y)e^{i2\pi(f_x x+f_y y)} \right)^* \right)}_{Linear} + \underbrace{\left| \iint_{f_x,f_y \neq -f_{off-x}} O(f_x f_y)P(f_x,f_y)e^{i2\pi[(f_x-f_{off-x})x+f_y y]} \right|^2}_{Quadratic} \quad (1.1)$$

and similarly for I_{off-y}. The quadratic term is summed over all frequencies except the DC plane wave. $O(f_x, f_y)$ is the object/mask spectrum, $P(f_x, f_y)$ is the pupil function of a coherent optical imaging system, i.e. a uniform, radial low-pass filter to cutoff frequency NA/λ and f_{off} is the frequency offset of the off-axis illumination point source, $f_{off-x,y} = NA/\lambda$. The first term is the bias, the linear term is the desired quantity, and the quadratic term is the nonlinear image that causes image degradation. In addition, the pupil filter (function), $P(f_x, f_y)$, in each exposure can be reshaped by apodization (pupil filtering).

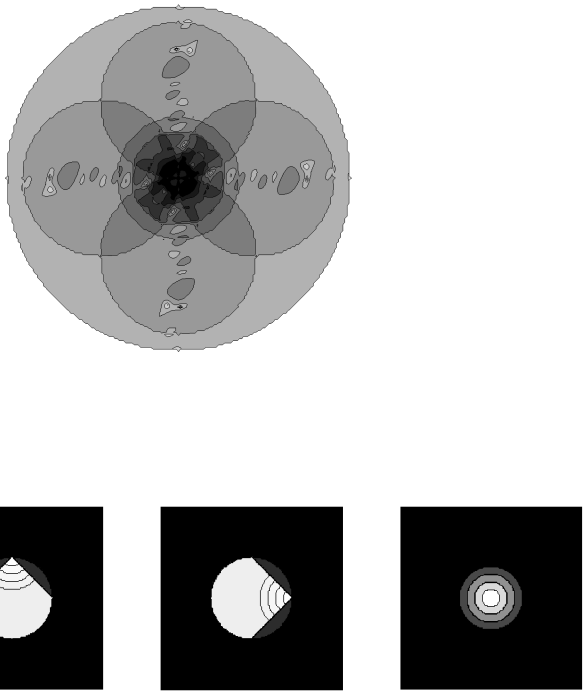
2.1. Algorithms

A detail description of the algorithms used to place the frequency coverage and parsing are described in [5]. The basic principle is based on transform coding commonly used in the signal processing community. Essentially, in coding the most significant coefficients in a transformer coder are kept and the rest are set to zero. A transform coder is analogous to the diffraction limited optical system where a lens is viewed as a finite Fourier transform filter. It is optimal to keep the most significant coefficients such that the most significant planes waves are passed through the pupil optimally. Algorithms are developed to search the frequency space to maximize the sum of squares of the coefficients in each exposure. From a lithography perspective, the source illumination is designed to place the diffraction orders in the optimal places in the pupil for the least error. Several computation algorithms with different complexity were explored. Greedy and dynamic programming types are most useful for speed and optimality respectively. Each exposure selects a frequency coverage area and hence that determine which Fourier coefficients pass through the optical system. With high NA's, the algorithms will converge to the typical solution shown in Figure 1 (a), however at smaller NA's, variations in frequency coverage are possible since smaller NA's cover less frequency space area and the promise of reintroducing the reference beam at the image plane was possible with low NA optical systems. In addition, alternative frequency coverages are discussed for experimental use and the algorithms can be used to search redundant expansions instead of typical Fourier expansions to possibly find better approximations using frequency coverages with correct overlap. It is known that non-orthogonal expansions allow better approximations. For example, overlapping frequency coverages can amplify certain frequency components to increase exposure latitude.

2.2. Filters and Coverage Figures

In this section the dynamic programming algorithm result at high NA is shown in Figure 1. The typical coverage for the off-axis illumination is at the expected locations on vertical and horizontal axes since a majority of spatial frequency

space is covered with a few high NA exposures. The four and five exposure pupil filters are shown on the right side of the figure. The four exposure method uses the left two pupil filters and the third includes the on-axis pupil filter for the five exposure system setup. An addition on-axis exposure is added to reduce the iso/dense bias from the off-axis exposures. The reference beam location in the pupil plane is at the apex of the triangle in the pupil filters. Comparisons of aerial images resulting from these configurations are discussed in the next section. Some addition frequency coverages with different geometries at low NA and with our experimental high NA lens with an obscuration in the center of the pupil are discussed in [5]. Special case NA's results in typical coverage geometries with tilted coverage to recover corner frequencies and to move the obscuration slightly away from the vertical and horizontal axes for the four and five exposure systems.



(a) (b)

Figure 1. (a) Frequency space picture of five exposure IIL coverage where optical limit frequency is $2/\lambda$. Concentric circles are various on-axis pupils of NA = .2, .4, .6 and .8. (b) Left figure is vertical frequency pupil filter. Middle figure is horizontal frequency pupil filter. Right is on-axis pupil filter for various on-axis NA. White represents transparent area, gray is opaque area, and black is the area outside the Fourier pupil for off-axis and on-axis exposures. Partial, concentric circles are on-axis pupil sizes.

Next, typical quadratic or dark field images results from three different pupil filtering configurations are shown in Figure 2 for the mask pattern in Figure 3 (a). Figure 2 (a) shows the dark field image from one off-axis exposure without any pupil filter and hence the interference is strong and causes a bias in the square box region leading to ringing defects as has been seen previously [1,2]. Figure 2 (b) illustrates the dark field image with the four exposure pupil filters in Figure 1 (b). The dark field has a less impact than previously with binary filters. Figure 2 (c) shows the dark field image of the off-axis image when an on-axis exposure is included. The bias in the box region is removed and only a small effect in the dense region of the pattern is apparent.

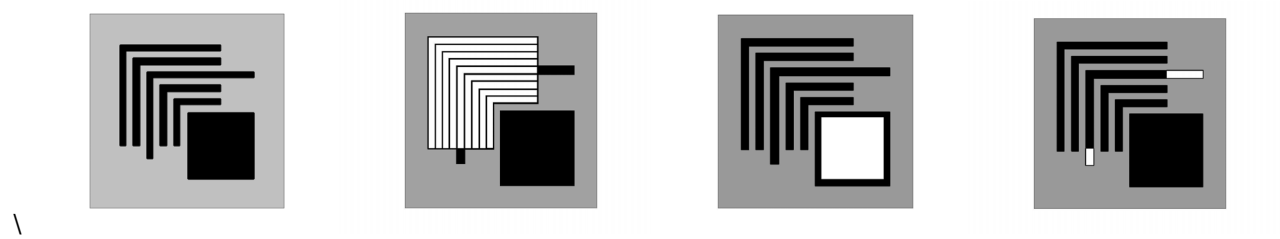


(a) (b) (c)

Figure 2. (a) Typical quadratic term without pupil filter. Strong DC bias in box region is evident. (b) Typical quadratic term (dark field) of four exposure IIL system where pupil filter is triangle shape. (c) Typical quadratic term from five exposure IIL system where pupil filter is moon shape. Peak value in dark field image is smaller and has a small DC bias in dense line area.

3. COMPARISONS OF AERIAL IMAGES

Simulations of IIL aerial images in vector and scalar models are discussed. Although there are differences in the precise aerial image calculation, the scalar model still illustrates the general trend in the aerial image at high NA's since polarization control is used in each separate exposure for the rectilinear patterns of interest. Figure 3 illustrates the various error weights used to emphasize error in dense and isolate parts of the pattern for aerial image comparisons in the figures to follow.



\

(a) (b) (c) (d)

Figure 3. These images represent the weights on the pattern to emphasize error in important areas. White corresponds to weighted region. Black are the typical lines of the pattern and the gray background is the space area of the mask. (a) Optimal mask, identity (unweighted) weight (b) Dense weight mask emphasizing dense area with weight = 100. (c) Box weight emphasizing large line area with weight = 100. (d) Isolated weight mask emphasizing isolated lines with weight = 600.

A weight, $w(x,y)$, is calculated as on the error between the binary mask, $M(x,y)$, and the threshold aerial image, $T(x,y)$ over the spatial coordinates x,y.

$$Error = \sum_{x,y} w(x,y)\big(M(x,y) - T(x,y)\big)$$

3.1.1 Scalar Model Comparisons

The IIL intensity scalar model typically shows a stronger biased aerial image that requires higher threshold values than single exposure aerial images because of the increase number of exposures. Later in the vector model section, it is illustrated that the bias is lessen by optimizing the exposure ratio dosage parameters. Scalar aerial images for various on-axis pupils tend to be the same for all error measures except the box weight error curve in Figure 4 (b). It is seen in section 4 that the addition of apodization can alleviate the sensitivity of the box error weight with respect to the on-axis pupil size. The inclusion of an on-axis exposure with pupil filtering make the error curves behaves closely to a linear

system. The similarity is illustrated in the linear term nearness to the various five exposure error curves to that of the four exposure error curves in Figure 4.

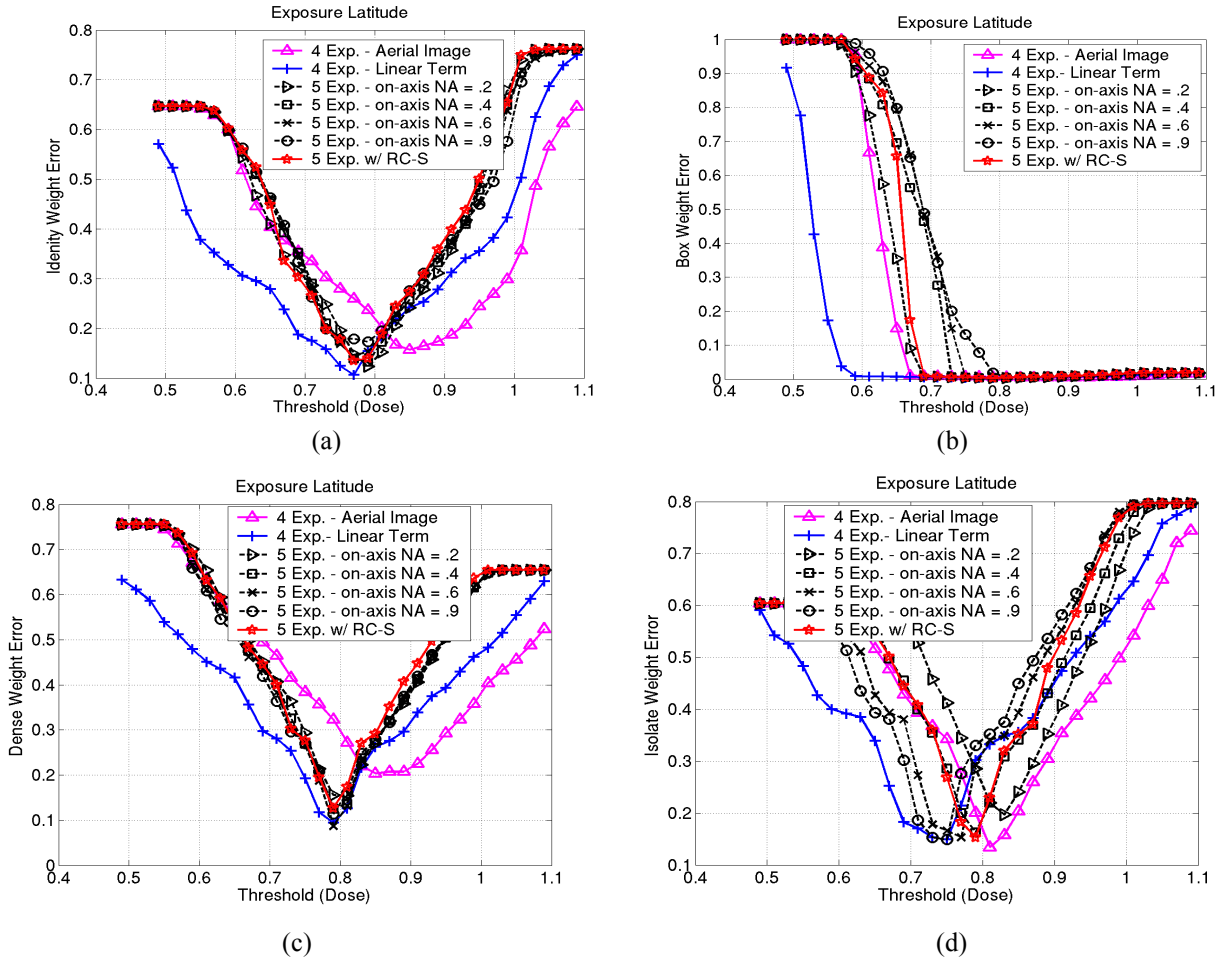
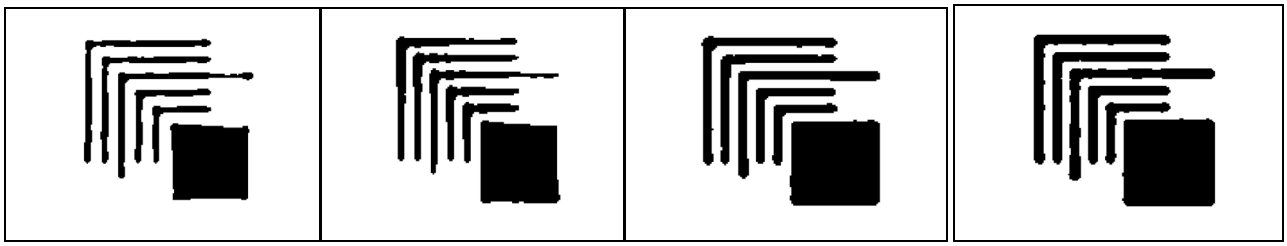


Figure 4. Error curves of threshold aerial images (a) Identity weighted error curves. (b) Box weighted error curve (c) Dense weighted error curve. (d) Isolate line error curve

There exists more exposure latitude in the four exposure system since the dense weight error curve in Figure 4(c) appears wider, but with a greater minimum error. The isolated line error curve in Figure 4 (d) shows that large on-axis pupil filter tends to print the isolated lines better. This is expected since the on-axis image defines the low frequency portions of the mask. The RC-S error curve is discussed Section 4.

The iso/dense bias affects the dense lines unevenly in the vector model and this creates difficultly to print dense features correctly than in the scalar model shown in the threshold aerial images in Figure 5. Figure 5 (a) and (b) display the resulting threshold image from the scalar and vector models. The corner region shows a strong iso/dense bias compared to the isolated lines. This is somewhat expect since the high NA accurately accounts for the high spatial frequencies better at the large propagation angles. Notice the top of the large box has the angular falloff that is still seen in the scalar model. The five exposure images in Figure 5 (c) and (d) from the vector and scalar model are similar, primarily because the five exposure system behaves more like a linear imaging system than the four exposure method.

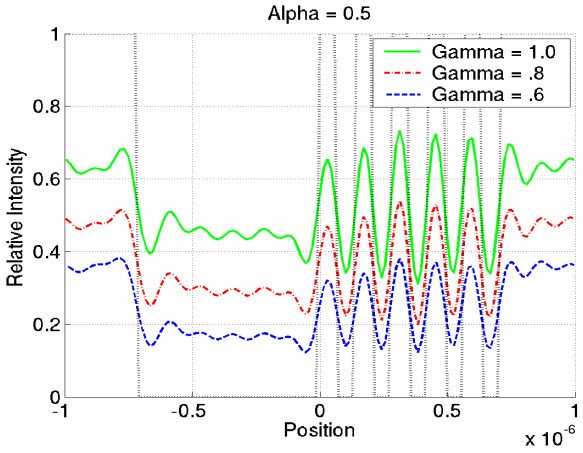


(a) (b) (c) (d)

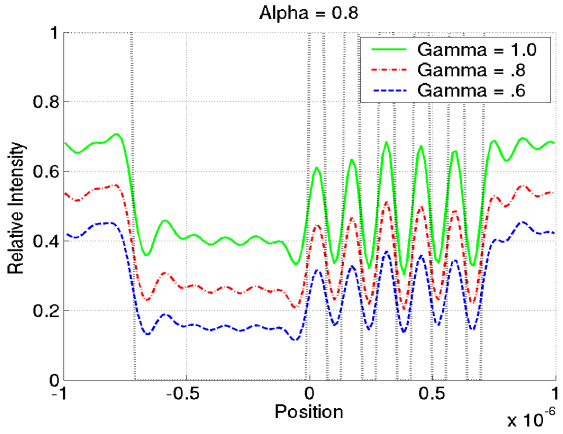
Figure 5. Threshold aerial images (a) Scalar model 4 exposures. (b) Vector model, 4 exposures. (c) Vector model, 5 exposures (d) Scalar model, 5 exposures

3.1.2 Vector Model Comparisons

Optimal values of exposure dosages typically depend on the pattern of interest and the simulation tool. Here we illustrate the effect of changing the α, γ parameters. Many aerial images are generated from PROLITH vector model with various combinations α, γ and here we show aerial image profiles and compute the same error curves in Figure 3 to understand quickly the image quality to print the best pattern possible for various combinations of α, γ .



(a)



(b)

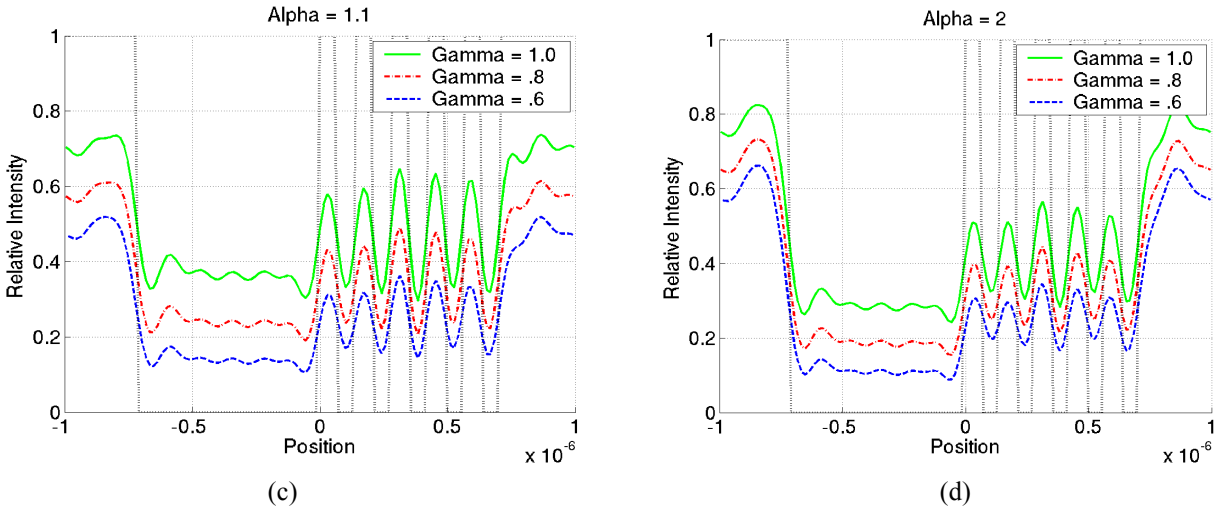
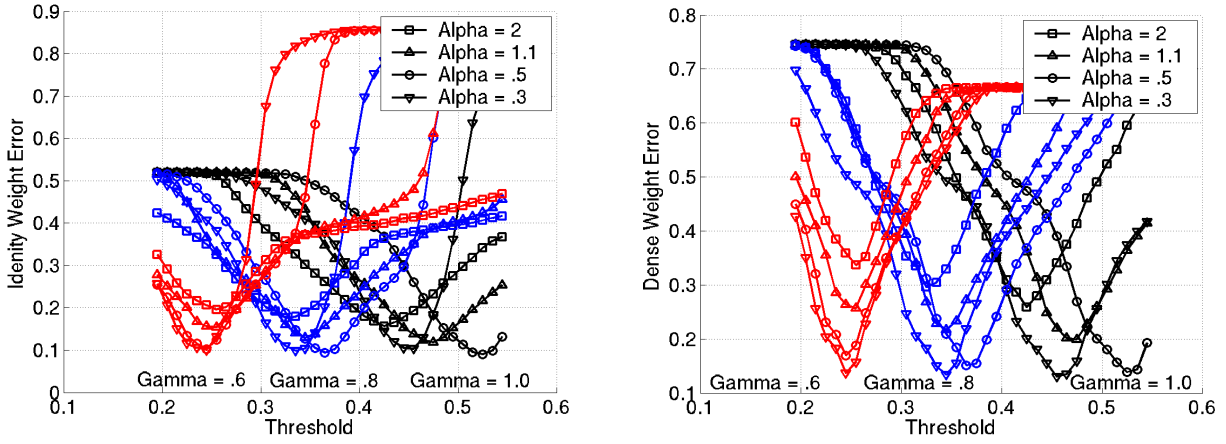


Figure 6. PROLITH Vector aerial image profiles for various α and γ values. (a) α = .5 (b) α = .8 (c) α = 1.1 (d) α = 2.

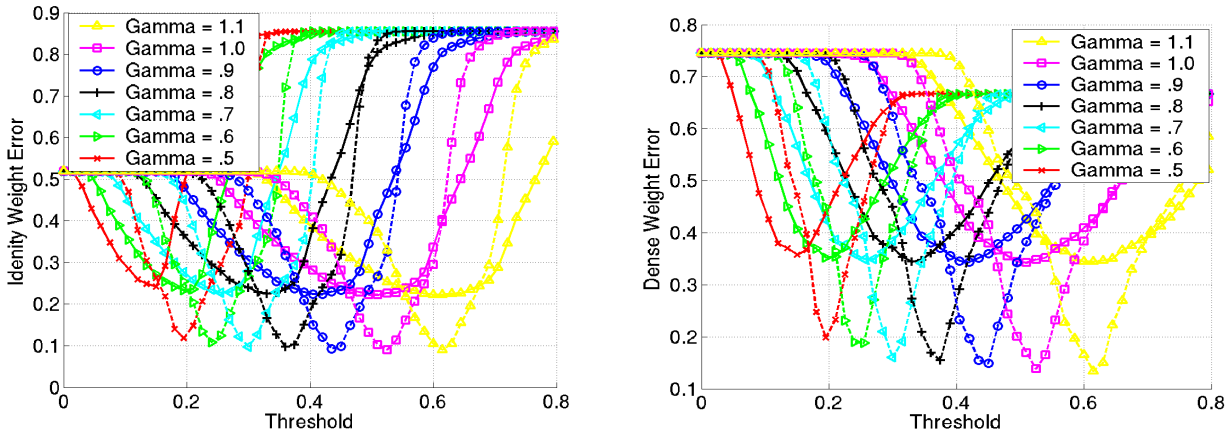
Figure 6 shows aerial image profiles for four α's with varying γ's values. It is observed that the varying the γ values increases the bias and increases the exposure latitude of the dense lines. This behavior matches well with the scalar aerial image intensity in eqn. 1.1. The square of γ is proportional to the bias and γ multiplies the linear term; this effect is observed in the vector model in Figure 6 and consequently in the error curves of Figure 8. As γ is increased the bias shifts the error curve right and tends to widen the error curve. From observations in Figure 6, 7 and 8, the γ control over the DC background bias is important since small changes in γ affect the aerial image bias significantly. Typical dosages should be between $5E_o$ and $2E_o$ (.2 and .5 threshold values). Because the simple resist model is based on a threshold function, it is important to control the bias of the multiple exposure system through γ since biasing also limits resolution. The gain by using a high γ values is offset by a large DC background without a large gain in the exposure latitude in the linear term amplification. Primarily, γ is used to put the bias in the resist threshold range so that the dense structures are at the correct threshold values. From observations in PROLITH, optimal value of γ, α are .8, .5 respectively. This gives a threshold value of approximately .37. Comparisons in Figure 6, show that the on-axis aerial image should have similar contrast to the off-axis aerial images to create a robust photoresist image with similar contrast for on-axis and off-axis intensities. For a fixed γ, a large α value amplifies the isolate structures and decreases the exposure latitude of the dense lines. Using a small α decreases the exposure latitude of the on-axis image (isolated structures) and hence reduces the image quality of the photoresist image. In Figure 7, the middle group of curves with γ=.8 and with α equal to .5, the error curve is widest in comparison with the four other curves with other α values. For example, the error curve becomes larger with decreasing α, but after α =.5 the error curve becomes narrow again. A simple method to determine an optimal α during simulation is to measure the maximum intensity of the off-axis and on-axis aerial images. Then α can be determined to equalize the intensities of the off-axis and on-axis aerial images. Then α is the ratio between $I_{\max-off} / I_{\max-on}$. This will adjust the intensities between the off-axis and on-axis images such that the overall aerial image contrast matches well and a wide threshold region is found for photoresist development.

Figure 7 shows the effect of varying α for fixed γ values using the general and dense error measures. The γ determines the region for aerial image thresholds and the α helps to vary the exposure latitude locally in the threshold region. For large and small α values, the error curve is becomes narrower as compared to the optimal values of .5 that is observed to be optimal for the photoresist image developed from the aerial image [1]. A similar trend is seen in the dense error measure in Figure 7 (b).



(a) (b)

Figure 7. Error curves for five exp., γ =.6, .8, 1.0. Each group of γ consists α = 2, 1.1, .5, .3. (a) Unweighted error (b) Dense error.



(a) (b)

Figure 8. Error curves for 4 exp. (solid line) and 5 exp. (dash line) aerial images for fix α =.5 and various γ . (a) Unweighted error curve. (b) Dense error metric.

Figure 8 (b) further illustrates that dense lines print better in five exposures instead of four exposures by a factor of approximately two in the minimum error. Although, there is more contrast due to the quadratic term to widen the error curve, it degrades and increases the relative minimum error.

4. APODIZATION

High NA's allow a significant amount of frequency overlap between on-axis and off-axis exposures as illustrated in Figure 1. Designing filters for overlapping frequency spectrum between various frequency bands of signals is large facet of filterbank theory and signal processing. The use of subapertures in the different IIL exposures allows the use of smooth pupil functions such that the transitions in the frequency spectrum are smooth in transition instead of the brickwall shape of the MTF in coherent illumination. Of course, at the highest spatial frequencies of the system, the imaging physics still dictates hard cutoffs with coherent illumination. Smooth transitions in the overlap regions allow higher resolution with lowered image degradation. Radially, nonsymmetrical filters are considered in [6,7] to improve aerial image robustness and DOF, however here the filters are designed with a two channel filterbank idea in mind to improve image fidelity.

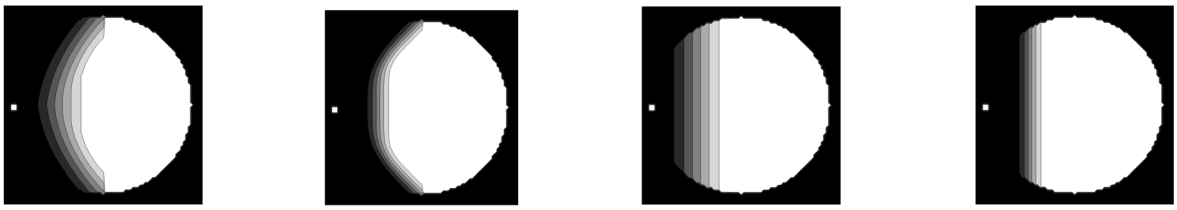
Figure 9 displays two on-axis pupil filters with the sharp raise cosine and raise cosine filter with square symmetry. The sharp raised cosine has a fast varying, narrow transition band and a flatter central peak. The raised cosine is smoother and contains a wider transition region.



(a) (b)

Figure 9. On-axis pupil filters with square symmetry. (a) Sharp raised cosine (b) Raised cosine

The same functions for the transition areas of the filters are used, however the off-axis filters fan out in diagonal or straight directions. Figure 10 consists of nonsymmetrical off-axis pupil filters. The horizontal frequency filters are shown because the vertical frequency filters are 90 degree rotations of the horizontal pupil filters. Figure 10 (a) and (b) have a triangle shape and even out the overlap in frequency coverage between the horizontal and vertical off-axis exposures. The filters in Figure 10 (c) and (d) overlap and amplify the corner frequencies and may provide higher contrast for corner frequencies.

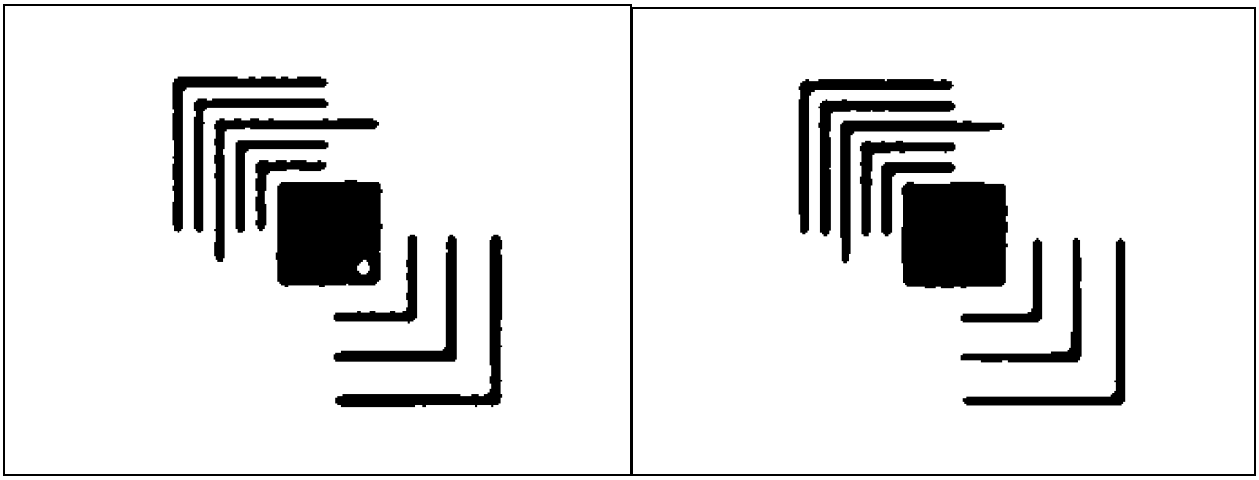


(a) (b) (c) (d)

Figure 10. Off-axis pupil filters. (a) Triangle shape – raise cosine (b) Triangle shape – sharp raise cosine (c) straight shape – raise cosine (d) straight shape – sharp raise cosine

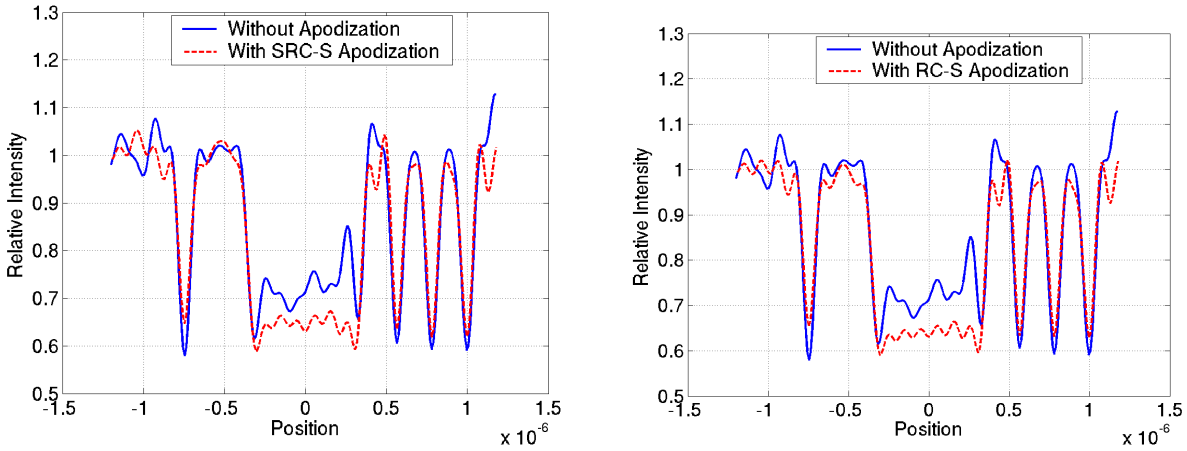
Further use of other filters such as the Bartlett, Gaussian windows, and other filters are possible. The transition regions may not mirrors of each other between the on-axis and off-axis filters and optimizing the transition region is possible to improve image quality.

Threshold images of the aerial image at the same threshold level are shown in Figure 11. Figure 11 (a) displays the defect caused by the ringing that is well-known property of imaging in coherent light. The threshold image resulting from apodization with the raise cosine filter is shown in Figure 11 (b). The threshold images are similar for all the apodization filters discussed, hence only one image is shown. In all cases, the defect has been mitigated by the addition of apodization. Aerial image profiles are show in Figure 12, taken vertically through the defect region of the image in Figure 11. The raise cosine filter reduces the defect slightly better than the sharp raise cosine pupil filter due to the smoother response. In addition, the apodization is carried out with the mask pattern in Figure 3 (a) and the error curves for the raise cosine with square symmetry are include in Figure 4 for comparison. It is observed that the isolated, dense and unweighted error curves are all similar to the error curves with binary filtering; improvement in the box error curve is attained by reduced biasing in the large box region that is worse in the case of binary pupil filtering.



(a) (b)

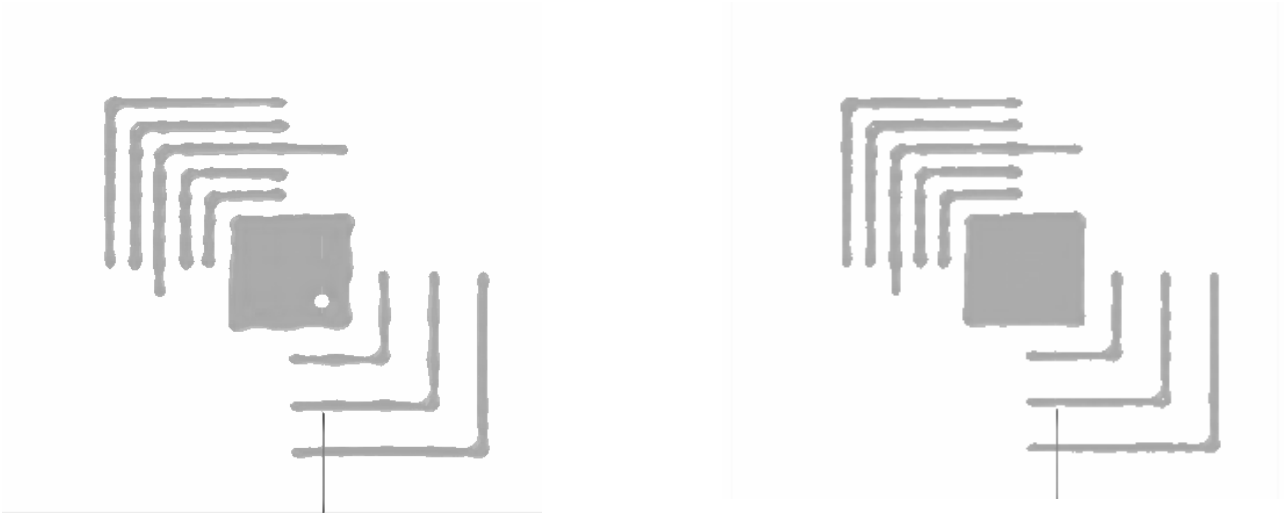
Figure 11. Threshold scalar aerial images. (a) Without apodization using binary filters of Figure 1 (b) With raise cosine apodization.



(a) (b)

Figure 12. Scalar aerial image slices with and without apodization (a) Sharp raise cosine with square symmetry (SRC-S) (b) Raise cosine with square symmetry (RC-S)

Figure 13 shows photoresist images using PROLITH vector imaging model. The hole in the box region reproduces well with the scalar model threshold image in Figure 11 (a). The defect is removed by the apodization filters to reduce the ringing and the resulting photoresist image is shown in Figure 13 (b) Note, the linewidths are straighter and some iso/dense bias is apparent.



(a) (b)

Figure 13 Photoresist images with and without apodization using vector imaging model at high NA (a) Photoresist image with binary filters of Figure 1. (b) Photoresist images with raised cosine filters.

5 CONCLUSION & FUTURE WORK

Algorithms have been developed to optimize the frequency coverage and image quality in IIL. They optimize the coverage such that the most significant plane wave making up the pattern are passed through the system. At high NA, the solution converges to the typical solution, however they provide insights at low NA's and provide alternatives and options to overlap coverages with higher number of exposure for more accurate aerial images.

Comparisons between scalar and vector model simulations are made. The behavior of the aerial image in scalar and vector tend to hold with the Manhattan pattern and polarization control. The quadratic term helps to increase contrast in some cases; however it often degrades the image quality beyond acceptability. Controlling the dosage ratios between the various exposures helps to increase exposure latitude and contrast. These parameters of IIL provide aerial image robustness necessary for second generation RET.

Lastly, while a binary pupil filter approach has been shown to control the quadratic term in the four exposure to an extent, additional on-axis (low frequency) coverage with a binary pupil filter reduces the iso-dense bias and impacts the pattern printability significantly. At high NA, the five exposure IIL allows apodization to mitigate defect printing between exposures. Apodization can help reduce the ringing without reducing the low κ_1 imaging offered by IIL. Vector simulations at high NA and full photoresist modeling with apodization are shown to behave similarly to the scalar and threshold resist models. Implementation issues need to be investigated for apodization and further experimental work is necessary to fully understand the impact of apodization at high NA for improved image fidelity, leading to high resolution imaging without the need for phase-shift masks and extensive OPC.

ACKNOWLEDGMENTS

We thank Abani Biswas and Christian Schwarz for discussions on PROLITH and lithography. This work is support by ARO/MURI under Grant DAAD19-99-1-0196.

REFERENCES

1. A. Biswas and S. R. J. Brueck, "Simulation of the 45-nm Half-Pitch Node with 193-nm Immersion Lithography" *J. of Microlithography, Microfabrication, and Microsystems* **3**, pp 35-43, 2004

2. X.L. Chen and S.R.J. Brueck, “Imaging interferometric lithography: A wavelength division multiplex approach to extending optical lithography” *J. of Vac. Sci. & Tech. B.* **16**, pp. 3392-3397, 1998.
3. S.R.J. Brueck and X.L. Chen, “Spatial frequency analysis of optical lithography resolution enhancement techniques” *J. Vac. Sci. & Tech. B*. **17**, pp. 908-920
4. T.M. Tridhavee, B. Santhanam, and S.R.J. Brueck, “Optimal Frequency Coverages and Parsing for Imaging Interferometric Lithography”, *J. of Microlithography, Microfabrication, and Microsystems.* – Submitted 2004
5. E. Wu, B. Santhanam, and S.R.J. Brueck, “Characterization of Imaging Interferometric Lithography in Simulation”, *J. of Microlithography, Microfabrication, and Microsystems* – Accepted 2004
6. R.M. Bunau, H. Fukuda, and T. Terasawa, “Effects of radially non-symmetric pupil filters and multiple-pupil exposure”, *Optical Microlithography IX, Proc. SPIE* 2726, pp 375-385, 1996
7. R.M. Bunau, G. Owen, and R.F. Pease, “Optimization of Pupil Filters for Increased Depth of Focus”, *Jpn. J. Appl. Phys*. **32** pp. 5850-5855, 1993
8. D. Gottlieb and C.W. Shu, “On the Gibbs phenomenon and its resolution”, *SIAM REVIEW* **39**, 4 pp. 644-668, 1997
9. A.K. Wong, *Resolution Enhancements Techniques in Optical Lithography*, SPIE PRESS **TT47,** 2001.

Deep-UV Immersion Interferometric Lithography

Alex K. Raub[a], S. R. J. Brueck[a], Will Conley[b], Ralph Dammel[c], Andy Romano[c], Mitsuru Sato[d]

[a] Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106
[b] Motorola, Austin, TX 78729
[c] AZ Clariant, Somervill, NJ 08876
[d] TOK, Koza-Gun, Kanagawa-Ken 253-0114, Japan

alex.raub@chtm.unm.edu

ABSTRACT

Liquid immersion lithography (LIL) can extend the resolution of optical lithography well beyond today's capabilities. The half-pitch limit is given by the well-known formula $\Lambda = \lambda/(4NA)$, where λ is the optical wavelength and $NA = n\sin(\theta)$ is the numerical aperture of the exposure device with n the refractive index of the exposure medium. Through the use of exposure media such as purified water (refractive index of 1.44 at 193 nm), it is possible to reduce minimum pitches by a factor of as much as 44% - a full technology node. Beyond this simple observation, there is a good deal of work necessary to fully understand the impact of LIL on a lithography processes. This paper will address issues concerning resist chemistry and the impact of water immersion on the imaging capabilities of different resist formulations. All resists were evaluated by imaging dense line-space structures at a 65-nm half-pitch both in air and with water immersion. Studies of the dense 65-nm lines made by immersion imaging in HPLC grade water with controlled variations in resist components were performed. Significant differences were observed and will be discussed.

Keywords: 193-nm resist, liquid immersion, interferometric lithography, 65-nm half-pitch.

1. INTRODUCTION

As the microelectronic industry pushes towards smaller feature sizes, greater demands are placed on lithography. Modern lithography now commonly produces features smaller than the optical wavelength. Through the use of optical enhancement techniques such as off-axis illumination (OAI), optical-proximity correction (OPC), phase-shift masks (PSM), and imaging interferometric lithography (IIL), industry can extend the current limits of state-of-the-art lithography tools. These techniques are quickly being pushed to practical limits while next generation lithography technologies are still under development. Industry needs to find new ways to extend their lithography capabilities to enable continued scaling of ICs. The ultimate minimum feature optical lithography can produce is a line and space whose pitch is determined by the well-known formula, $\Lambda = \lambda/2NA$, where λ is the wavelength of the light and $NA = n\sin(\theta)$ is the product of the immersion medium refractive index and the sine of the half angle of the light at the exit of the lens.

Historically, the exposure wavelength λ has been steadily decreased and the numerical aperture NA has been steadily increased to achieve higher resolution. Current leading edge exposure tools operate at a wavelength of 193 nm with NA ~ 0.7 with projections to NA ~ 0.85 to 0.9 in the near future. Currently there is no projected optical node beyond 157 nm where significant material issues remain outstanding. Recently, interest has grown dramatically in exploring the use of an immersion medium between the exit face of the lens and the photoresist surface to extend optical lithography.[1] Currently air is the only exposure medium being used for commercial exposure tools with $n = 1$. The use of immersion media to extend optical resolution is well known from oil-immersion optical microscopy. Through the use of immersion media, such as de-ionized water with $n \approx 1.44$ at 193 nm[6], much smaller minimum feature sizes can be achieved than are possible in air. This can be thought of either as an effective wavelength $\lambda_{eff} = \lambda_0/n \sim 134$ nm or (not and) alternatively as an increase to an effective $NA_{eff} = n\sin\theta \le 1.44$. Thus, the use of water immersion lithography could poten-

tially extend 193-nm wavelength optical lithography by 44%, twice the resolution enhancement offered by the problematic transition to 157 nm. Liquid immersion lithography also provides an improvement in depth of field. At a fixed resolution, $DOF = k_2\lambda_0/n\sin^2(\theta)$.[7] Application of immersion techniques at 157 nm will have a similar impact; however a suitable liquid has yet to be determined.

There are some issues that need to be understood before immersion lithography can be used in manufacturing. One issue is the interaction between the immersion fluid and the last lens element. For instance de-ionized water (DI H_2O) will etch Calcium-fluoride (CaF) lens elements, so a protective coating will need to be put on the CaF final lens element that will not be influenced be the presence of the DI water. Another issue, which is the focus of this paper, is the interaction between the immersion fluid and the resist-coated wafer. With liquid immersion lithography there are two new interactions of concern, first any uptake of the immersion liquid into the resist and the impact on resolution, and second the adsorption of resist components into the immersion liquid and contamination of the lens surface. Both of these processes can change the photoresist chemistry and the optical properties of the immersion liquid.

This paper presents exposure results for immersion lithography at a wavelength of 213 nm, close to the current 193-nm exposure wavelength. A maskless, interferometric lithography approach that gives essentially ideal contrast fringes was used.[8] Exposures were carried out on a number of resists with controlled variations in resist components. The images of dense 65-nm lines obtained from the resist exposed in air and in HPLC grade water were studied. Significant differences were observed and will be discussed. The use of a protective coating for the resist was also studied and will be discussed.

2. EXPERIMENTAL SETUP

The light source was a Coherent Infinity 40-100 Nd:YAG pulsed (~ 3 ns) laser followed by nonlinear crystals to generate a fifth-harmonic beam at a wavelength of 213 nm. This source has a much higher coherence length (~ 0.6 m) than that of a typical 193-nm excimer laser, which greatly simplifies the experimental arrangement for interferometric lithography. 193-nm resists generally have comparable absoption at 193- and 213-nm and work well at this wavelength. A simple telescope with an aperture expands the ¼-inch beam to about 1.5-inches at the entrance of the immersion interferometer setup.

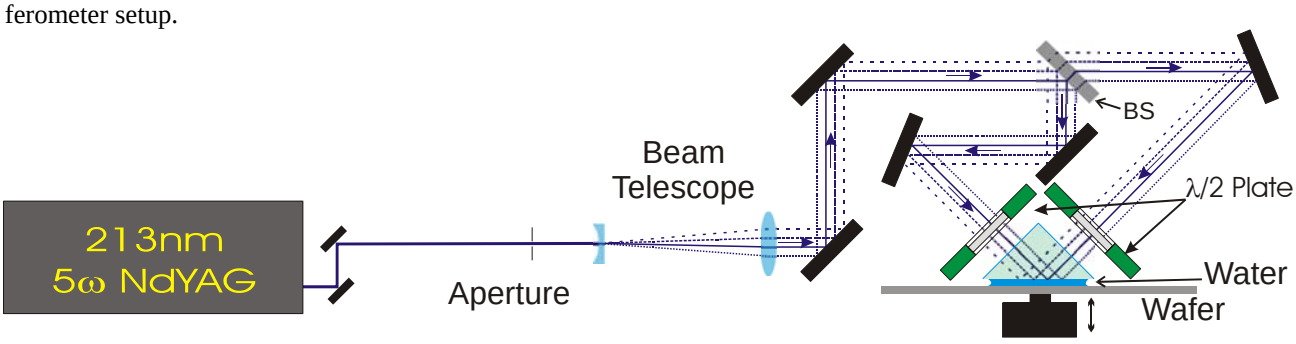


Figure 1 – Michelson interferometer immersion exposure setup.

A Michelson interferometer with a coupling prism was designed and constructed for performing the immersion exposures. Figure 1 shows a diagram of the exposure setup. The simplest way to couple the light into an immersion liquid was to use an isosceles prism. The top faces of the prism are antireflection coated. A drop of water is used between the bottom of the prism and the resist coated wafer. Both laboratory deionized water and HPLC water were used with comparable results. Figure 2 illustrates the two beam interference at the liquid-resist interface. Here, n_1 and n_2 denote the indices of the prism and immersion liquid, and θ_1 and θ_2 are the corresponding propagation angles with respect to the wafer normal, as drawn n_1 is larger than n_2. The liquid layer is about 1 mm thick. The period of the grating exposed in the resist is given by $\Lambda = \frac{\lambda}{2 \cdot n_1 \cdot \sin(\theta_1)} = \frac{\lambda}{2 \cdot n_2 \cdot \sin(\theta_2)}$. In this configuration, the grating period is *independent* of the refrac-

tive index of the plane-parallel liquid layer: the change in wavelength according to $\lambda=\lambda_0/n$ is exactly canceled by the change in θ due to refraction at the glass-liquid interface. In practice it is desirable for the glass and liquid indices to be approximately equal, in order to minimize reflections at the interface.

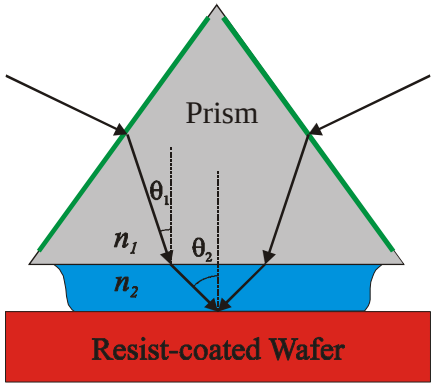


Figure 2 – Prism interface with immersion liquid.

An advantage of this arrangement is that the beam is folded directly onto itself, allowing for interference with beams with only small spatial and temporal coherence lengths. Disadvantages are the cumbersome alignment required to change the pitch, and the lack of a direct measurement of the liquid refractive index. Fig. 3 shows a picture of the Michelson interferometer immersion exposure setup used for the experiments.

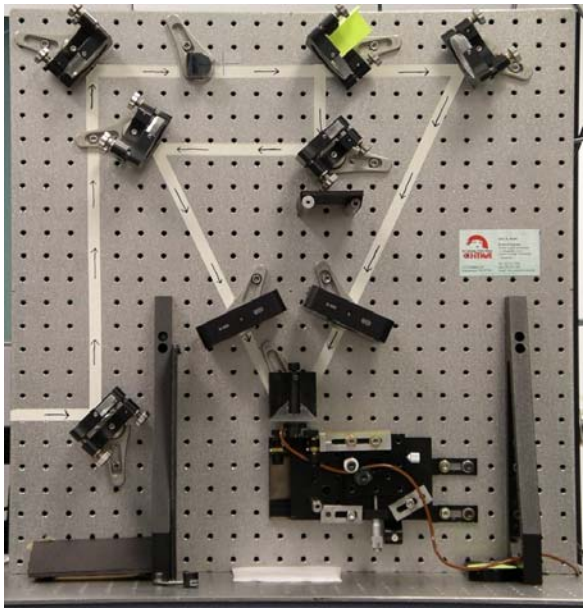


Figure 3 – Michelson interferometer immersion exposure setup.

Half-wave plates are mounted above the prism and used to rotate the polarization of the incident light. This allows for full control of the polarization state of each interfering beam. TE polarized light always gives the best contrast in the interference pattern and was used for all of the results reported here. By incorporating the half-wave plates experiments can be executed that mimic the imaging performance of a actual lithography system, in terms of illumination polarization state.

3. EXPERIMENTAL RESULTS

Two resist manufactures provided resists in which the material was fully disclosed as to the chemical components and makeup of the resist. The first company was TOK that provided six samples of resist. These resist were formulated using two different polymers and three different bases. The second company was AZ Clariant that provided eight samples of resist. Their resist were formulated using two different polymers, two different PAGs, and two different bases

3.1. TOK resist liquid immersion imaging results

The TOK resist used Acrylic polymer backbones. Three of the resists ILP01, ILP02, and ILP03 were based on an acrylic ter-polymer, and the other three, ILP04, ILP05, and ILP06, on an acrylic tetra-polymer. All six resists used the same photoacid generator (PAG), Tri-Phenylsufonium nanofluorobutanesulfonate (TPSNf). Three different base quenchers were used. Resists ILP01 and ILP04 used a small molecular sized base triethanolamine. Resists ILP02 and ILP05 used a medium sized base tripentylamine. Resists ILP03 and ILP06 used a large sized base tri-n-dodecylamine. Table 1 summarizes the matrix of TOK-ILP resist formulations. All resist were spun at a thickness of 100 nm unless noted otherwise, atop a 32-nm thick Brewer Science ARC28-4 anti reflective coating that was optimized to give a minimum substrate reflection for 65-nm half-pitch dense patterns.

		Base Quencher		
		Triethanolamine (Bs)	Tripentylamine (Bm)	Tri-n-dodecylamine (BL)
Polymer Backbone	Acrylic ter-polymer	TOK-ILP01	TOK-ILP02	TOK-ILP03
	Acrylic tetra-polymer	TOK-ILP04	TOK-ILP05	TOK-ILP06

Table 1 – TOK resist formulation matrix

3.1.1. Image comparison of resist formulations

Figure 4 shows the three Acrylic ter-polymer resists (ILP01, ILP02, and ILP03) exposed in air and in HPLC H_2O. All three samples gave reasonable exposures in both air and water. The samples exposed in air had noticible top rounding. The samples exposed in HPLC H_2O a smaller degree top rounding. All the samples exposed in HPLC H_2O had better line profiles than the corresponding air-exposed samples, with the sample containing the tri-n-dodecylamine base (ILP03) giving the best results.

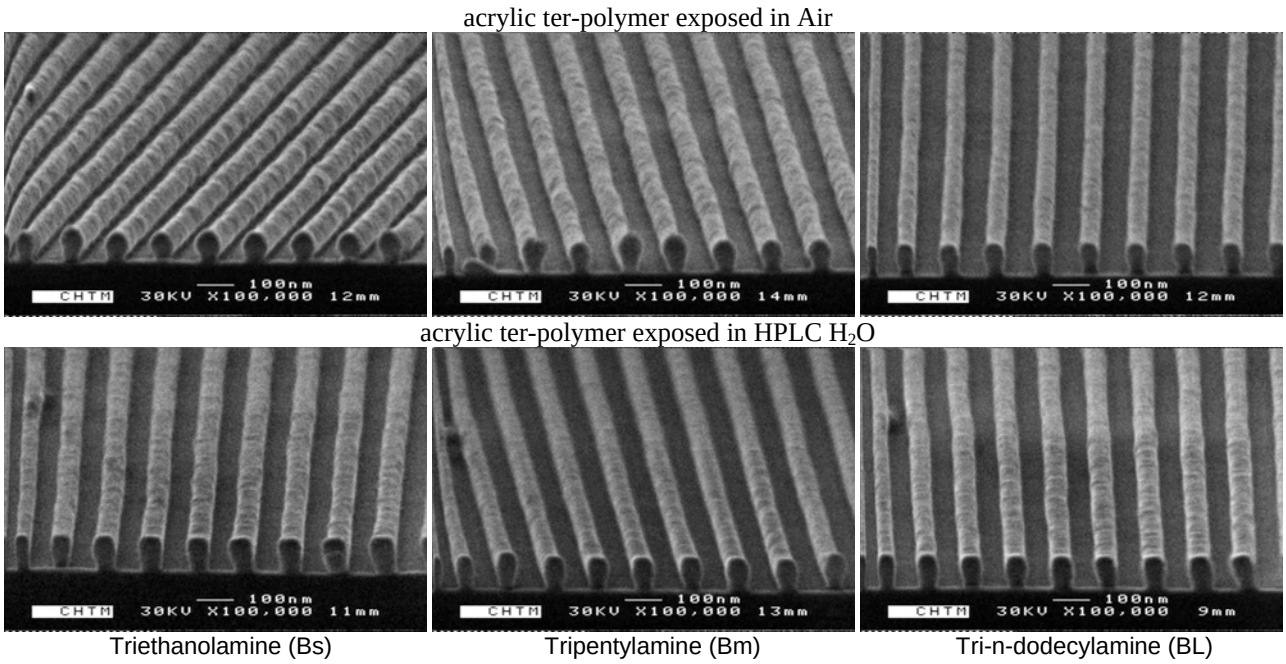


Figure 4 – 131-nm pitch TOK acrylic ter-polymer resist LIL results.

Figure **5** shows the three Acrylic tetra-polymer resists (ILP04, ILP05, and ILP06) exposed in air and in HPLC H_2O. All three samples gave reasonable results when exposed in air. Two of the samples gave reasonable exposure in the HPLC water. The sample with the small Triethanolamine base (ILP04) had micro bridging and pattern collapse in the HPLC H_2O exposure. The samples exposed in air had top rounding similar to the acrylic ter-polymer resists. The samples exposed in HPLC H_2O showed a reduced top rounding, and the smaller the base the less rounded the line profile. Two of the samples exposed in HPLC H_2O had better line profiles than the corresponding air-exposed samples, with the sample containing the tri-n-dodecylamine base (ILP03) giving the best results.

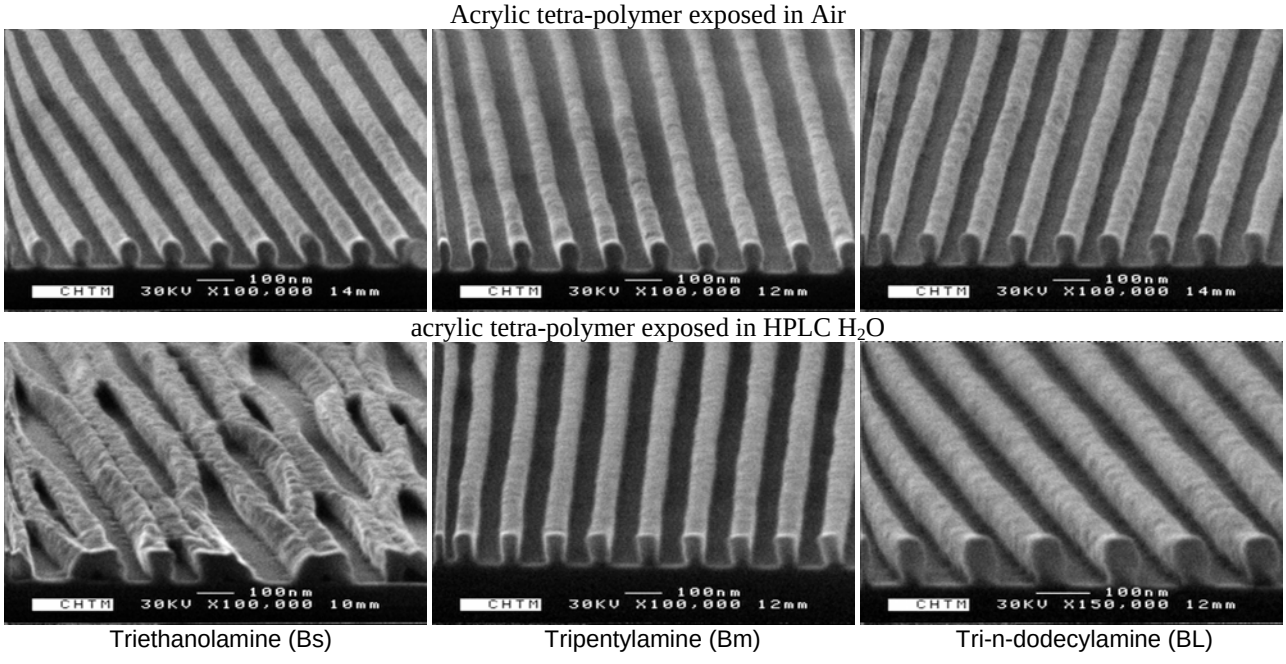


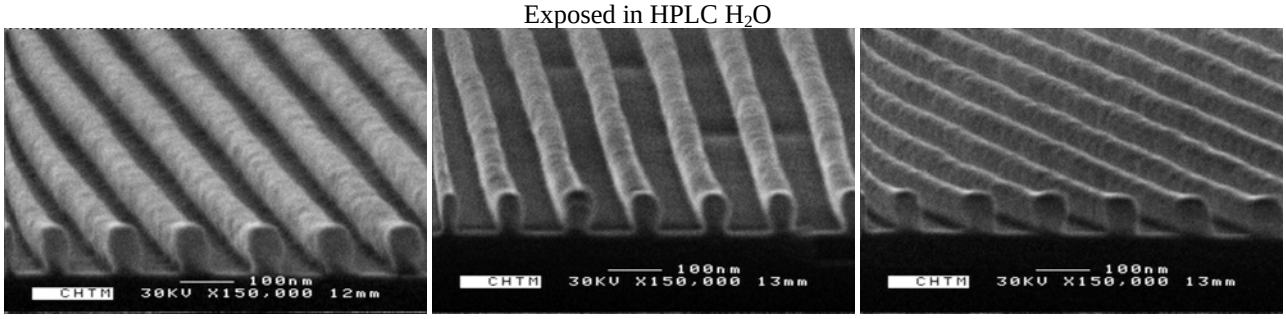
Figure 5 – 131-nm pitch TOK acrylic tetra-polymer resist LIL results.

The samples made from the acrylic ter-polymer gave slightly better results than the samples made form the acrylic tetra-polymer when exposed in HPLC H_2O. The triethonalamine base did not perform well as the other bases in the HPLC H_2O exposure, especially with the acrylic tetra-polymer backbone. The best TOK resist for immersion lithography was TOK-ILP03 closely followed by TOK-ILP06 both containing the tri-n-dodecylmine base.

3.1.2. Impact of resist soak times

Using one of the more successful TOK resists (TOK ILP06), the impact on the time the sample soaked in HPLC water was investigated. First a sample was exposed as soon as it was immersed in the water and dried off right after the exposure followed by a post-exposure bake, for a maximum water contact time of less than 30 seconds. A second sample was soaked in the water for 10 minutes before it was exposed and immediately dried off and PEBed. A third sample was exposed as soon as it was immersed in the water and then soaked in the water for 10 minutes at before it was dried off and PEBed. Figure 6 shows the exposure results for the three samples.

Exposed in HPLC H_2O



<30 se. H_2O contact time | 10 min pre-soak | 10 min. post soak

Figure 6 – Impact of soak times on 131-nm pitch TOK-ILP06 resist.

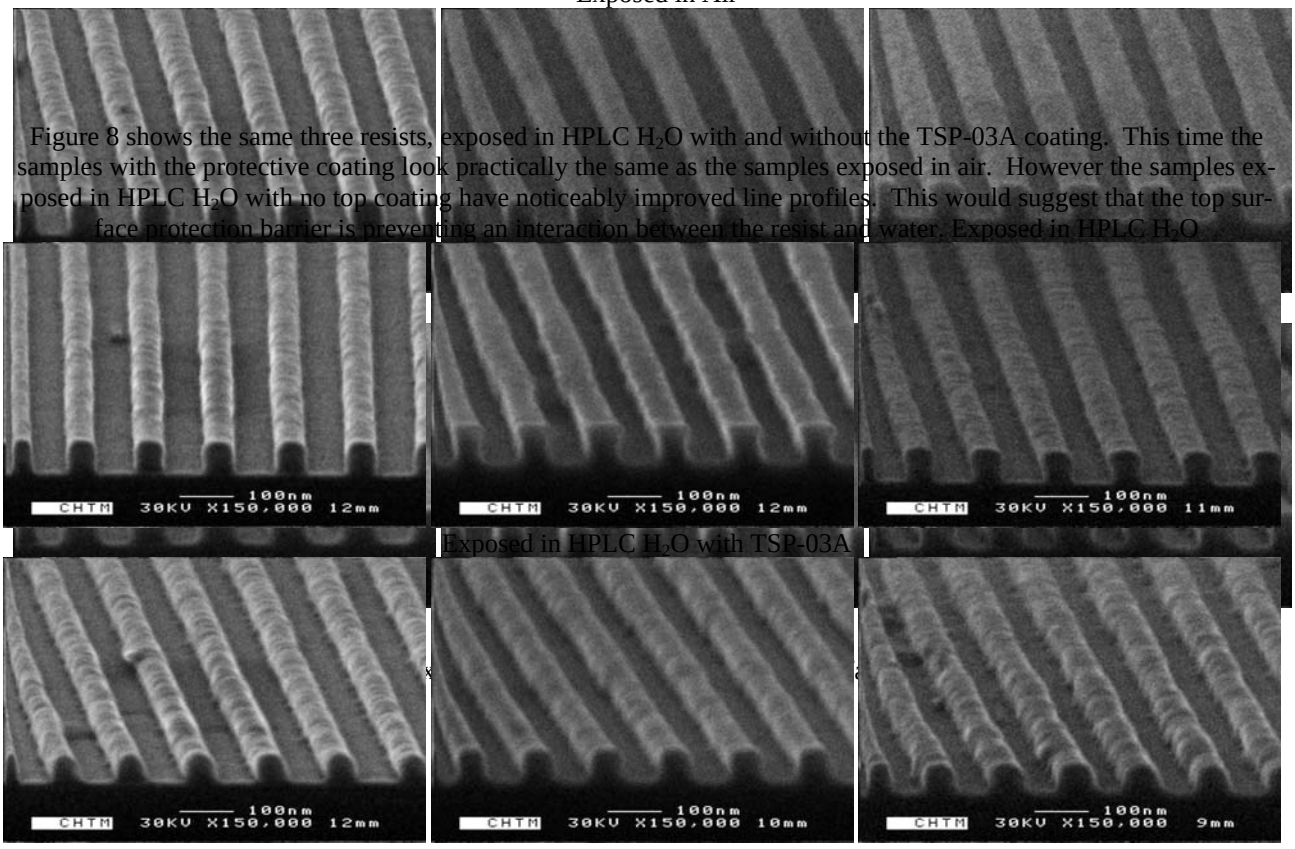
The 10-minute pre-soak sample had slightly better exposure latitude and uniformity, which cannot be seen from Fig. It is clearly seen that regardless of soak time in the water all samples gave similar and good line profiles. It is believed that from this and other experiments that whatever interaction happens between the resist and water take place in a very short time (less than 1 second)

3.1.3. Impact of top surface protection

One proposed technology to isolate the resist from the immersion liquid is the use of a top surface protection barrier layer. The barrier layer needs to be transparent to the exposure light and to be inert to both the immersion liquid and the resist. TOK has formulated a floro-carbon spin-on top protective coating TSP-03A, which meets these requirements for a barrier layer. Two additional steps are added to the processing. First after coating and post-application baking the wafer with resist, an additional spin-on and bake sequence is used to coat the TSP-03A material atop the resist. Second after the post exposure bake, the TSP-03A coating is removed using a special solvent TSP-3 remover before the resist development.

Figure 7 shows TOK-ILP01, TOK-ILP03, and TOK-ILP06 resist exposed in air with and without the TSP-03A coating. As expected, the samples with the protective coating look practically the same as the samples without the coating when exposed in air.

Exposed in Air



TOK-ILP01 TOK-ILP03 TOK-ILP06

Figure 8 – 131-nm pitch TOK resist exposed in HPLC H_2O with and without a top surface protection coating.

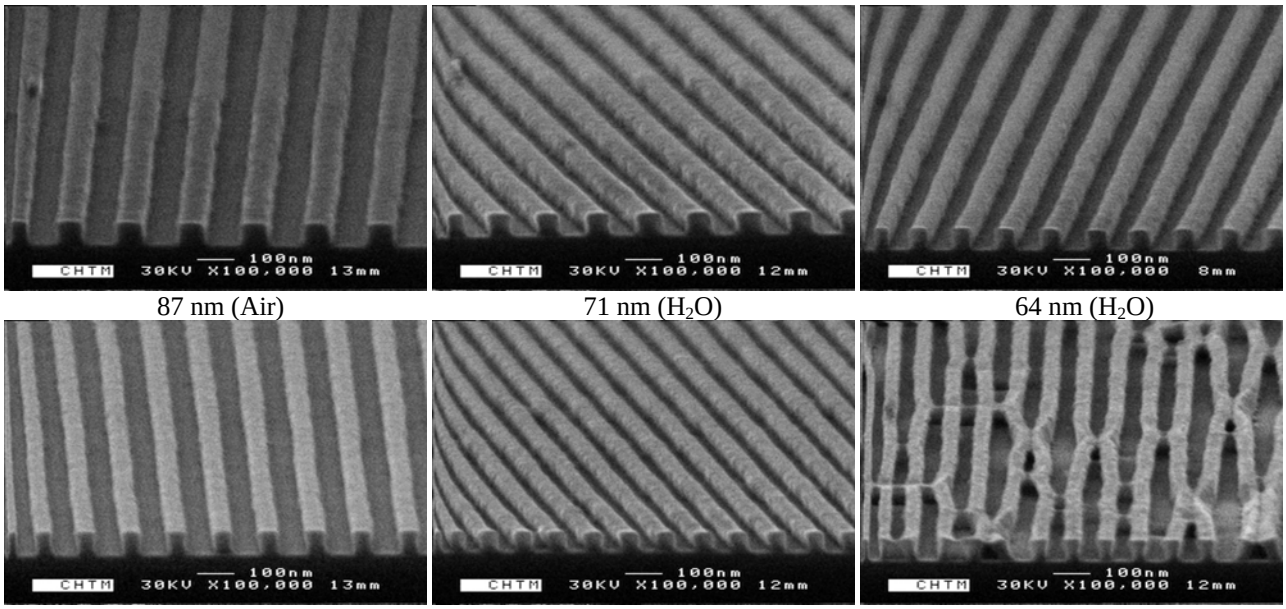
Figure 8 shows the same three resists, exposed in HPLC H_2O with and without the TSP-03A coating. This time the samples with the protective coating look practically the same as the samples exposed in air. However the samples exposed in HPLC H_2O with no top coating have noticeably improved line profiles. This would suggest that the top surface protection barrier is preventing an interaction between the resist and water. Exposed in HPLC H_2O

Other experiments suggest that the photo-acid generated upon exposure is leached out of the top few monolayers of the resist into the water. This would also support the squaring of the line profiles for the immersion exposure, as the top of the resist becomes less sensitive as compared to the bulk of the resist.

Although a top surface protection layer does prevent resist water interaction, there are resists that perform well when exposed in water. From an imaging point of view there is no real need for the added cost and complexity of using a top surface protection layer. Even at this early stage of development it looks possible to engineer the resist with components that work well in a liquid immersion lithography system.

3.1.4. Evaluation through half-pitch

Using one of the better TOK resists (TOK-ILP03), the resist profiles were evaluated as the half-pitch was decreased until the resist was not resolving the dense line and space structure or having poor imaging capabilities. Figure 9 shows the resist as the half-pitch decreases from 87 nm in air to 45 nm in HPLC H_2O. The dense line-space profiles of the images are square and well defined down to a 50 nm half-pitch in HPLC H_2O. As the half-pitch is reduced to 45 nm the dense lines-spaces exhibit severe micro bridging and pattern collapse. 50 nm lines and spaces is about the usable resolution limit of the TOK-ILP03 resist under the present exposure and process conditions.



87 nm (Air) | 71 nm (H_2O) | 64 nm (H_2O)

64 nm (H_2O) | 50 nm (H_2O) | 45 nm (H_2O)

Figure 9 – TOK-ILP03 resist exposures through half-pitch.

3.1.5. Impact of a rinse surfactant

One of the problems with imaging 65-nm half-pitch lines is that at aspect ratio larger than 2.5 there is frequently complete pattern collapse. This is due to surface tension of water drying unevenly between the lines after the development rinse. The simplest way to reduce this pattern collapse is by reducing the surface tension of water via adding a surfactant to the water. At a 65-nm half-pitch pattern collapse due to adhesion tension will not occur if the resist thickness is at or below 100 nm, however at 150 nm and larger there is often pattern collapse due to adhesion tension.

Figure 10 shows images of 65-nm half-pitch dense lines exposed in air and DI H_2O with and without a rinse surfactant for both 100-nm and 150-nm thick lines in TOK-ILP03 resist. There was no pattern collapse in any of the images. Based on other experience at these with these resists, collapse was expected that the 150-nm thick lines exposed in air and DI H_2O without the rinse surfactant would show pattern collapse due to water adhesion tension during wafer drying. A much more controlled investigation of pattern collapse limits at these scales is needed to gain a fuller understand of the limitations and control strategies.

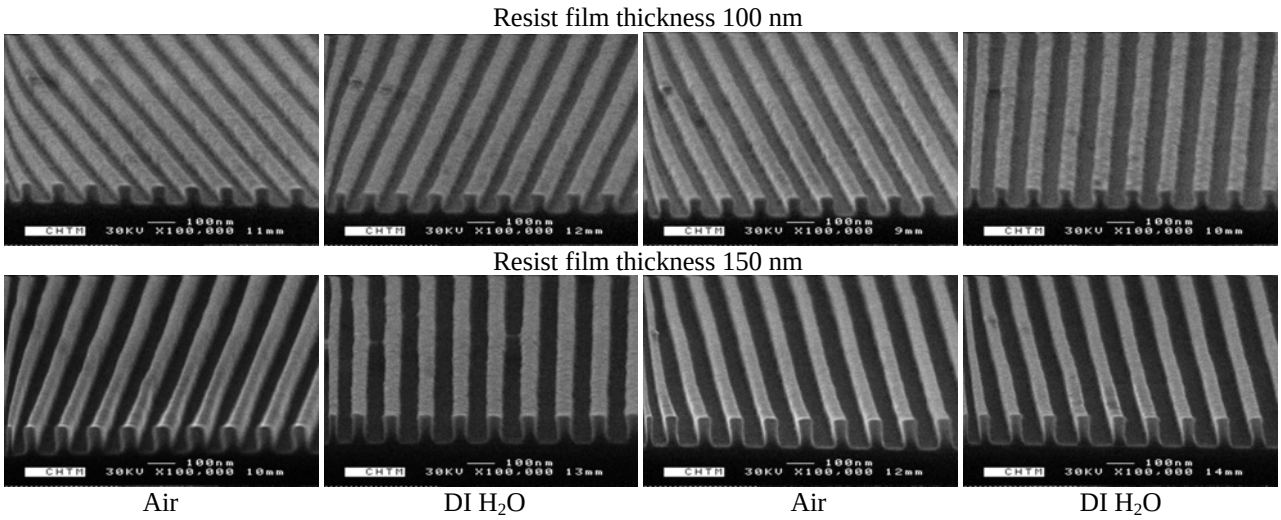


Figure 10 – 129-nm pitch TOK-ILP03 resist exposed with and without a rinse surfactant.

3.2. AZ Clariant resist liquid immersion imaging results

The AZ Clariant provided eight resist formulations. Four of the resists used a COMA hybrid polymer T2030 backbone and the remaining four resists used an acrylic polymer T518 backbone. Figure 11 shows the chemical composition of the two polymers used in the AZ Clariant resists. Half of the resists used a large molecule TPSNF PAG similar to the TOK resist, and the other half used a small TPSTf PAG. AZ Clariant also used two base quenchers, a large molecule trioctylamine base and a small diethanolamine base. Table 2 summarizes the matrix of AZ Clariant resist formulations. All resists were spun on a 32-nm thick Brewer Science ARC28-4 anti reflective coating that was optimized to give a minimum substrate reflection for 65-nm half-pitch dense patterns.

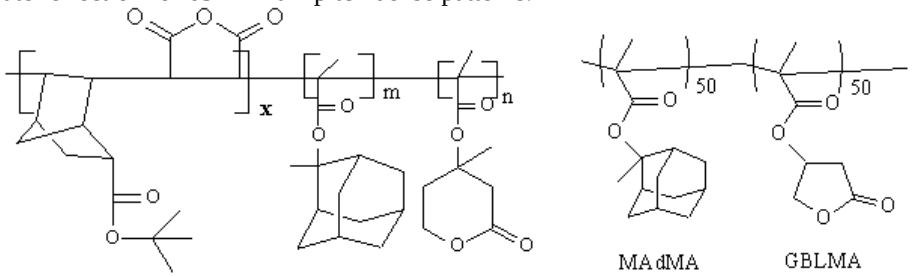


Figure 11 – AZ Claraint polymer backbone chemical composition.

PAG		TPSTf (Ps)		TPSNf (PL)	
Base Quencher		Diethanolamine (Bs)	Trioctylamine (BL)	Diethanolamine (Bs)	Trioctylamine (BL)
Polymer Backbone	Hybrid polymer T2030	T2030 Ps Bs	T2030 Ps BL	T2030 PL Bs	T2030 PL BL
	Acrylic polymer T518	T518 Ps Bs	T518 Ps BL	T518 PL Bs	T518 PL BL

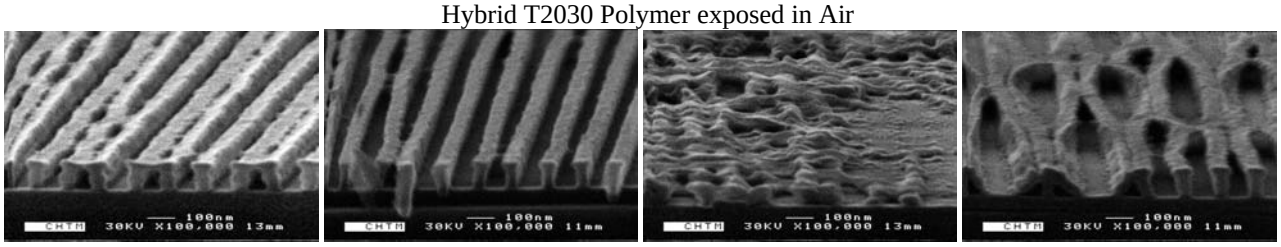
Table 2 – AZ Clariant resist formulation matrix

3.2.1. Image comparison of resist formulations

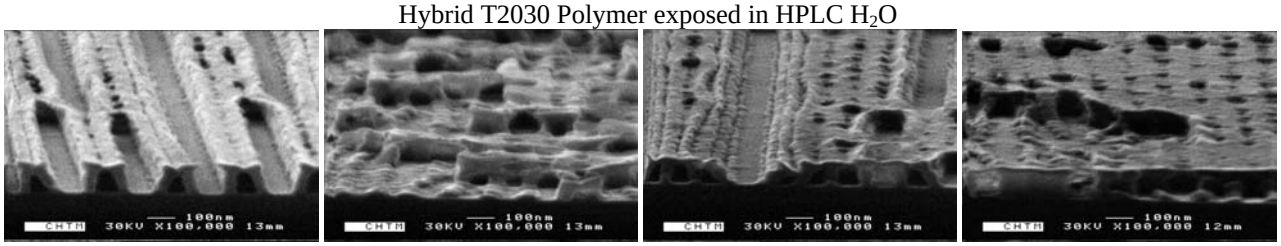
Figure 12 shows 65-nm half-pitch dense lines in 130-nm thick hybrid polymer T2030 resist. There were issues of pattern collapse at this resist thickness, and no rinse surfactant was used. In air there was noticeable T-toping of the line

profiles, with the large-PAG, small-base resist formulation having imaging issues. If the pattern collapse is neglected the smaller PAG gave better profiles for the air exposures. In the HPLC H_2O exposure, the small-PAG resist formed a skin and once the developer penetrated this surface layer, the remaining resist lines were completely developed away. This suggests that the small PAG is completely leaching away from the surface of the resist into the water. Again ignoring the pattern collapse issues, the resist with the large PAG and large base gave the best results in HPLC H_2O for the hybrid T2030 polymer resist.

Hybrid T2030 Polymer exposed in Air



Hybrid T2030 Polymer exposed in HPLC H_2O

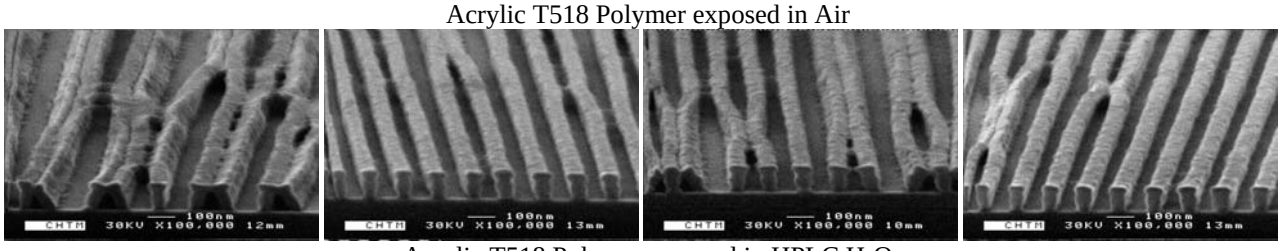


PL, BL Ps, BL PL, BL Ps, Bs

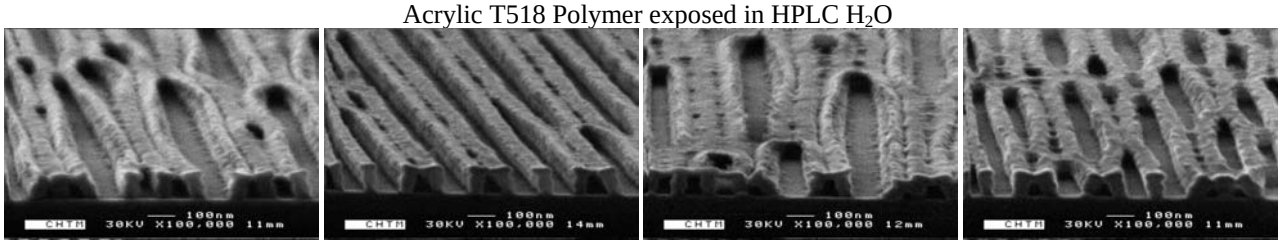
Figure 12 – 131-nm pitch AZ Clariant hybrid T2030 polymer resist LIL results.

Figure 13 shows 65-nm half-pitch dense lines in 120-nm thick Acrylic polymer T518 resist. Like the T2030 polymer resist there were issues of pattern collapse as the resist thickness was thicker than the desired 100 nm. The AZ Clariant resists all imaged well in air similar to the TOK Acrylic resist, ignoring the pattern collapse. The resist with the smaller PAG had reduced T-toping and gave better results in air. In the HPLC H_2O exposures, ignoring pattern collapse, the large PAG had slightly more T-toping then the small PAG, and no real noticeable difference between the small and large base resist images.

Acrylic T518 Polymer exposed in Air



Acrylic T518 Polymer exposed in HPLC H_2O



PL, BL Ps, BL PL, BL Ps, Bs

Figure 13 – 131-nm pitch AZ Clariant Acrylic T518 polymer resist LIL results.

From the AZ Clariant resist results it is clear that the acrylic resists work better for liquid immersion lithography. The size of the PAG made noticeable differences for both the hybrid and acrylic polymer resist. The best AZ Clariant resist

was the Acrylic polymer T518 with the small PAG and large Base. The Acrylic resist should be reevaluated at a thinner formulation that allows for 100 nm resist film thickness.

3.2.2. Evaluation through half-pitch

Using an improved Acrylic AZ Clariant resist (EXP M3000) spun at 100-nm thick, the resist profiles were evaluated as the half-pitch was decreased until the resist was not resolving the dense line and space structure or having poor imaging capabilities. Figure 14 shows the resist as the half-pitch decreases from 87 nm in air to 45 nm in HPLC H_2O.

87 nm (Air)	71 nm (H_2O)	64 nm (H_2O)
64 nm (H_2O)	50 nm (H_2O)	45 nm (H_2O)

Figure 14 – AZ Clariant EXP M3000 resist exposures through half-pitch.

3.2.3. Evaluation of contact holes

Using the same Acrylic AZ Clariant resist (EXP M3000) spun at 100-nm thick, dense contact holes were exposed in the resist and the profiles were evaluated. Contact holes are achieved by performing two back-to-back exposures where the sample was rotated by 90 degrees between the two exposures. Figure 15 shows the 65 nm dense contact holes array in the AZ Clariant resist.

Exposed in Air

Exposed in DI H_2O

Figure 15 – 129-nm pitch contact holes in AZ Clariant EXP M3000 resist.

4. SUMMARY

4.1. Conclusions

Immersion lithography reduces the minimum accessible pitch by a factor of the index of refraction *n* of the immersion medium. The work in this paper has demonstrated that 45-nm half-pitch lines can be patterned into resist using liquid immersion lithography (LIL) with a NA=1.18. This would allow for 193-nm lithography to be extended out at least 2 technology nodes.

Some but not all existing 193-nm resists work well in LIL. Based on these preliminary results, it seems likely that 193-nm LIL resists can be formulated with fully comparable performance to today's 193-nm air-exposed resists. The acrylic polymer resist platform performs well in LIL. It is also clear that PAGs and bases play an important role, although more studies are needed to find an optimal set of PAGs and bases for 193-nm LIL resist.

Resist top surface protection coatings exist that work well with 193-nm resist and water immersion liquid. From a imaging stand point a top surface protection coating is most likely not necessary, and just adds cost and complexity to the processing.

As LIL enables the printing of ever-smaller lines and spaces, the issue of pattern collapse due to adhesion tension during post develop wafer drying will become increasingly important. Rinse surfactants should reduced pattern collapse issues, however resist thicknesses likely will continue to decrease, posing issues for subsequent pattern transfer steps. The problem of pattern collapse needs further study; a new resist with larger Young's modulus will help combat these issues.

Based on all the work presented here there are no showstoppers to liquid immersion lithography. There are known solution to deal with the interactions between resist and the water immersion liquid. The resist can be engineered and developed for 45-nm half-pitch dense lines in LIL and beyond.

4.2. Future studies

Work will need to be continued to evaluate new 193-nm resists formulations for liquid immersion lithography. These resist will also need to be evaluated at and below the 45-nm half-pitch node to understand the true limits to the resist. The resist will also need to be evaluated for contact holes, as resist performance tends to differ between contact holes, dense lines and isolated lines.

Detailed studies will need to be performed on pattern collapse. The limit of resist aspect ratios for a given resist system needs to be studied at the 65-nm and 45-nm nodes. Evaluation of rinse surfactants and possible developer surfactant needs to be performed at the LIL technology nodes.

Polarization effects are going to become very important on hyper NA systems. These effects need to be quantified on an LIL system and their corresponding impacts on resist imaging evaluated.

Currently purified water with a refractive index of 1.44 has been identified as the preferred liquid immersion medium. The resolution limits of 193-nm lithography can be further improved if a viable new liquid is found with an even higher refractive index. Investigation for new liquid immersion media needs to be continued.

A hyper NA imaging system needs to be developed and demonstrated in LIL. Evaluation of the improved depth of focus (DOF) and process latitude needs to be performed. New exposure strategies will be needed for high NA imaging. Polarization sensitivity will impact existing reticle enhancement techniques, and may require new strategies that involve separation high spatial frequencies in the orthogonal *x*- and *y*- directions. Arbitrary 65-nm and 45-nm patterns must be evaluated in the new 193-nm immersion resist.

REFERENCES

1. C. V. Shank, R. V. Schmidt, “Optical Technique for Producing 0.1-μm Periodic Surface Structures”, Applied Physics Letters, Vol. 23 No. 3, Aug. 1973, p154.

2. W. T. Tsang, S. Wang, “Simultaneous Exposure and Development Technique for Making Gratings on Positive Photoresist”, Applied Physics Letters, Vol. 24 No.4, Feb. 1974, p196.

3. J.A. Honffnagle, W. D. Hinsberg, M. Sanches, F. A. Houle, “Liquid Immersion Deep-Ultraviolet Interferometric Lithography”, J. Vac. Sci. Technol. B, Vol 17 No.6, Nov/Dec 1999, p3306-3309

4. M. Switkes, and M. Rothschild, “Immersion Lithography at 157nm”, J. Vac. Sci. Technol. B, Vol 19 No. 6, p2353-2356, Nov/Dec 2001.

5 A. K. Raub and S. R. J. Brueck, “Deep UV Immersion Interferometric Lithography*,”* Proc. SPIE **5040**, Optical Microlithography XVI, 667-678 (2003).

6. John H. Burnett and Simon Kaplan, “Measurement of the refractive index and thermo-optic coefficient of water near 193 nm”, Proc of SPIE, Vol 5040, pg. 1742-1749.

7. Burn J. Lin, “The k3 coefficient in noparaaxial /NA scaling equations for resolution, depth of focus, and immersion lithography”, JM3, vol. 1 issue 1, pg. 7-12.

8. S. Zaidi and S. R. J. Brueck, “High Aspect-Ratio Holographic Photoresist Gratings,” Appl. Optics 27, 2999 (1988)

Simulation of the 45-nm Half-Pitch Node with 193-nm Immersion Lithography

Abani M. Biswas, Andrew Frauenglass and Steven R. J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106
Tel: (505) 272-7800; Fax: (505) 272-7801; e-mail: abiswas@chtm.unm.edu

ABSTRACT

The minimum half pitch available to a lithography tool is given by the well-known equation $\Lambda_{min\text{-}half.} = \lambda/4n\sin\theta$, where λ is the exposure wavelength, n is the refractive index of the immersion medium and 2θ is the angle between the propagation directions of the two beams. Using water (n = 1.44) as an immersion medium, the highest spatial frequency available with ArF-based (193 nm) lithography tools with an *NA* of 1.3 (1.44×0.9) corresponds to a half-pitch of 37 nm suggesting that the 45-nm half-pitch node should be accessible. A detailed simulation study (PROLITH™ 8) is reported for two different approaches to printing for this node. Dipole illumination (with two masks) as well as imaging interferometric lithography (with a single mask and multiple exposures incorporating pupil plane filters) is shown to be capable of printing arbitrary structures under these conditions. Because of the loss of contrast for TM-polarization at the high spatial frequencies at this node the high spatial frequencies in the *x*- and *y*-directions need to be printed with different polarizations in order to retain the necessary contrast. This, in turn, will require modification of the illumination system or a multiple exposure approach to allow the necessary polarization control.

1. INTRODUCTION

The current industry roadmap calls for initial volume manufacturing of the 45-nm half-pitch node in the 2005-2007 time frame, which in turn demands an immediate decision on the manufacturing technology. It now appears that next generation lithography techniques will be hard pressed to meet this time-line and attention has turned to extensions of optical lithography. Recently, water immersion lithography[1-6] at 193 nm has emerged as the industry consensus for the next major optical lithography extension. The minimum half pitch available to a lithography tool is given by the well-known equation,

$$\Lambda_{min\ half.} = \lambda/\ 4n\sin\theta \qquad (1)$$

where λ is the exposure wavelength, n is the refractive index of the immersion medium and the two beams interfering to create a sinusoidal pattern at the resist are propagating in the water at angles of ±θ with respect to the wafer surface normal. For 193-nm exposures with an immersion medium of water (n = 1.44)[7] and a maximum $\sin\theta$ of 0.9, this minimum half-pitch is 37 nm, suggesting that a system with these capabilities should be capable of printing arbitrary 45-nm half-pitch patterns.

At the high angles in the resist that these dimensions and wavelength imply, there are significant differences in the resist response to TE and TM polarizations; in particular the TM polarization aerial image loses contrast and even inverts polarity (exchanging lines and spaces). To circumvent these adverse effects, the high spatial frequencies in each direction need to be printed with optimized (orthogonal) polarizations. This, in turn, will require modification of the illumination system or some form of multiple exposure lithography to allow the necessary polarization control. Two imaging schemes[8] that lend themselves to this constraint are imaging interferometric lithography[9,10] (IIL), which uses off-axis illumination at the limits of the pupil along with pupil-plane filters to control the image fidelity, and dipole illumination, where different masks and off-axis illumination directions are used for the high spatial frequency content in the *x*- and *y*-directions of a Manhattan geometry pattern. Both of these techniques offer independent control of the polarization for the x- and y-direction exposures. In section 2, the polarization effects in immersion lithography are discussed. Section 3 presents modeling results for partial coherent illumination, IIL, and dipole illumination for a simple nested-ell test structure. Finally conclusions and future research directions are outlined in Section 4.

Proc. SPIE **5377**, 1579-1586 (2004).

2. POLARIZATION EFFECTS IN IMMERSION LITHOGRAPHY

The polarization effects can readily be understood in terms of a simple interferometric lithography (IL) exposure with beams incident on the wafer from the extremes of the lens pupil. It is straightforward to calculate the interference pattern for two–beam IL with equal intensity beams with TE- and TM-polarizations.

$$I_{TE}(x) = 2I_0\left[1+\cos\left(\frac{4\pi n_{resist}\sin\theta_{resist}}{\lambda}x\right)\right] \qquad (1)$$

$$I_{TM}(x) = 2I_0\left[1+\cos(2\theta_{resist})\cos\left(\frac{4\pi n_{resist}\sin\theta_{resist}}{\lambda}x\right)\right]$$

where I_0 is the intensity of each beam and the subscripts emphasize that the refractive index and internal angles within the resist are used in evaluating these expression. The TE result retains full contrast over the entire accessible half-pitch range independent of θ_{resist}. The additional factor of $\cos(2\theta_{resist})$ that appears in the TM expression has a dramatic impact on the aerial image. For $\theta_{resist} = 45°$, the contrast of the interference pattern goes to zero; at higher angles (smaller half-pitches) there is a phase shift so that lines and spaces are interchanged. These polarization effects have not yet become a significant issue with current lithography tools. Unpolarized illumination, which averages the effects of the two possible polarizations, is adequate for most imaging. This is because the immersion fluid today is air ($n = 1$) while the index of a typical resist is considerably higher ($n_{resist} \sim 1.7$). As a consequence of Snell's law, the propagation angles in the resist are considerably smaller than in the air. For a period corresponding to the 90-nm half-pitch node, $\sin\theta_{air} \sim 0.54$, the propagation angle in air is 32° so θ_{resist} is only 18° and the corresponding contrast for TM polarization is 84%. However, for the 45-nm half-pitch node, the propagation angle of 48° in the immersion fluid corresponds to $\theta_{resist} \sim 39°$ where the contrast for TM polarization is only 21%!

This is illustrated in Fig. 1(a), which shows the intensity distribution versus position in the resist for periods corresponding to the 90-, 65- and 45-nm half-pitch nodes. The top trace is for TE polarization that retains ideal contrast independent of the pitch. The TM curves for the 90-, 65- and 45-nm half-pitch nodes are also shown. Figure 1(b) shows the TM visibility, $(I_{max} — I_{min})/(I_{max} + I_{min})$, as a function of the half pitch. For the 90-nm node, the TM visibility is 84%, dropping to 62% for the 65-nm node and 21% for the 45-nm node. The line-edge roughness is strongly dependent on the visibility[11] so that this loss of contrast impacts the quality of the image in several ways. The immersion fluid also impacts the reflection at the liquid-photoresist interface, generally offering partial antireflection coating and improving the energy transfer across the interface.

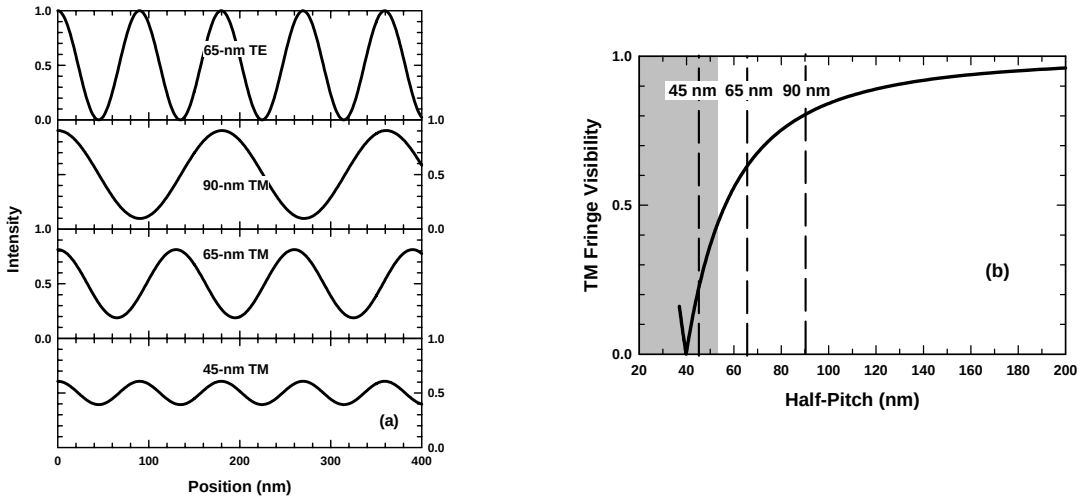


Figure 1: (a) Intensity profiles in resist for 65-nm half pitch TE polarization (top) with full contrast, and for 90-, 65- and 45-nm half-pitch patterns with TM polarization. The decrease of contrast is severe. (b) Visibility of the TM polarization pattern in the resist as a function of half-pitch. The shaded region is only accessible with immersion.

3. IMMERSION LITHOGRAPHY SIMULATIONS FOR ARBITRARY PATTERN

Figure 2 shows the mask pattern that has been used in all of the simulations reported here. It consists of a series of five nested "ells" at a 45-nm half-pitch CD and a large box at 10×CD. For the nested features, the middle feature extends beyond the others (a different amount in the *x*- and *y*-directions) to give some indication of dense-isolated biases. For imaging interferometric lithography, this mask is used without any OPC or other RETs. For dipole illumination, the features are split into two separate masks as described below. Simulations were carried out with PROLITH™ Version 8 (KLA-Tencor) in full vector electro-magnetic mode and with the full physical model for the photoresist (PEB diffusion, 3D etching). In all cases, the simulations used the same film stack. Substrates were bare silicon with a 48-nm thick bottom anti-reflection coating (BARC) ($n = 1.7059 + i0.335$) and a 50-nm thick, positive photoresist ($n = 1.703 + i0.0172$; Dill parameters $A = 0$, $B = 1.122\ \mu m^{-1}$, $C = 0.068\ cm^{-2}$ /mJ). Standard post-exposure bake (110°C for 60 s) and develop conditions (60 s) were used for all simulations. As the previous section indicated, separating the exposures of the high spatial frequencies in the *x*- and *y*-directions and using orthogonal polarizations for the two cases, so that all exposures are dominantly TE polarization, will be key to imaging at the limits of immersion lithography. The simulations presented below bear this out. In the following sections, simulation results for two approaches to imaging the 45-nm half-pitch node at 193 nm using water immersion are presented.

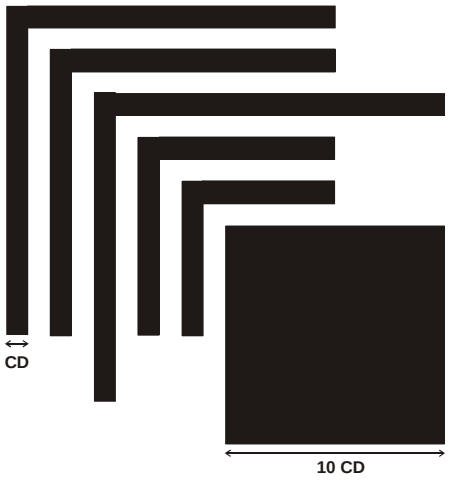


Figure 2: Test pattern used for the simulations

The highest spatial frequencies possible in an optical system correspond to two beams at opposite ends of a pupil diameter, mimicking 2-beam interferometric lithography, e.g. to coherent illumination ($\sigma = 0$) with the incident beam tilted to one edge of the pupil and the high frequency diffracted beam from the mask captured at the opposite pupil edge. Partial coherence necessarily degrades this resolution as it involves beams with lower tilt that do not capture the highest frequencies and lower the overall image transfer function. Imaging interferometric lithography (IIL), essentially off-axis illumination at the edge of the pupil along with imaging pupil-plane filters to ensure uniform weighting of all spatial frequencies is discussed first. Second is dipole illumination at the edges of the pupil, with two masks that separate out the high *x*- and high *y*-spatial frequency parts of the image. In both cases the polarization is rotated between exposures to ensure predominately TE polarization exposures.

3.1 Imaging interferometric lithography

Imaging interferometric lithography (IIL) was introduced as a technique to cover the extremes of frequency space by tilting the illumination as far as possible off-axis. With the modest NA's of ~ 0.5 to 0.65 that were available at the time, this required tilts that were beyond the limits of the imaging pupil and consequently reintroduction of a zero-order beam on the image side of the lens. With today's very high NA lenses of $\sin\theta \sim 0.9$, the gain for the significant added optical complexity of building interferometers around the lens is small. With transmission of the zero-order beam through the imaging optics, IIL is similar to quadrupole off-axis illumination with a very small partial coherence for each off-axis illumination beam, and with the quadrupole oriented along the principal (*x*, *y*) axes of a Manhattan geometry pattern to maximize the spatial frequency capture. The polarization of each of the off-axis illumination beams can be independently controlled to optimize the imaging. Quadrupole off-axis illumination inherently emphasizes low frequencies at the expense of high frequencies[10,12] because each of the exposures includes the same low-frequency information while the individual high frequencies are covered only in selected exposures. Pupil plane filters eliminate these multiple coverages and provide an improved image, close to the theoretical frequency space limits. Thus, IIL for $\sin\theta \sim 0.9$ lenses becomes quadrupole OAI with tilts to the edges of the imaging pupil along with pupil-plane filters. Figure 3 shows the result of this procedure for the 45-nm half-pitch pattern without using Phase Shift Mask (PSM) or any OPC. The pattern, Fig. 3(a), is well resolved. The impacts of the limited resolution include rounding of the line edges and some line edge-ripple. The pupil plane filters used in this simulation are shown in Fig. 3(b).

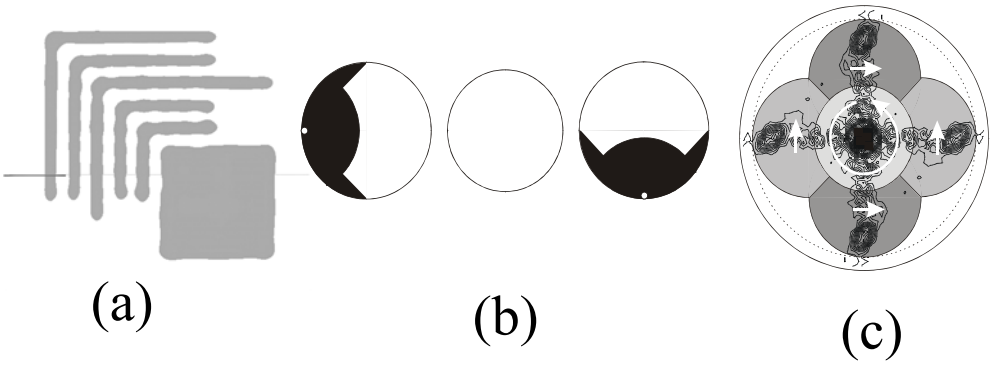


Figure 3: Modeling of IIL. (a) gives the final real-space image. (b) shows the pupil filters used in the off-axis exposures. (c) shows the resulting frequency-space coverage along with the intensity Fourier transform of the desired mask pattern. The gray shadings are simply to denote the frequency coverage, the effective MTF for the linear imaging terms is uniform across the image. The polarization of the individual exposures is indicated.

In the simulation, two offset exposures from opposite directions are included to restore a telecentric response. The polarization is controlled in each of the exposures as indicated in Fig. 3(c) to ensure a dominant TE polarization, especially for the highest spatial frequencies. Figure 3(c) shows the resulting spatial frequency coverage with the intensity Fourier transform of the image superimposed. The different gray levels in this figure are simply to distinguish individual exposure coverages. The effective modulation transfer function for the linear imaging terms is uniform across the entire accessible frequency space. The impact of polarization control is shown in Figure 4. This result is for the same pupil filters as the result of Fig. 3, the only difference is that a fixed polarization has been used for all exposures. The impact of having the "wrong" polarization for the *x*-directed, high-spatial frequencies is clear.

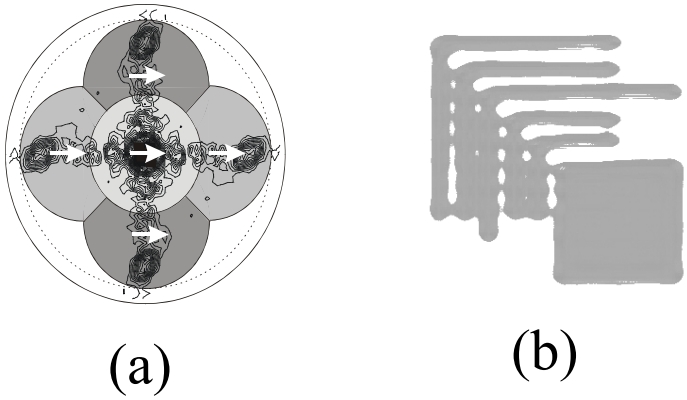


Figure 4: Exposures identical to Fig. 3 with except for the polarization. The impact of using the "wrong" polarization for the *x*-directed high-spatial frequencies is clear.

Symmetric results were obtained if *y*-directed polarization is used for all exposures, the opposite parts of the ell did not print. This result confirms the simple grating analysis given above – it will be critical to independently optimize the polarization for the high spatial frequencies involved in printing 45-nm node patterns by immersion with a 193-nm source. The impact of defocus on the IIL pattern is shown in Figure 5. At a 80-nm defocus, the pattern is largely maintained. At a 100-nm defocus, dense/isolated bias is evident in the line-end shortening and thinning of isolated features. Finally, at a 130-nm defocus the lines begin to merge together and a hole has developed in the large box.

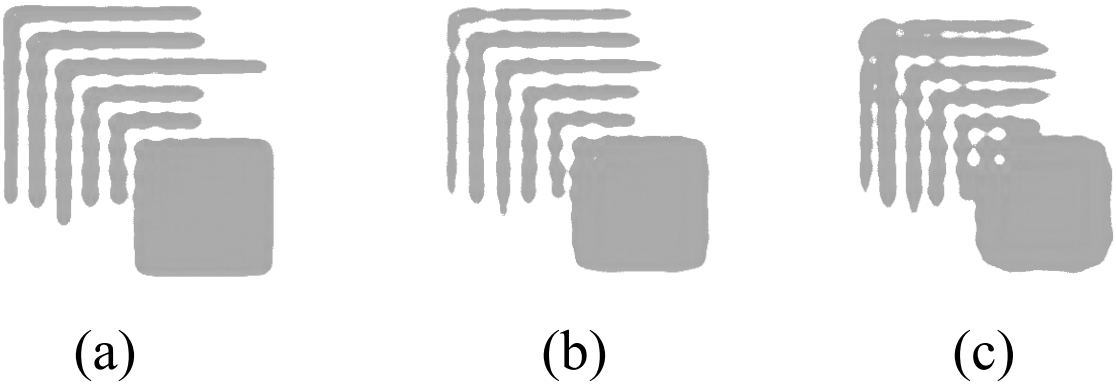


Figure 5: Evolution of the IIL image of Fig. 5 though focus showing (a) 80 nm out of focus (b) 100 nm out of focus and, (c) 130 nm out of focus. The depth-of-field is of order 160 nm.

The addition of OPC should extend the depth-of-field (DOF) which is ~ 160 nm from these simulations. It is important to recognize that the large off-axis angles of IIL (or dipole illumination, below) maximize the DOF compared with conventional on-axis illumination for these high spatial frequency structures. In the limit of a periodic structure where the off-axis illumination is chosen to symmetrize the zero- and first-order diffracted beams on the image pupil, *i. e.* interferometric lithography, the DOF approaches infinity since both beams are inclined at the same azimuthal angle (θ). The frequently cited Rayleigh result ($DOF \propto \kappa_2 / \lambda\, NA^2$) does not hold for this situation.

The importance of incorporating the pupil-plane filters is shown by the simulation result in Fig. 6 for the same pattern using only the offset exposures but without using any pupil plane filters. The dense features are not completely resolved and there is a significant dense-isolated bias. The frequency space coverage shown in Fig. 6(b) shows that the low frequency regions of the pattern are included in multiple exposures while the high frequency components are only counted once. The resulting overemphasis on the low frequencies is the reason for the inability to resolve the dense features. This exposure strategy is equivalent to dipole illumination, but with a single mask with all of the features included. The next section will show that somewhat better results are available by parsing the features between two masks

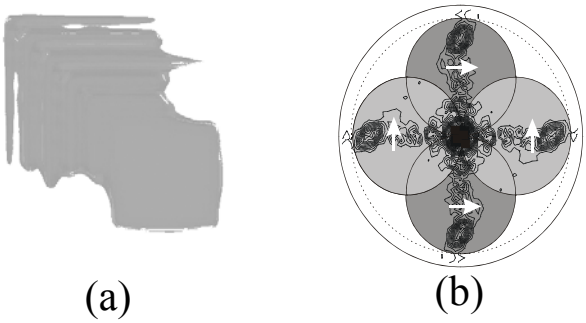


Figure 6: (a) Result of IIL exposures without pupil plane filters. The dense features are not resolved. (b) Corresponding frequency space coverage. The multiple coverage of the low frequency regions leads to the lack of resolution.

3.2 Dipole illumination

Dipole illumination, where the mask is split into two sub-masks each with the high spatial frequencies in one direction is an alternative exposure technique that offers the necessary polarization flexibility. The wide range of spatial frequencies in the present test pattern poses some difficulties for this approach. Figure 7 shows an example of coherent ($\sigma = 0$) dipole illumination at an offset of 0.9 just as was used for the IIL example above.

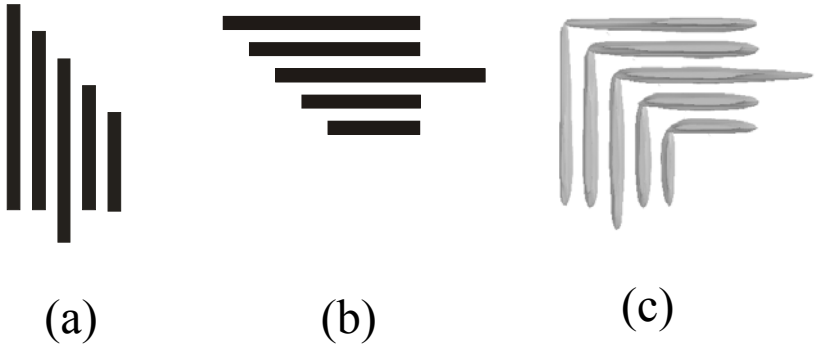


Figure 7: Dipole illumination with two masks (a) and (b) separating the high small features in the *x*- and *y*-directions. The large box was eliminated, since it was not possible to print the pattern when it was present. The illumination offsets were at the edge of the imaging pupil and the frequency space coverage was identical to that shown in Fig. 6. Because the large box was absent, the frequency space overlaps were tolerable.

The large box has been removed from the pattern and the remaining features split between separate masks for the *x*- and *y*-directions as shown. The polarization is optimized separately for exposure of each mask. The dense/isolated bias is more pronounced in this configuration and the corners of the ells are less robustly defined, even though these regions were left unexposed in both masks. Alignment of the two masks between exposures, and precise control of line-edge placement are added requirements for this technique. Figure 8 shows the result of a single mask exposure including the large box (left column). The individual lines close to the box are not resolved. The two right columns show two different dark field (inverted image) mask strategies.

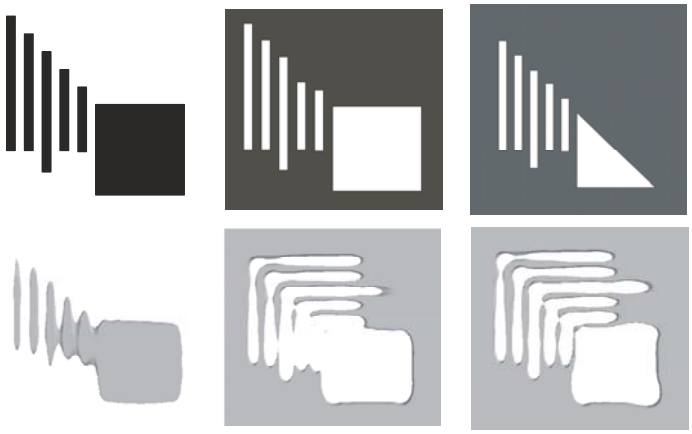


Figure 8: Results of three different dipole mask strategies. A single, bright-field mask printing only the lines parallel to the *y*-axis and the large box. It was not possible to print both the lines and the box in a single exposure. Also shown are two different attempts at dark field (inverted) images. The second mask for the first case (full large box) included only the lines parallel to the *x*-axis with no box. The second mask for the second case (triangular box segment) was symmetric to the mask shown.

In the center column, the mask includes the entire box while the second mask (not shown) has only the dense features parallel to the x-axis. Again in this case there is difficulty in separating the dense lines from the box, and there is significant distortion of the pattern. The right-hand column shows a more symmetric division of the mask. This produced the best-printed pattern, but still shows significant distortion and line-width variation. More detailed investigations are necessary to optimize these dipole illumination strategies. In particular, there is overcounting of the low frequencies in this case as well and pupil-plane filters may help the imaging.

4. CONCLUSIONS

A detailed simulation of arbitrary pattern lithography for the 45-nm half-pitch node using a 193-nm source with a *NA* = 1.3 water-immersion lithography tool has been presented. A requirement of this the physics of interference at high angles is to optimize the polarization individually for different parts of the image, which in turn requires a major change to illumination optics and will impact tool strategies and cost of ownership evaluations. In order to access the highest available spatial frequencies – necessary for printing 45-nm half-pitch structures, off-axis illumination near the edges of the imaging pupil will be required. Two strategies for separating the exposures to allow for the necessary polarization control have been presented. Imaging interferometric lithography, related to off-axis illumination with the addition of pupil plane filters to adjust the details of the image, provides relatively robust results. Different pupil plane filter strategies have been presented along with details of variation of the image through focus. Dipole illumination with the division of the pattern into two masks, each with patterns primarily oriented in orthogonal directions, was also investigated. Initial results are not as promising as the IIL strategy, but more investigation is required. These simulations show that 45-nm half-pitch lithography is indeed possible by extension of 193-nm tools to water immersion. Much more investigation is required to optimize the imaging over the range of patterns required for actual circuits, and to add OPC to provide more robust results. Experimental confirmation of the simulations using micro-stepper platforms is also an urgent need.

Acknowledgements: Support for this work was provided as part of the ARO/MURI in Deep Subwavelength Optical Nanolithography.

REFERENCES

[1] C. V. Shank and R. V. Schmidt, "Optical Technique for Producing 0.1-μm Periodic Surface Structures," *Appl. Phys. Lett*. **23**, 154 (1974).

[2] W. T. Tsang and S. Wang, "Simultaneous Exposure and Development Technique for making Gratings on Positive Photoresist," *Appl. Phys. Lett*. **24**, 196 (1974).

[3] J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, "Liquid Immersion Deep Ultraviolet Interferometric Lithography," *J. Vac. Sci. Technol*. **B17**, 3306 (1999).

[4] M. Switkes and M. Rothschild, "Immersion Lithography at 157 nm," *J. Vac. Sci. Technol*. **B19**, 2353 (2001).

[5] A. K. Raub and S. R. J. Brueck, "Deep UV Immersion Interferometric Lithography," Proc. SPIE **5040**, Optical Microlithography XVI, **,** 667-678 (2003).

[6] Alex K. Raub, A. Frauenglass, S. R. J. Brueck, Will Conley, Ralph Dammel, Andy Romano, Mitsuru Sato and William Hinsberg, "Deep-UV Immersion Interferometric Lithography," this volume

[7] John H. Burnett and Simon Kaplan, " Measurement of the refractive index and thermo-optic coefficient of water near 193 nm" , Proc. SPIE **5040**, Optical Microlithography XVI, 1742-1749 (2003).

[8] A. Biswas and S. R. J. Brueck, "Simulation of the 45-nm Half-Pitch Node with 193-nm Immersion Lithography," Jour. Microlithography, Microfabrication, and Microsystems **3**, 35-43 (2004).

[9] X. Chen and S. R. J. Brueck, "Imaging Interferometric Lithography - Approaching the Resolution Limits of Optics," *Opt. Lett.* **24**, 124-126 (1999).

[10] S. R. J. Brueck and X. Chen, Spatial Frequency Analysis of Optical Lithography Resolution Enhancement Techniques," *Jour. Vac. Sci. Technol*. **B17**, 908-921 (1999).

[11] W. Hinsberg, F. A. Houle, J. Hoffnagle, M. Sanchez, G. Wallraff, M. Morrison and S. Frank, "Deep-ultraviolet interferometric lithography as a tool for assessment of chemically amplified photoresist performance," *J. Vac. Sci. Technol*. **B16**, 3689 (1998).

[12] Xiaolan Chen and S. R. J. Brueck, "Experimental Comparison of Off-Axis Illumination and Imaging Interferometric Lithography," *Jour. Vac. Sci. Technol*. **B17**, 921-929 (1999).

Extension of 193-nm immersion optical lithography to the 22-nm half-pitch node

Steven R. J. Brueck and Abani M. Biswas
Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106
Tel: (505) 272-7800; Fax: (505) 272-7801; e-mail: brueck@chtm.unm.edu

ABSTRACT

Initial volume manufacturing of the 16- to 22-nm half-pitch integrated circuit node is targeted for the year 2018. Lithography is under tremendous pressure to extend its capabilities to meet this deadline. Recently, immersion lithography, particularly using water as an immersion fluid at 193 nm, has attracted much attention as a promising optical lithography extension. However based on simple optical bandwidth considerations, 193-nm-based optical lithography alone will not have the bandwidth necessary for printing the 22-nm half-pitch mode with any foreseeable combination of immersion liquids and conventional resolution enhancement techniques (RETs). The approach to reaching this node presented here is to combine all available RETs with processing nonlinearities and spatial-frequency doubling using two photoresist layers and an image storage layer. Appropriate combinations of multiple exposures/processes can access the 22-nm node; thus reaching current end-of-roadmap values for half-pitch while retaining the current 193-nm lithography infrastructure. A detailed simulation (PROLITH^{TM} 8) study of this approach is reported.

1. INTRODUCTION

The relentless demands of Moore's law on lithography are well known. The semiconductor industry has been aggressively moving to print ever-decreasing feature size, imposing tremendous pressure on lithography to extend its present capability. By 2018, at the culmination of current industry roadmap, volume manufacturing of the 16- to 22-nm half-pitch node will be needed.[1] Immersion lithography, particularly using water as an immersion fluid at 193 nm, has recently emerged as the industry consensus for the next major optical lithography extension.[2-5] Taking advantage of a water (n = 1.44 at 193 nm)[6] immersion lens with $NA = n\sin\theta$ of up to 1.3 (1.44×0.9), the 45-nm half-pitch node will be directly accessible.[7] The resolution limits of optical lithography are often discussed in the context of the famous Raleigh criterion:

$$CD = \kappa_1 \frac{\lambda}{NA} \qquad (1)$$

where CD is the critical dimension or smallest printed feature (line or space). The parameter κ_1 depends on process specifications, but there is a hard limit at a κ_1 of 0.25 (assuming equal lines/spaces) for a single photoresist layer, corresponding to the highest spatial frequency of $2NA/\lambda$ available from the optical system. Resolution enhancement techniques (RETs)[8] including: phase shift masks (PSM), off-axis illumination (OAI), optical proximity correction (OPC), and imaging interferometric lithography (IIL) can be used to reduce the value of κ_1, approaching the theoretical limit of 0.25. These RETs are not mutually exclusive and can be combined to address the image specific issues. For water immersion this gives a limiting spatial frequency response of 37 nm, validating the claim that the 45-nm half pitch node should be accessible for these parameters. As shown below for a substantial extrapolation to an index of 2, simple scaling gives a limiting pitch of 26 nm, allowing sufficient bandwidth for the 32-nm half-pitch node. Continuing this scaling, an index of 2.9 is required for the 22-nm half-pitch node. These are clearly aggressive goals for the liquid index, and alternative techniques for enhancing resolution need to be investigated in parallel.

In this paper, we present an approach to extend optical lithography beyond the 22-nm node by combining process nonlinearities with a spatial frequency doubling step.[9,10] The basic concept is based on the fact that the fundamental single-image limit in optical lithography is on pattern period (e.g. on pattern density), not on the individual CDs (linewidths). As a result of nonlinear processes in the exposure, development and process transfer steps, CDs can be much smaller than half the smallest available period. As an example, oxygen plasma thinning of photoresist lines is

Proc. SPIE **5377**, 1315-1322 (2004).

already an accepted process in advanced microprocessor manufacturing. If the results of such a nonlinear imaging process are stored in a sacrificial layer on the sample, and a second similar imaging step is performed and interpolated with the first image, a factor of two extension in the maximum available spatial frequency is achieved. In principle, this nonlinear interpolation can be carried out multiple times, leading to the conclusion that there are no fundamental limits to optical lithography. In practice, both the image fidelity demands and the image placement requirements are increased, leading to the obvious corollary that there are increasingly challenging process and manufacturing issues that need to be addressed.

Simulations were carried out with Prolith™ Version 8 (KLA-Tencor) in full vector electromagnetic mode and with the full physical model for the photoresist (PEB diffusion, 3D etching, etc.). Substrates were bare silicon with a bottom anti-reflection coating (BARC) (n = 1.7059 + i0.335) and a positive photoresist (n = 1.703 + i0.0172; Dill parameters A = 0, B= 1.122 μm^{-1}, C = 0.068 cm^{-2} /mJ). Standard post-exposure bake (110°C for 60 s) and develop conditions (60 s) were used for all simulations.

2 IMMERSION IMAGING INTERFEROMETRIC LITHOGRAPHY (I^3L) TO THE 32-NM HALF-PITCH NODE

Imaging interferometric lithography (IIL) is a relatively new concept[11] for accessing the fundamental linear systems resolution limits of optics, which uses off-axis illumination at the limits of the pupil along with pupil-plane filters to control the image fidelity. I^3L at 193-nm with water as immersion fluid (NA = 1.3) will access the 45-nm half-pitch node.

New liquids are also being studied to extend the refractive index to 1.6 or beyond, compared with 1.44 for water. If we take the limit of practical optical systems as $\sin\theta = 0.9$ and therefore $NA = n_{fluid} \times 0.9$, it is relevant to ask what NA is required to reach the next, 32-nm half-pitch, node at 193 nm. PROLITH simulation results of arbitrary dense pattern with CDs down to 32-nm using IIL exposures at 193 nm assuming NAs up to 1.8 and corresponding spatial frequency coverage are shown in the Fig. 1. Fig. 1(a) shows the resulting spatial frequency coverage of five-exposure IIL with the limiting frequency of $2NA/\lambda$. The polarization is controlled in each of the exposures as indicated in the figure to ensure a dominant TE polarization, especially for the highest spatial frequencies. The scaling is as expected, it takes a factor of 1.4 improvement in NA to print a factor of 0.7 smaller CD. Continuing this scaling to the 22 nm node requires a probably unrealistic liquid index of 2.9.

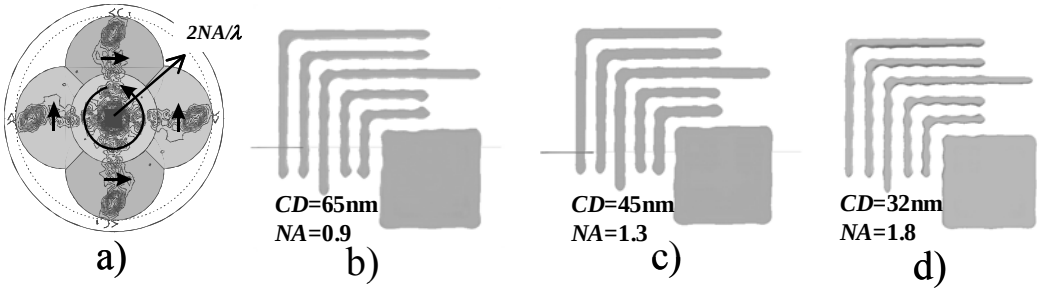


Fig.1 Modeling of IIL. (a) shows the pupil filters with frequency-space coverage. The gray shadings are simply to denote the frequency coverage, the polarization of the individual exposures is indicated. (b), (c), and (d) show the final real-space images.

3 EFFECT OF LITHOGRAPHIC PROCESS NONLINEARITIES ON THE RESOLUTION

The minimum half pitch available to a lithography tool is given by the well-known equation $\Lambda_{\text{min half}} = \lambda / 4n\sin\theta$, where λ is the exposure wavelength, n is the refractive index of the immersion medium and θ is the interfering angle of the two beams. These limits are on the half pitch, not on the CD. Using nonlinear properties of the photoresist and post-development processing, which is already standard industry practice, much smaller CDs can be realized[9].

The higher harmonics added by the process nonlinearities extend optical lithography beyond the linear system limits imposed by diffraction. However, process nonlinearities do not increase the pattern density. Since the harmonics are in phase with the fundamental components, they do not lead to increased density in the printed pattern and the effects are limited to the reduction of linewidth and to improved sidewall slopes. Figure 2 shows a sinusoidal aerial image and corresponding developed photoresist pattern.

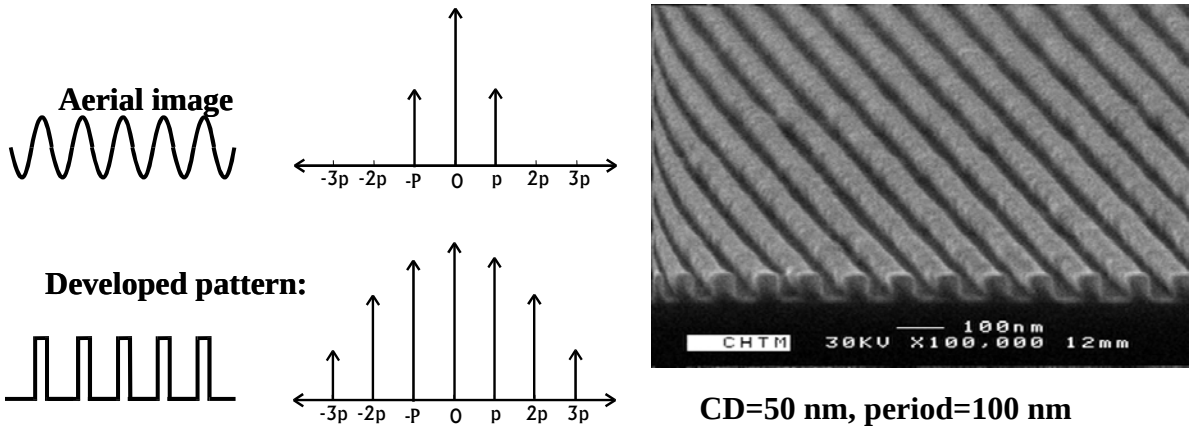


Fig. 2 Resist nonlinearity adds higher harmonic spatial frequencies. The developed photoresist pattern (immersion result at 213 nm courtesy of A. Raub; paper 26 this volume) exhibits nearly vertical sidewalls, corresponding to this higher frequency content.

A one dimensional example of the spatial frequency-doubling concept is illustrated in Fig.3. The top row of Fig.3 shows a photoresist pattern on top of a sacrificial layer above a substrate. The period of the pattern is limited by optics to $\lambda/2NA$; the CD, on the other hand, is not limited by the optics and is chosen to be less than $\lambda/8NA$. This can be achieved by manipulating the higher harmonic content of the image and does not violate any optical limits. In Fig. 3b the photoresist pattern has been transferred into the sacrificial layer using standard pattern transfer. A second photoresist layer has been spun, exposed and developed as shown in the Fig. 3c.

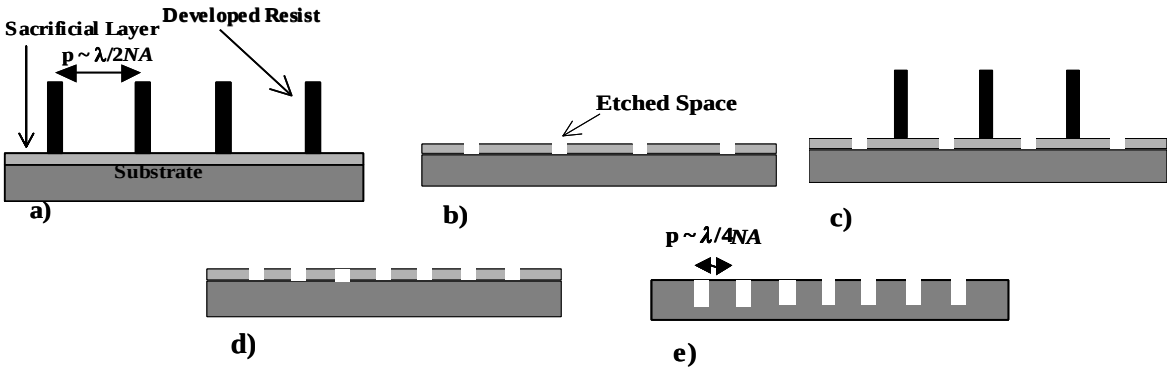


Fig. 3 Sequence of steps to double the 1D pattern density by spatial period division

The period and CD are the same as in the first case, but the pattern has been shifted by half the pitch ($\lambda/4NA$) to interpolate the lines of the second pattern midway between the lines of the first pattern. Finally, this pattern is also transferred into the sacrificial layer (Fig.3d). The sacrificial layer with the combined pattern is then used as a etch mask to transfer the total pattern into the substrate. Subsequently the sacrificial layer is removed leaving a final pattern with a pitch of $\lambda/4NA$, twice as dense as the single exposure optical limit.

For this simple case, the mathematics is trivial. The aerial image for a two-beam interferometric exposure at a period p is simply:

$$I(x) = 2\left[1 + \cos\left(\frac{2\pi x}{p}\right)\right] \tag{2}$$

Following the nonlinear processes in the exposure, development and pattern transfer the result is a simple Fourier cosine series:

$$P(x) = \mathcal{N}[I(x)] = \sum_n a_n \cos\left[\frac{2\pi n x}{p}\right] = \sum_n \left(\frac{\sin\left[\frac{\pi n a}{p}\right]}{\frac{\pi n a}{p}}\right) \cos\left[\frac{2\pi n x}{p}\right] \tag{3}$$

where the nonlinear transformation is indicated by $\mathcal{N}$ and a is the pattern CD. The final amplitude coefficients depend on the process variables. The addition of the second pattern, at the same period as the first but with a half-period, p/2, lateral shift is just:

$$\begin{aligned} P_{1\&2} = \mathcal{N}[I_1(x)] + \mathcal{N}[I_2(x)] &= \sum_n \left(\frac{\sin\left[\frac{\pi n a}{p}\right]}{\frac{\pi n a}{p}}\right) \cos\left[\frac{2\pi n x}{p}\right] + (-1)^n \sum_n \left(\frac{\sin\left[\frac{\pi n a}{p}\right]}{\frac{\pi n a}{p}}\right) \cos\left[\frac{2\pi n x}{p}\right] \\ &= \sum_{n'} \left(\frac{\sin\left[\frac{2\pi n' a}{p}\right]}{\frac{2\pi n' a}{p}}\right) \cos\left[\frac{4\pi n' x}{p}\right] \end{aligned} \tag{4}$$

where the $(-1)^n$ factor arises from the lateral shift of the second pattern. As a result of this factor, the odd terms in the original sin$x$/x summations cancel, the even terms survive and the result is a sin$x$/x expansion at twice the pitch – *e. g.* spatial-frequency doubling. Schematically, this result is shown in Figure 4. The Fourier coefficients are shown for a simple grating with lines parallel to the x-axis. The Fourier coefficients of the second exposure are identical

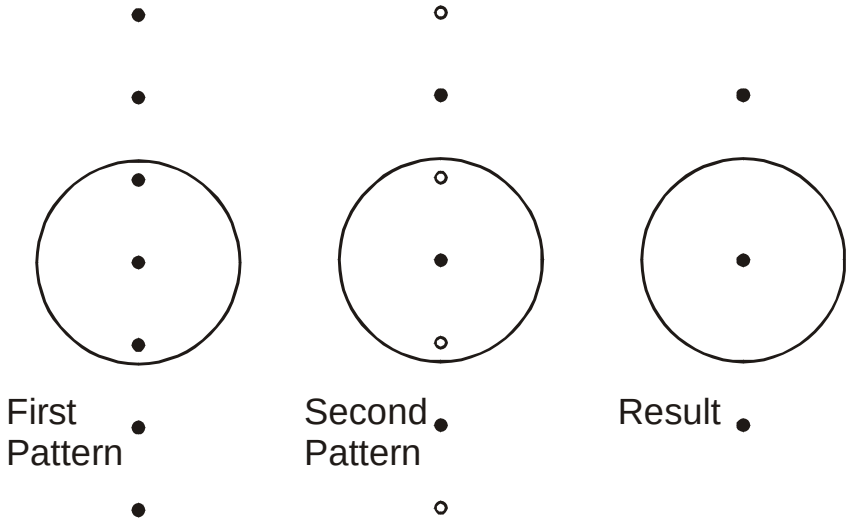


Figure 4: Graphical representation of a simple frequency doubling process.

except for the phase shift associated with the lateral translation. Unfilled circles indicate negative values. Finally the summation gives a grating at twice the spatial frequency. The circles with radius $2NA/\lambda$ indicate the frequency space limits of optics. Note that the final grating period cannot be produced by a single optical exposure since its fundamental frequency is greater than the optical system bandpass.

4 COMBINATION OF PROCESS NONLINEARITIES WITH RETs

The general concept of utilizing nonlinear processes in the multiple exposure and transfer method is to print a diffraction limited image in a single step; transfer this image to a sacrificial layer, and repeat this sequence one or more times accumulating a composite, interpolated image in the sacrificial layer which is subsequently used as an etch mask to transfer the complete pattern into the underlying circuit layers. A Prolith simulation of this process for the test pattern introduced above for a dense CD of 22 nm (λ/6n) with a water immersion *NA* of 1.3 and an exposure wavelength of 193 nm is shown in Fig.5. Two different masks are used with each having every other feature of the composite pattern. The smallest period for each exposure is 90 nm, corresponding to the single exposure frequency space limit at these parameters. In the modeling, the resist exposure dose is increased to give a CD of 22 nm, rather than the 45-nm CD for a equal line:space pattern. The mask CD is unchanged in this process, so the mask making requirements are not more difficult than for the 45-nm node. In the simulation, the sacrificial layer was simply the final photoresist pattern that was stored in the computer. In an actual process, a thin sacrificial layer such as a nitride

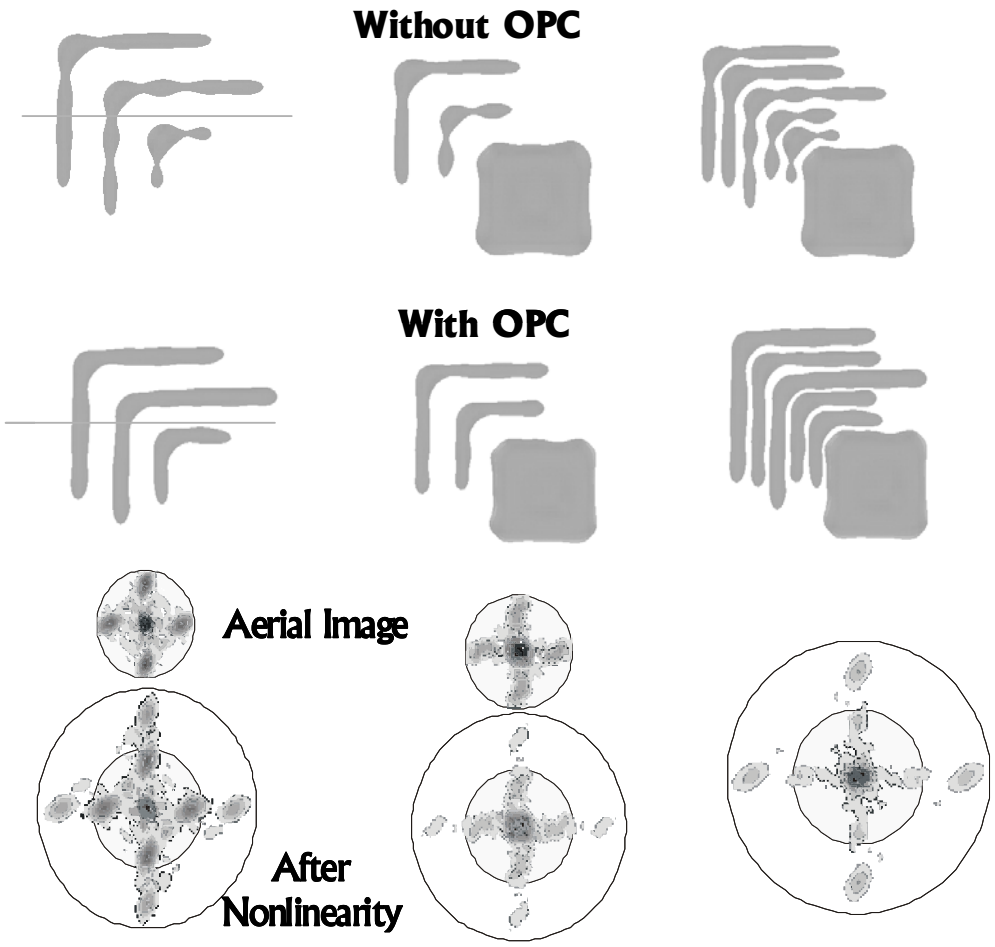


Fig. 5 Simulation for interpolating two images, each near the single exposure limit of optics to provide a final pattern at twice the usual optics limits. The modeling is for 22-nm dense CD at a wavelength of 193 nm with an immersion *NA* of 1.3.

or polymer layer would be required. This layer must be robust to removal of the photoresist and recoating and must have sufficient resolution to store the multiple, interpolated images.

It is critical that the nonlinearities to be applied separately to each exposure. If multiple exposures are accumulated in a single resist layer before the nonlinearity is applied [equivalent to $\mathcal{N}(I_1+I_2)$ rather than $\mathcal{N}(I_1)+\mathcal{N}(I_2)$in Eq. 4], the patterns merge and enhanced resolution is not realized. The top row shows the results of two exposures and their combination using an straightforward pair of WYSIWYG masks without any OPC. The line edge ripple is increased substantially from the 45 nm node case shown in Fig. 1c). This is due to the increased exposure uniformity requirements associated with the overexposure as discussed below. The second row shows improved results with the addition of some optical proximity correction (OPC) mask features (see discussion below). The two bottom left figures show the aerial image frequency content for the first exposure, limited to $2NA/\lambda$ by the transmission limit of optics, and the final frequency content of the image stored in the sacrificial layer after the nonlinear exposure, development and pattern transfer steps. The middle column shows similar results for the second exposure (top: WYSIWYG image; middle: image with OPC; and bottom: Fourier content in aerial image and following nonlinearity and pattern transfer. The rightmost column shows the combined image results (top: WYSIWYG; middle: with OPC; bottom: Fourier content of composite image.) Note that just as was the case for the simple grating example, the fundamental frequency components are out of phase and largely cancel leaving the higher-frequency second-harmonic components as the dominant spectral features in the composite image. Note again, that these second harmonic frequencies are not transmitted by the optics, but arise solely from the nonlinearities in the exposure, develop, and pattern transfer steps.

These results imply that the process control issues will be more significant at these high exposure levels as the linewidth variation for a small change in intensity becomes more severe. This is explained in the Fig.5 which shows the aerial image intensity profiles along the cuts indicated by the dotted lines in the left hand images of the Fig.4, passing through two vertical "ell bars".

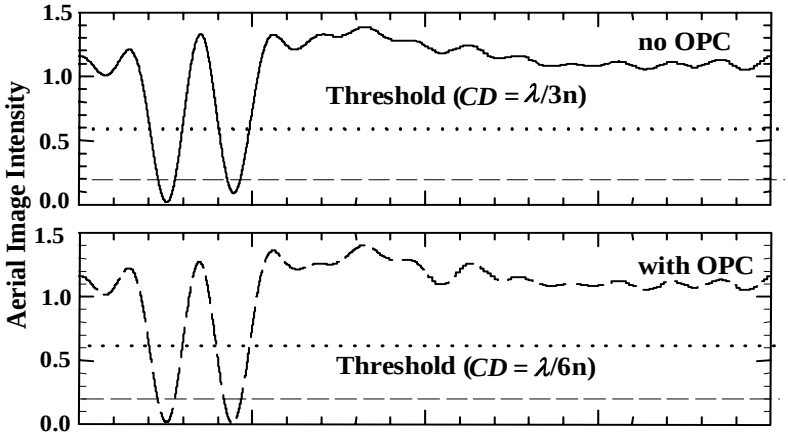


Fig. 6 The aerial image intensity profiles along the solid lines in the lefthand images of Fig. 4 (see text).

Because the amplitudes of the two lines are not precisely equal, there is a relative narrowing of the second line in the final developed pattern. This is more pronounced at the higher exposure level (indicated by the lower dashed horizontal line labeled Threshold (CD = $\lambda/6n$)) than at the lower exposure level ($\lambda/3n$) that is used for single exposure, half-pitch lithography (e.g. Fig. 1). This sensitivity is the reason for the enhanced sidewall ripple in the no OPC case in fig. 5. The imbalance between the exposures for these lines can be addressed by widening of the mask pattern at this point as shown in the OPC mask (Fig. 6). The frequency of the intensity oscillation is set by the cutoff limit of optics and is the same for both lines, so the intensity (depth of the null) is the only possible adjustment, each line is the same width. If the threshold is moved up to the mid-intensity levels, as it shown by the dotted lines labeled "threshold (CD~λ/3n)" the impact of these variations on the final printed image is much less pronounced. For this simulation, the definition of this OPC was an ad-hoc process of adding and subtracting bars in the areas that exhibited the most pronounced deviations.

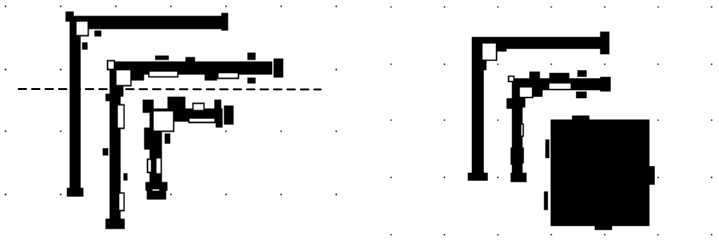


Fig.7: OPC additions (black) and subtractions (white) to the masks used for the simulations.

An extensive investigation of the OPC was not carried out and a better result is certainly available. These results are, however, sufficient to prove the concept. The message is that mask correction, mask CD control, and process control all will become more demanding as increasing orders of nonlinear interpolations are attempted.

5 CONCLUSIONS

There are no fundamental limits to optical lithography. There are only process, implementation and cost issues that will have to be weighed against the similar issues in new technologies. Spatial-frequency doubling after applying process nonlinearities independently to each sub-image has sufficient flexibility to extend immersion 193-nm optical lithography to at least the end of the current industry roadmap. Combining multiple nonlinear exposures requires transferring each image to a sacrificial layer and multiple resist layers. This represents an increased processing burden. In addition, the demands on the mask while not increasing in terms of resolution do increase in terms of linewidth control and the need for OPC. In principle, these nonlinear interpolation techniques can be applied multiple times. In practice, manufacturing considerations will impose a limit that can only be adequately evaluated in comparison with alternative technologies.

Acknowledgements: Support for this work was provided as part of the ARO/MURI in *Deep Subwavelength Optical Nanolithography*.

REFERENCES

[1] ITRS available at http://public.itrs.net.

[2] J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, "Liquid Immersion Deep Ultraviolet Interferometric Lithography," *J. Vac. Sci. Technol*. **B17**, 3306 (1999).

[3] M. Switkes and M. Rothschild, "Immersion Lithography at 157 nm," *J. Vac. Sci. Technol*. **B19**, 2353 (2001).

[4] A. K. Raub and S. R. J. Brueck, "Deep UV Immersion Interferometric Lithography," *Proc. SPIE* **5040**, Optical Microlithography XVI, 667-678 (2003).

[5] See also Special Section of the *Journal of Microlithography, Microfabrication, and Microsystems*, January 2004

[6] John H. Burnett and Simon Kaplan, "Measurement of the refractive index and thermo-optic coefficient of water near 193 nm," *Proc. SPIE* **5040**, Optical Microlithography XVI, 1742-1749 (2003).

[7] Abani Biswas and S. R. J. Brueck, "Simulation of the 45-nm Half-Pitch Node with 193-nm Immersion Lithography-imaging interferometric lithography and dipole illumination," *J. Microlith. Microfab. and Microsys*. **3**, 35-43 (2004).

[8] Alfred K. Wang, *Resolution Enhancement Techniques in Optical Lithography*, SPIE Press TT47, 2001.

[9] S.H. Zaidi and S .R. J. Brueck, "Nonlinear Processes to Extend Interferometric Lithography," *Proc. SPIE* 3676, 371-378 (1999).

[10] T. Ebihara, M. D. Levenson, W. Liu, J. He, W. Yeh, S. Ahn, T. Oga, M. Shen and H. M'saad, "Beyond k_1=0.25 lithography: 70nm L/S patterning using KrF scanners," *Proc. SPIE* **5626**, 985-994 (2003).

[11] X. Chen and S. R. J. Brueck, "Imaging Interferometric Lithography - Approaching the Resolution Limits of Optics," *Opt. Lett.* **24**, 124-126 (1999).

Imaging capabilities of resist in deep ultraviolet liquid immersion interferometric lithography

Alex K. Raub,[a] A. Frauenglass, and S. R. J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

Will Conley
Freescale Semiconductor Assignee/SEMATECH, Austin, Texas 78735

Ralph Dammel and Andy Romano
AZ Electronic Materials, Somerville, New Jersey 08876

Mitsuru Sato
TOK, Koza-Gun, Kanagawa-Ken 253-0114, Japan

William Hinsberg
IBM Almaden Research Center, San Jose, California 95120

(Received 20 July 2004; accepted 6 October 2004; published 14 December 2004)

Liquid immersion lithography (LIL) extends the resolution of optical lithography to meet industry demands into the next decade. Through the use of exposure media such as purified water (n of 1.44 at 193 nm), it is possible to reduce minimum pitches compared with traditional air/vacuum exposures media by a factor of as much as 44%—a full technology node. Beyond this simple observation, there is a good deal of work necessary to fully understand the impact of LIL immersion lithography on a lithography processes. This article addresses the impact of water immersion on the imaging capabilities of different resist formulations. All resists were evaluated by imaging dense line-space structures at a 65-nm half-pitch both in air and with water immersion. Studies of dense 65-nm lines made by immersion imaging in HPLC grade water with controlled variations in resist components were performed. Significant differences were observed and will be discussed. © *2004 American Vacuum Society.* [DOI: 10.1116/1.1824951]

I. INTRODUCTION

As the microelectronic industry pushes towards smaller feature sizes, ever greater demands are made on lithography. Current lithographic approaches are being pushed to practical limits, while next generation lithography technologies are still under development. Recently, interest has grown dramatically in exploring the use of a high-index liquid-immersion medium between the exit face of the lens and the photoresist surface to further extend optical lithography.[1–7] The use of immersion media to extend optical resolution is well known from oil-immersion optical microscopy. Through the use of immersion media, such as deionized water with $n \approx 1.44$ at 193 nm and 1.42 at 213 nm,[7,8] smaller minimum feature sizes can be achieved than are possible in air. This can be thought of either as an effective wavelength, $\lambda_{\text{eff}} = \lambda_0/n \sim 134$ nm, or alternatively as an increase to an effective $\text{NA}_{\text{eff}} = n \sin\theta \leqslant 1.44$. Thus, the use of water immersion lithography could potentially extend 193-nm wavelength optical lithography by 44%, twice the resolution enhancement offered by the problematic transition to 157 nm. Liquid immersion lithography provides an improvement in depth of field for a given NA. At a fixed resolution, $\text{DOF} = k_2\lambda_0/n \sin^2(\theta)$.[9] An air exposure system with a NA of 0.80 will have a smaller DOF than an immersion exposure system with the same NA. Application of immersion techniques at 157 nm would have a similar impact should a suitable liquid be identified.

There are some issues that need to be addressed before immersion lithography can be used in manufacturing. One concern is the interaction between the immersion fluid and the last lens element. For instance deionized water (DI H_2O) will etch calcium fluoride (CaF_2) lens elements, so a protective coating will be needed. The focus of this paper is the interaction between the immersion fluid and the resist-coated wafer. With liquid immersion lithography there are two new interactions of concern: first uptake of the immersion liquid into the resist and its potential impact on resolution; and second desorption of resist components into the immersion liquid and possible contamination of the lens surface. Both of these processes can change the photoresist chemistry and the optical properties of the immersion liquid.

This paper presents exposure results for immersion lithography at a wavelength of 213 nm, close to the current 193-nm exposure wavelength. A maskless, interferometric-lithography approach that gives fringes with essentially ideal contrast was used.[10] Exposures were carried out on a number of resists with controlled variations in resist components. Developed resist patterns of dense 65-nm lines obtained from the resist exposed in air and in HPLC grade water were studied. Significant differ-

[a]Electronic mail: alex.raub@chtm.unm.edu

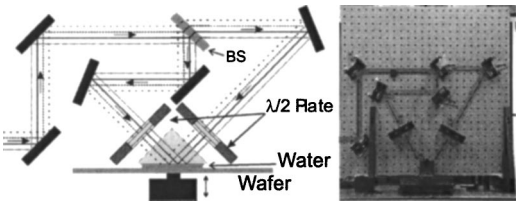


FIG. 1. Michelson interferometer immersion exposure setup.

ences related to resist formulations were observed and are discussed. The use of a protective coating for the resist was also studied and is discussed.

II. EXPERIMENTAL SETUP

The light source was a Coherent Infinity 40-100 Nd:YAG pulsed (~3 ns) laser followed by nonlinear crystals to generate a fifth-harmonic beam at a wavelength of 213 nm. This source has a much higher coherence length (~0.6 m) than that of a typical 193-nm excimer laser, which greatly simplifies the experimental arrangement for interferometric lithography. 193-nm resists generally have comparable absorption at 193- and 213-nm and work well at this wavelength. A Michelson interferometer with a coupling prism was designed and constructed for performing the immersion exposures. Figure 1 shows a diagram and a picture of the immersion exposure setup. The light was coupled into the immersion liquid using an isosceles prism. The top faces of the prism were antireflection coated. A drop of water is placed between the bottom of the prism and the resist coated wafer (thickness ~1 mm). Both laboratory deionized water and HPLC water were used with comparable SEM cross-section resist line profile results.

The period of the grating exposed in the resist is given by $\Lambda=\lambda/2\cdot n_1-\sin(\theta_1)=\lambda/2\cdot n_2-\sin(\theta_2)$. Here, n_1 and n_2 denote the indices of the prism and immersion liquid, and θ_1 and θ_2 are the corresponding propagation angles with respect to the wafer normal; as drawn n_1 is larger than n_2. The grating period is independent of the refractive index of the plane-parallel liquid layer: the change in wavelength according to $\lambda=\lambda_0/n$ is exactly canceled by the change in θ due to refraction at both the glass–liquid and the liquid–resist interfaces. In practice the resist has a higher index than the lens, and it is desirable for the glass and liquid indices to be approximately equal in order to minimize interface reflections.

III. EXPERIMENTAL RESULTS

Two resist manufacturers, TOK and AZ Clariant, provided resists for which the chemical components and makeup were fully disclosed. TOK provided six different resists. These resists were formulated using combinations of two polymers and three bases. AZ Clariant provided eight samples of resist formulated using combinations of two polymers, two PAGs, and two bases.

A. TOK resist liquid immersion imaging results

The TOK resists were based on acrylic polymer backbones. Three of the resists ILP01, ILP02, and ILP03 were formulated with an acrylic ter-polymer, and the other three: ILP04, ILP05, and ILP06 with an acrylic tetra-polymer. All six resists used the same photoacid generator (PAG), triphenylsufonium nanofluorobutanesulfonate (TPS-Nf). Three different base quenchers were used. Resists ILP01 and ILP04 used the small molecular-size base, tri-ethanolamine. Resists ILP02 and ILP05 used the medium sized base tri-pentylamine. Resists ILP03 and ILP06 used the large sized base tri-octylamine. Table I summarizes the matrix of TOK-ILP resist formulations. All resist were spun at a thickness of 100 nm unless noted otherwise, atop a 32-nm thick Brewer Science ARC28-4 antireflective coating optimized to give a minimum substrate reflection at 213 nm for 65-nm half-pitch dense patterns.

1. Image comparison of resist formulations

The three acrylic ter-polymer resists (ILP01, ILP02, and ILP03) gave reasonable patterns in both air and water.[11] The small tri-ethanolamine base, used with the acrylic tetra-polymer resists (ILP04), gave good pattern results only in air and exhibited microbridging and pattern collapse when exposed in HPLC H_2O. This suggests that the base has a more significant interaction with the immersion water when incorporated into the acrylic tetra-polymer resist than in the acrylic ter-polymer resist, which did give reasonable line profiles. The other two resists based on the acrylic tetra-polymer however did result in good patterns in the immersion exposures.[11] When the samples were exposed in HPLC H_2O they exhibited lessened top rounding than for the corresponding air exposures, and the smaller the base the less rounded the profile. The samples using the acrylic ter-polymer gave slightly better results than the samples made from the acrylic tetra-polymer when exposed in HPLC H_2O.

TABLE I. TOK resist formation matrix.

		Base quencher		
		Tri-ethanolamine (Bs)	Tri-pentylamine (Bm)	Tri-octylamine (BL)
Polymer backbone	Acrylic ter-polymer	TOK-ILP01	TOK-ILP02	TOK-ILP03
	Acrylic tetra-polymer	TOK-ILP04	TOK-ILP05	TOK-ILP06

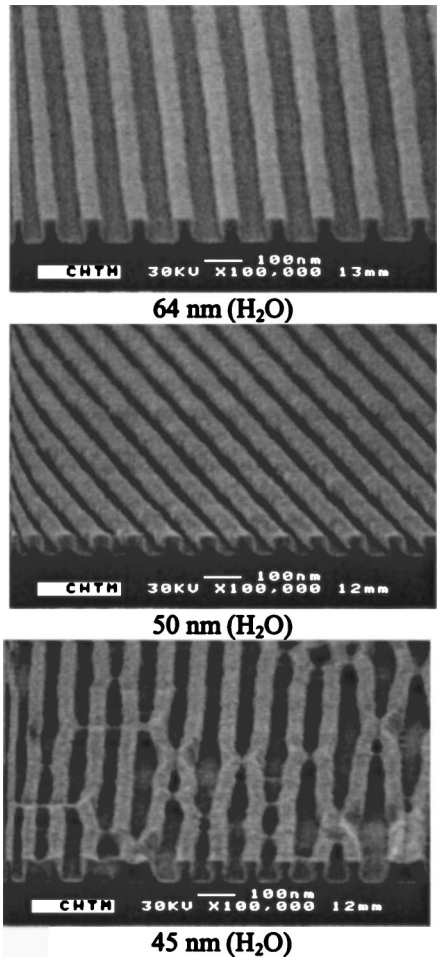


FIG. 2. TOK-ILP03 resist exposures through half-pitch.

The best TOK resist for immersion lithography was TOK-ILP03 closely followed by TOK-ILP06 both containing the large tri-octylamine base.

2. Evaluation through half-pitch

Using one of the better TOK resists (TOK-ILP03), the resist profiles were evaluated as the half-pitch was decreased until the resist was unable to resolve the dense line-space structure. Figure 2 shows the results as the half-pitch was decreased from 64 nm to 45 nm in HPLC H_2O. The dense line-space profiles of the images are square and well defined down to a ~50-nm half-pitch. As the half-pitch was reduced to 45 nm, the dense features exhibited severe micro-bridging and pattern collapse. The usable resolution extends to ~50-nm lines and spaces for the TOK-ILP03 resist under the present exposure and process conditions.

B. AZ Clariant resist liquid immersion imaging results

AZ Clariant provided eight resist formulations. Four of the resists used a COMA hybrid polymer T2030 backbone and the remaining four resists used an acrylic polymer T518 backbone. Half of the resists used a large molecule TPS-NF PAG similar to the TOK resist, and the other half used a small molecule TPS-Tf PAG. AZ Clariant also used two base quenchers, a large molecule tri-octylamine base and a small di-ethanolamine base. Table II summarizes the matrix of AZ Clariant resist formulations. All resists were spun on a 32-nm thick Brewer Science ARC28-4 antireflective coating optimized to give a minimum substrate reflection for 65-nm half-pitch dense patterns.

1. Image comparison of resist formulations

There were issues of pattern collapse in the hybrid polymer T2030 with a resist thickness of 130 nm, and without the use of a rinse surfactant. Beyond the pattern collapse issue, there was noticeable T-toping of the line profiles exposed in air. For the large-PAG, small-base resist formulation patterning issues were clearly evident; the smaller PAG gave better profiles for the air exposures. In the HPLC H_2O exposure, the small-PAG resist formed a skin and once the exposure dose was significant enough for the developer to clear this surface layer, the resist lines were completely overexposed and developed away. This suggests that the small PAG is substantially leaching away from the surface of the resist into the water. Again ignoring the pattern collapse issues, the resist with the large PAG and large base gave the best results in HPLC H_2O for the hybrid T2030 polymer resists.

As with the T2030 polymer resist there were issues of pattern collapse in the Acrylic polymer T518 resist, as the 120-nm thickness exceeded the desired 100 nm. The AZ Clariant acrylic resists all imaged well in air similar to the TOK acrylic resist, again ignoring pattern collapse issues. The resist with the smaller PAG had reduced T-toping and gave better results in air. In the HPLC H_2O exposures, ignoring pattern collapse, the large PAG had slightly more T-toping than the small PAG, and no substantial difference between the small and large base resist images.

TABLE II. AZ clariant resist formulation matrix.

		TPS-Tf (Ps)		TPS-Nf (PL)	
PAG Base quencher		Di-ethanolamine (Bs)	Tri-octylamine (BL)	Di-ethanolamine (Bs)	Tri-octylamine (BL)
Polymer backbone	Hybrid polymer T2030	T2030 Ps Bs	T2030 Ps BL	T2030 PL Bs	T2030 PL BL
	Acrylic polymer T518	T518 Ps Bs	T518 Ps BL	T518 PL Bs	T518 PL BL

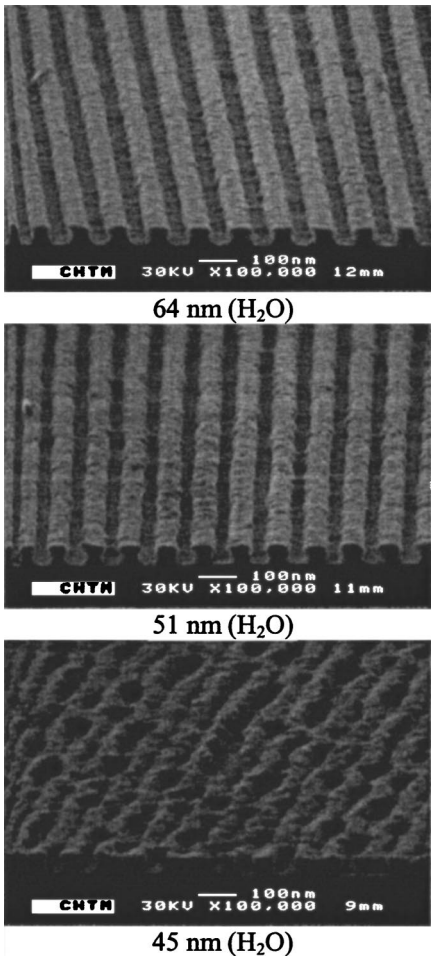


FIG. 3. AZ Clariant EXP M3000-50 resist exposures through half-pitch.

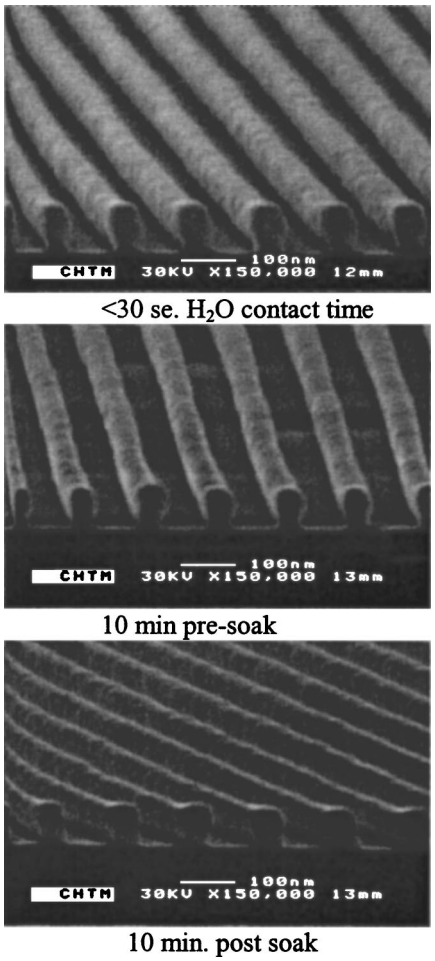


FIG. 4. Impact of soak times on 131-nm pitch TOK-ILP06 resist.

Based on the limited sample space, it seems that the acrylic resists are more suitable for liquid immersion lithography. The size of the PAG resulted in noticeable differences for both the hybrid and acrylic polymer resists. The best AZ Clariant resist was the acrylic polymer T518 with the smaller PAG and large base.

2. Evaluation through half-pitch

Using an improved acrylic AZ Clariant resist (EXP M3000-50A) spun 100-nm thick, resist profiles were evaluated as the half-pitch was decreased until the resist was unable to resolve the dense line and space structure. Figure 3 shows the resist profiles as the half-pitch is decreased from 64 nm to 45 nm in HPLC H_2O. Similar to the TOK acrylic resist, the AZ Clariant resist dense line-space profiles were also square and well defined down to a 50-nm half-pitch in HPLC H_2O. However the line edge roughness was more pronounced for the AZ Clariant resist. The AZ Clariant resist also exhibited severe micro-bridging and pattern collapse as the half-pitch was reduced to 45 nm.

C. Impact of resist soak times

Using one of the best performing TOK resists (TOK ILP06), the impact of sample soak-times in HPLC water was investigated. First a sample was exposed as soon as it was immersed in the water and dried off right after the exposure followed by a postexposure bake, for a maximum water contact time of less than 30 s. A second sample was soaked in the water for 10 min before it was exposed and immediately dried off and postbaked. A third sample was exposed as soon as it was immersed in the water and then soaked in the water for 10 min before it was dried off and postbaked. Figure 4 shows the exposure results for these three samples.

The 10-min presoak sample had slightly better exposure latitude and uniformity, which was evaluated across the full wafer and cannot be seen in the figure. Regardless of soak time in the water all samples gave similar, good line profiles. It is believed that from this and other cited work that any interactions between the resist and water take place in a very short time (less than 1 s).

D. Impact of top surface protection

One suggested technology to isolate the resist from the immersion liquid is the use of a top-surface-protection (TSP)

barrier layer. The barrier layer needs to be transparent to the exposure light and to be inert to both the immersion liquid and the resist. TOK has formulated a fluorocarbon spin-on top-protective coating TSP-03A, which meets these requirements. Two additional steps are added to the processing. First after resist spin-coating and postapplication bake, an additional spin-on and bake sequence is used to apply the TSP-03A material atop the resist. Second after the postexposure bake, the TSP-03A coating is removed using a special solvent, TSP-3 remover, before the resist development.

Three TOK resists (TOK-ILP01, TOK-ILP03, and TOK-ILP06) were exposed in air with and without the TSP-03A coating. For the samples with the protective coating, the developed resist line shapes are very similar to those for samples without the coating when exposed in air. Similarly the same three resists were exposed in HPLC H_2O with and without the TSP-03A coating. The samples with the protective coating look practically the same as the samples exposed in air. However the samples exposed in HPLC H_2O without any top coating have noticeably squarer line profiles with reduced rounding. This would suggest that the top surface protection barrier is preventing an interaction between the resist and water. Other experiments suggest that the photoacid generated upon exposure is leached out of the top few monolayers of the resist into the water.[12,13] This would also support the observed squaring of the line profiles for the immersion exposure, as the top of the resist becomes less sensitive as compared to the bulk of the resist.

E. Impact of a rinse surfactant

One of the issues with imaging 65-nm half-pitch lines is that at aspect ratios larger than 2.5 there is frequently complete pattern collapse. This is due to the impact of the water surface tension as it dries unevenly between the lines after the development rinse. The simplest way to reduce this pattern collapse is to reduce the surface tension of the water by adding a surfactant. At a 65-nm half-pitch pattern collapse due to adhesion tension will not occur if the resist thickness is at or below 100 nm, however at 150 nm and larger there is often pattern collapse. At a 45-nm half-pitch with 100-nm thick resist, the pattern collapse was severe using a DI H_2O rinse. This pattern collapse was not noticeable with a surfactant/water rinse, but microbridging was still observed due to the resolution limits of the resist. A much more controlled investigation of pattern collapse limits at these scales is needed to understand the limitations and control strategies associated with pattern collapse issues.

F. Evaluation of contact holes

Using the same acrylic AZ Clariant resist (EXP M3000-50A) spun at a 100-nm thickness, dense contact holes were exposed in the resist and the profiles were evaluated. Contact holes are achieved by performing two sequential exposures where the sample is rotated by 90° between the two exposures. This technique allows for using TE polarized light in both the *x*- and *y*-directions of the contact hole pattern. Figure 5 shows the 65-nm diameter dense contact holes array in the AZ Clariant resist. Work continues to optimize this process and to understand and reduce the hole-to-hole variations evident in these first exposures.

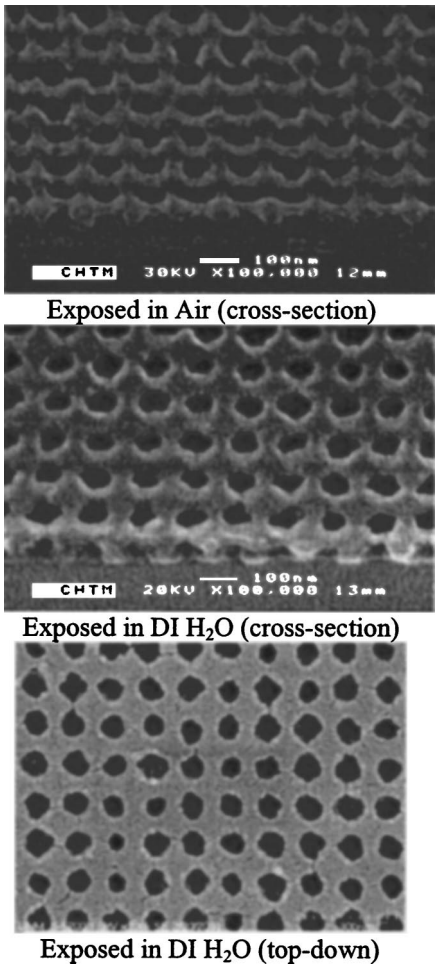


FIG. 5. 129-nm pitch contact holes in AZ Clariant EXP M3000 resist.

IV. SUMMARY

The work in this paper has demonstrated that 45-nm half-pitch lines can be patterned into resist using liquid immersion lithography (LIL) with a NA = 1.18 and $\lambda = 213$ nm. This would allow the extension of 193-nm lithography by almost two technology nodes.

Some but not all existing 193-nm resists work well in LIL immersion lithography. Based on these preliminary results, it seems likely that 193-nm LIL immersion resists can be formulated with fully comparable performance to today's 193-nm air-exposed resists. The acrylic polymer resist platform performed well in LIL immersion lithography. It is also clear that PAGs, bases, and polymers combinations play an important role. Studies are needed to find an optimal set of PAGs and bases for a given polymer system to be used as a 193-nm LIL immersion resist. These studies will need to be performed for contact holes, dense lines, and isolated lines, as the resist performance tends to differ between them.

Resist top surface protection coatings work well with 193-nm resist and water immersion liquid. Although a top

surface protection layer does prevent resist water interaction, there can be interactions between resist and the protection layer. Still there are resists that perform well when exposed in contact with water and provide a simpler process flow.

As LIL liquid immersion lithography enables the printing of smaller lines and spaces, pattern collapse due to adhesion tension during post develop wafer drying will become increasingly important. Rinse surfactants should reduce pattern collapse issues, however resist thicknesses likely will continue to decrease, posing issues for subsequent pattern transfer steps. The problem of pattern collapse needs further study to quantify the limit of resist aspect ratios for specific resist systems. A resist with a larger Young's modulus will help combat these issues. Evaluation of rinse surfactants and possible developer surfactant needs to be performed at the LIL immersion lithography technology nodes.

All the exposure made in this article used the optimal TE polarization. This might not be the case for an imaging immersion lithography system, and polarization effects are going to become very important for hyper-NA imaging. These effects need to be quantified along with their corresponding impacts on resist profiles. The exposures also used HPLC grade water. Although purified water with a refractive index of 1.44 has been identified as the preferred liquid immersion medium, the resolution limits of 193-nm lithography can be further improved if a viable liquid is found with an even higher refractive index.

No major issues were found concerning the resist-liquid interaction for liquid immersion lithography. There are known solutions to deal with the interactions between the resist and the water immersion liquid. Resists can be engineered and developed for 45-nm half-pitch dense lines in LIL immersion lithography and beyond.

[1]C. V. Shank and R. V. Schmidt, Appl. Phys. Lett. **23**, 154 (1973).
[2]See, for example, the special section on Immersion Lithography, in J. Microlithogr., Microfabr., Microsyst. **3**, 8 (2004).
[3]Burn J. Lin, J. Microlithogr., Microfabr., Microsyst. **3**, 377 (2004).
[4]W. T. Tsang and S. Wang, Appl. Phys. Lett. **24**, 196 (1974).
[5]J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, J. Vac. Sci. Technol. B **17**, 3306 (1999).
[6]M. Switkes and M. Rothschild, J. Vac. Sci. Technol. B **19**, 2353 (2001).
[7]A. K. Raub and S. R. J. Brueck, Proc. SPIE **5040**, 667 (2003).
[8]John H. Burnett and Simon Kaplan, Proc. SPIE **5040**, 1742 (2003).
[9]Burn J. Lin, J. Microlithogr., Microfabr., Microsyst. **1**, 7 (2002).
[10]S. Zaidi and S. R. J. Brueck, Appl. Opt. **27**, 2999 (1988).
[11]A. K. Raub, A. Frauenglass, S. R. J. Brueck, Will Conley, Ralph Dammel, Andy Romano, Mitsuru Sato, and William Hinsberg, Proc. SPIE **5377**, 306 (2004).
[12]J. Christopher Taylor, Charles R. Chambers, Ryan Deschner, Robert J. LeSuer, Will Conley, Sean D. Burns, and C. Grant Willson, Proc. SPIE **5376**, 34 (2004).
[13]William Hinsberg, Gregory Wallraff, Carl Larson, Blake Davis, Vaughn Deline, Simone Raoux, Dolores Miller, Frances Houle, John Hoffnagle, Martha Sanchez, Charles Rettner, Linda Sundberg, David Medeiros, Ralph Dammel, and Will Conley, Proc. SPIE **5376**, 21 (2004).

244-nm imaging interferometric lithography

A. Frauenglass, S. Smolev,[a] A. Biswas, and S. R. J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

(Received 22 July 2004; accepted 27 September 2004; published 14 December 2004)

Imaging interferometric lithography, combining off-axis illumination, multiple exposures covering different regions of spatial frequency space, and pupil plane filters to ensure uniform frequency-space coverage, is a relatively new imaging concept that provides an approach to accessing the fundamental, linear-systems-resolution limits of optics. With an air medium between the lens and the wafer, the highest spatial frequency available with 244-nm exposure tool with a numerical aperture of 0.9 corresponds to a half-pitch of 68-nm. Allowing for ~10% subbands above this central frequency, this suggests that ~75-nm half-pitch patterns should be accessible. A 22× reduction imaging interferometric lithography testbed demonstration of printing a non-periodic (arbitrary) 86-nm half-pitch pattern is reported. This result was achieved with a simple chrome-on-glass mask without the use of any mask-based resolution-enhancement techniques such as phase-shift or optical proximity correction. Scaling this result to a 193 nm wavelength and an immersion numerical aperture of 1.3 directly addresses the 45-nm half-pitch node. © *2004 American Vacuum Society.* [DOI: 10.1116/1.1821505]

I. INTRODUCTION

The tremendous growth of the microelectronics industry has enabled today's digital revolution. A driving force has been the steady improvement in semiconductor manufacturing technology, especially in lithographic pattern definition, enabling the continuous increase in the functionality and speed of integrated circuits. Moore's Law[1,2] has both predicted and driven the remarkable advances in increased chip area, decreased feature size and improved circuit design. The National Technology Roadmap for Semiconductors, developed by the Semiconductor Industry Association (SIA), serves as the industry standard version of Moore's Law.[3] The current industry roadmap, looking forward to the 65-nm technology node by 2007 and 45-nm technology node by 2010, demands a decision on lithographic technology for these nodes in the near future.

A number of new or next-generation lithography technologies are being developed to meet the SIA roadmap demands. Some of the technologies being considered as alternatives to traditional optical lithography include extreme ultraviolet (EUVL), electron projection, and nanoimprint lithographies. While these NGL technologies hold promise, it is unclear which, if any, will become the prevailing technology for mass production in the future, especially given the near-term requirements of the roadmap.[4]

Additional attention is being placed on extensions of optical lithography. One particularly promising technique is liquid immersion lithography[5] that allows an increase in the resolution of up to a factor of the liquid refractive index ($n_{H_2O}=1.44$ at 193 nm).[6] This represents a 44% resolution improvement, larger than the 28% wavelength reduction from 248- to 193-nm, or the problematic 23% reduction from 193- to 157-nm. It is therefore important to investigate the ultimate resolution capabilities of optical lithography and to investigate imaging techniques that offer practical approaches to reaching those ultimate limits.

Interferometric lithography provides a simple technique that allows access to features approaching the resolution limits of optics.[7,8] Integration of interferometric lithography with conventional optical lithography[9] offers an optics-based lithographic capability for at least the 45-nm half-pitch generation.[10]

The minimum half-pitch feature interferometric lithography can access is given by the well-known equation, $\Lambda_{\text{half-min}}=\lambda/4n\sin\theta$, where λ is the exposure wavelength, n is the refractive index of the immersion medium and the two beams propagate at angles of $\pm\theta$ with respect to the wafer surface normal. For 244-nm exposures in air ($n=1.0$) and a maximum NA$=n\sin\theta$ of 0.9, this minimum half-pitch is 68-nm. For an imaging system, the ultimate resolution does not reach this level as a result of the need to capture the information in the pattern-dependent sidebands around the fundamental frequency. Assuming that a bandwidth of ~10% around the center frequency is needed for a typical pattern, a system with these capabilities will be able to print ~75-nm half-pitch patterns. There remains the issue of realizing this resolution in the face of the nonlinear properties of optical imaging; optical proximity correction and phase-shift masks are well-studied directions for enhancing the resolution of an lithographic image.

Imaging Interferometric Lithography (IIL), using off-axis illumination at the limits of the pupil along with pupil-plane filters, to ensure a uniform frequency-space coverage, is a relatively new imaging concept that provides an approach to accessing the fundamental linear system resolution limits of optics. The resolution limits of conventional optical lithography reflect the low-pass spatial-frequency characteristies of the imaging system. IIL eases many of the constraints that limit the capabilities of conventional optical lithography and

[a]Electronic mail: smolev@chtm.unm.edu

offers a path to the ultimate linear systems which limits of optics extends to dense CDs of $\sim\lambda/3\mathrm{NA}$ ($\mathrm{NA}=n\sin\theta$ for immersion) for arbitrary patterns. Significant benefits of IIL, especially for high-NA optics, include the ability to optimize the polarization for the highest spatial frequencies, and to vary relative intensities and contrast of the multiple images separately in the *x*- and *y*-directions for 2D mask patterns.

II. IIL

IIL was introduced[9,11] as a technique to cover the extremes of frequency space by tilting the illumination as far as possible off-axis. With the NAs of ~ 0.6 available at that time, this required tilts that were beyond the limits of the imaging pupil and consequently reintroduction of a zero-order beam on the image side of the lens. The first experiments showed a successful result for the 2-μm CD using $\lambda = 364$ nm and $\mathrm{NA}=0.04$ equivalent to a Rayleigh κ_1 of 0.22.[12] The next step towards higher NAs and smaller features was made using $\lambda = 364$ nm and $\mathrm{NA}=0.356$ with the principals of IIL applied in only one direction because of the difficult optical system resulting from the small working distance of the lens. A CD of 250 nm was achieved, equivalent to a $\kappa_1 \sim 0.25$.[13]

With today's very high NA lenses of $\sin\theta \sim 0.9$, it is sufficient to set the offset to the largest value allowed by the lens NA and to retain the zero-order beam within the lens, eliminating the need for an around-the-lens interferometer. Thus, IIL is similar to quadrupole off-axis illumination with a small partial coherence for each off-axis illumination beam, and with quadrupole oriented along the principal (x, y) axes of a Manhattan geometry pattern to maximize the extent of the spatial frequency capture (as in dipole illumination, but with only a single mask). The *x*- and *y*-axis spatial frequencies are covered with separate off-axis illumination exposures. As NAs are increased even further with the use of liquid immersion techniques, control over the light polarization will become a critical issue. TE polarization retains full contrast while the contrast of TM is reduced and phase shifted at higher angles.[10] The choice of proper polarization for the high spatial frequency components in orthogonal directions dramatically improves the resist patterns.

Quadrupole off-axis illumination inherently emphasizes low frequencies at the expense of high frequencies because each of the exposures redundantly covers the same low-frequency information, while the high frequencies are covered only in individual exposures. Pupil plane filters can be used to eliminate these multiple coverages and thereby provide a uniform transfer function and an improved image.[14]

In this paper, we present our initial results at demonstrating IIL at high NAs and at feature sizes approaching the ultimate limits outlined above. The experimental testbed uses a 244-nm, cw Ar-ion laser source and a 0.9 NA lens. The optical system includes a relayed imaging system pupil plane on the mask side of the imaging lens to allow insertion of pupil plane filters. As detailed below, our first results demonstrate resolution for arbitrary patterns to at least an 86-nm half pitch ($\kappa_I \sim 0.32$) with a simple binary mask without the need for either optical proximity correction or phase-shift mask-based resolution enhancement techniques. Scaling this result to 193-nm and to an immersion NA of 1.3 directly gives the 45-nm half-pitch node.

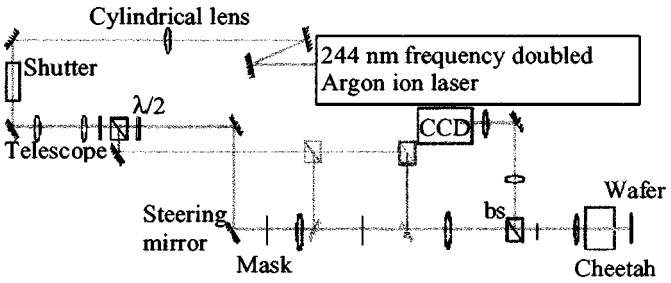


FIG. 1. Schematic layout of the experimental IIL tool with 244-nm laser source, 0.9 NA *Cheetah* objective lens and real-time focusing system.

III. EXPERIMENTAL SETUP

A 0.5-W, cw, intracavity doubled, Ar-ion laser at 244 nm was the illumination source for the IIL exposures. A cylindrical lens was used to correct for the frequency-doubling-induced astigmatism and to get a spatial distribution close to TEM_{00}. A polarizer and a half-wave plate were used to control the beam intensity. A second half-wave plate is used to control the polarization of the beam for the various IIL exposures. A pair of steering mirrors allows control of the off-axis angle-of-illumination of the mask for the various exposures. A schematic of the experimental layout is shown in Fig. 1.

The key element of the optical system is the 0.9 NA "*cheetah*" lens (TROPEL). Lens specifications are presented in Table I. The imaging pupil plane inside the lens is inaccessible. Therefore, a set of transform lenses is used to relay the pupil plane to an accessible position for the insertion of pupil plane filters. The lens is catadioptric with central obscuration of 15% of the NA. Specifically, this means that coherent illumination at normal incidence is not possible be-

TABLE I. Specifications for the 0.9 NA TROPEL *Cheetah* lens.

Trade name	Cheetah 244
NA	0.9
Field μm	1.00
Wavelength (nm)	244
Pupil diameter (mm)	3.6
Tube length	Infinity - custom tube lens
Working distance (mm)	1.1
Telecentric (deg.)	0.1 - object side
Transmission (%)	75 - normal
Optical material	Excimer SiO_2
Obscuration (%)	15 - linear
Cemented surfaces	None
Internal foci	None
Diameter (mm)	72 - maximum
Length (mm)	72 - maximum
Mass (g)	1200

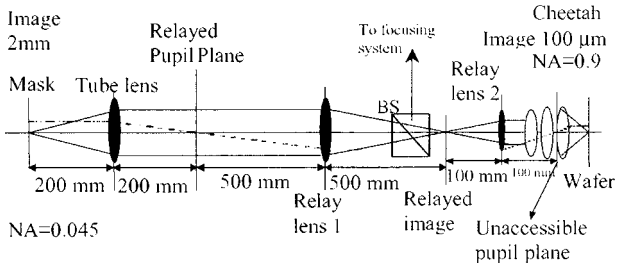


FIG. 2. Schematic layout of the 22× reduction telecentric imaging geometry showing imaging and illumination rays and relative positions of lenses and relayed planes.

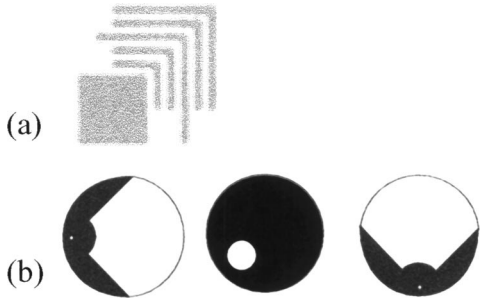


FIG. 3. PROLITH™ simulations of IIL were used prior to the experiment to provide insight into resolution and exposure strategies. (a) 3-exposure IIL for a CD of ~86 nm, ends of the lines are well defined, isolated lines are cleared; and (b) the pupil plane filter used in the off-axis exposures (left and right filters) and low-frequency exposure (center filter).

cause the lens blocks the zero-order beam. Thus, an off-axis strategy must be adopted for the low-frequency portions of the image as well as for the high frequencies in the IIL spectral decomposition.

An optical schematic of the 22× reduction imaging system is shown in Fig. 2. The optical system was simulated with standard ray-tracing software. The full $100\times100\ \mu m^2$ field-of-view at the wafer plane has a wavefront flatness of 0.82 waves. There are several addition elements in the setup that are associated with visual feedback of the image focus. The depth of focus for the high-frequency IIL exposures is only ~200-nm, requiring precise *z*-positioning of the wafer; since only a single level is printed, there is no *x*–*y* positioning requirement. Wafer flatness is not an issue for the small lens field-of-view. For visual feedback of the image focus we built an optical system associated with the beamsplitter cube in the primary optical path that provides a real-time view of the image, i.e., magnifies the reflection of the mask image from the photoresist surface back through the Cheetah and projects the image onto a CCD camera. Since the optical system onto the camera can accommodate an out-of-focus image, it is necessary to simultaneously image the reflected pattern as well as artifacts on the photoresist surface.

The mask patterns are Manhattan structure patterns [dense/isolated nested ells with a large box (10 CD) to present a range of spatial frequencies]. The binary chrome-on-glass mask does not incorporate any optical proximity correction (OPC) or phase shift mask (PSM) features. As a result of the 22× reduction, the mask fabrication was relatively straightforward and posed no special issues.

IV. EXPERIMENTAL PROCESS

A. Pupil plane filters

PROLITH™ simulations were used prior to the experiment to provide insight into resolution and exposure strategies. PROLITH can simulate the IIL image process using custom sources and pupil plane filters and different polarizations. From the simulation, the optimal parameters can be obtained as evaluated from the areal image and the developed photoresist patterns. Figure 3(a) shows a simulation result for a 3-exposure IIL for a CD of 86-nm. Figure 3(b) shows the pupil plane filters that were used for this simulation. The filter for the high *x*-spatial frequencies includes a small hole at the edge of the pupil that transmits the zero-order beam. In the experiment, the filter is put in place and the off-axis tilt is adjusted to ensure that the zero-order beam is cleanly transmitted through the 0.5-mm hole ($\sigma=0.003$). A very small size is required in order to maximize the spatial frequency content of the image. The maximum spatial frequency transmitted through the optical system with this filter in place is $2NA/\lambda$ and corresponds to the 68-nm half pitch discussed earlier. The black regions of the spatial filter block the low-frequency spectral components of the image because these components would otherwise be covered in both the high *x*-frequency (left filter) and high *y*-frequency exposures (right filter) and would therefore be overrepresented in the final image. The "batwings" at 45° from the zero-order aperture represent frequencies that are covered in one of the high frequency exposures and blocked in the other. The small circular regions around each zero-order aperture eliminate the low spatial frequencies that are covered in the low-frequency (center filter) exposure. The zero-order beam is directed through the center of the aperture in this exposure. The hole is offset to allow for the central lens obscuration. The same coverage could be accomplished with only two exposures by extending the transparent "vee" regions all the way to the zero-order aperture. There are two reasons that the present scheme was chosen. First, the alignment of the two apertures becomes critical if a two-exposure scheme is chosen since there are strong spectral components in the low frequency region that must be captured for a faithful image. This is not critical in the present scheme since there are no strong spectral components at the transitions from blocked to open in the present filters. Second, for the "vee" geometry (and for conventional imaging) the strong spectral components at low frequencies and those corresponding to the dense lines and spaces interfere with each other resulting in strong autocorrelation (dark-field) components, which can distort the final image. In the present scheme these components are separated into different exposures and the unwanted interference terms are completely eliminated. After simulating many different pupil plane filters geometries, we found that this set of filters gives the best imaging results for the present pattern.

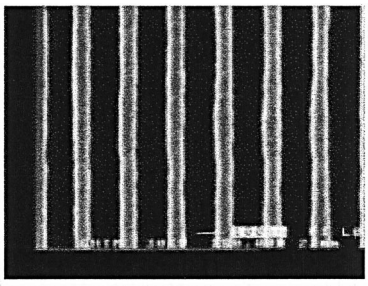


FIG. 4. SEM of PR lines with a CD of ~86-nm, demonstrating that the optical system has the expected spatial frequency bandwidth. Those lines were exposed using a mask and a single off-axis exposure without any pupil plane filter.

B. Photoresist/ARC stack

Because of the relatively small depth-of-focus and the small pattern CD, the ARC and PR thickness have to be very carefully optimized. Optimal thicknesses as obtained from the PROLITH simulation are: ARC (90±8)-nm, PR (100±10)-nm. The bottom ARC was DUV 64-3 from Brewer Science. The photoresist was DUV 210-0.6 from Shipley Corp. The DUV 210-0.6 photoresist was diluted with the ES solvent 11 for a final thickness of ~110-nm after spinning.

C. Exposures

The exposure process starts with insertion of the *x*-direction pupil plane filter and setting the focal position using a reduced intensity beam and the real-time focusing system. ***At this point the mask and wafer are fixed relative to each over, no further adjustment of the positions is allowed***. Then the first exposure is carried out, the pupil plane filter is rotated by 90° (to provide the *y*-pupil filter) and the second exposure is carried out. Then the low frequency pupil plane filter is inserted and the third exposure is made. A postexposure bake at 130 °C for 1-min on a vacuum chucked hot plate, followed by a 45-s CD-26 puddle develop completes the processing

V. EXPERIMENTAL RESULTS

A preliminary set of exposures with periodic grating patterns was carried out to assess the optical system resolution capability. Figure 4 shows the resulting photoresist pattern for a 80-nm half-pitch grating, demonstrating that the optical system has the expected spatial-frequency bandwidth.

A scanning electron micrograph SEM of the developed PR pattern for 3-exposure IIL for the 244-nm wavelength with 0.9 NA lens at CD ~86-nm is shown in Fig. 5(a). This was done using the 1.9-μm CD mask. The hole in the middle of the square is due to overexposure. A PROLITH™ simulation for an overexposed image shows similar results [see Fig. 5(b)]. The disappearance of the corner of the fifth line is due to overexposure as well as possible flare; work is underway to improve the pattern definition. Also deviation from the optimum ARC and PR thicknesses can cause similar problems. The image has well-defined line extensions (isolated lines); the dense lines are clear and well separated which indicates that the imaging system is performing well. The defects in the features are due only to photoresist and exposure issues, not to the IIL imaging capabilities.

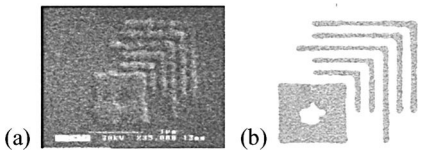


FIG. 5. (a) SEM of developed PR for 3-exposure IIL for a CD of ~86-nm at a 244-nm wavelength and 0.9 NA. Isolated lines and the square edge of the box are well defined, dense lines are cleared; (b) modeling of 3-exposure IIL for a CD of ~86 nm for an overexposed condition; shows the same hole in the middle of the square implying that this feature on the pattern is due to exposure issues not to the imaging.

VI. CONCLUSION

The 244-nm IIL setup has demonstrated a capability of printing patterns with a CD compatible with industrial needs. We have achieved ~86-nm half-pitch features at a 244-nm wavelength with 0.9 NA without any additional mask-based RETs (PSM or OPC). The implication is that IIL can minimize the use of mask-based RETs, thereby reducing mask cost and improving cost of ownership (COO). This conclusion is supported by a recent simulation analysis[10] modeling the printing of similar pattern at the 45-nm half-pitch node using immersion (NA~1.3) and a 193-nm source without the use of any mask-based RETs. Comparing the simulation results and the present PR patterns shows that IIL is capable of printing small patterns. The defects on the photoresist pattern are associated with photoresist and exposure issues, not with the resolution. Introducing the real-time focusing technique was a critical step that dramatically improved the results of the experiment. Control over the polarization was not critical at these dimensions but it will be very important as the CD is scaled further with immersion techniques.

The size of the printed pattern can be scaled using the simple relation $CD \propto \lambda/NA$. In the present experiment we have achieved ~86-nm with a 244-nm source and a 0.9-NA lens. Switching to a 193-nm source and a 1.3 immersion NA leads directly to a ~47-nm half-pitch—again without the need for any mask-based RETs such as phase shift and OPC. Simulation predicts that the resolution of the present system extends to ~75-nm. This would seales to a ~41-nm half-pitch for the 193-nm immersion system.

In the future we plan to optimize the exposure parameters for the 3-exposure IIL and evaluate the resolution limit of the system using IIL using both simulation and experiment. We would like to print different patterns on the current setup such as via holes. Apodization of the pupil plane filters can be used to improve image quality.[15] Scaling to the 45-nm half-pitch node will require use of 193-nm source and an immersion exposure medium in order to achieve 1.15/1.3 NA systems. With this extended testbed, polarization and hyper-NA effects will be investigated.

[1]G. E. Moore, Electronics **38**, 114 (1965).
[2]G. E. Moore, *Proceedings of the IEEE Int. Electron Devices Meeting* (IEEE, New York, 1975), pp. 11–13.
[3]*The National Technology Roadmap for Semiconductors* (Semiconductor Industry Association, San Jose, CA, 1999) (http://public.itrs.net)
[4]K. Derbyshire, Solid State Technol. **41**, 78 (1998).
[5]B. J. Lin, J. Microlithogr., Microfabr., Microsyst. **1**, 7 (2002).
[6]J. H. Burnett and S. Kaplan, Proc. SPIE **5040**, 1742 (2003).
[7]S. H. Zaidi and S. R. J. Brueck, J. Vac. Sci. Technol. B **11**, 658 (1993).
[8]A. K. Raub and S. R. J. Brueck, Proc. SPIE **5040**, 667 (2003).
[9]X. Chen and S. R. J. Brueck, J. Vac. Sci. Technol. B **16**, 3392 (1998).
[10]A. Biswas and S. R. J. Brueck, J. Microlithogr., Microfabr., Microsyst. **3**, 35 (2004).
[11]S. R. J. Brueck and X. Chen, J. Vac. Sci. Technol. B **17**, 908 (1999).
[12]X. Chen and S. R. J. Brueck, Opt. Lett. **24**, 124 (1999).
[13]C. Schwarz, Ph.D. thesis, University of New Mexico, 2003.
[14]X. Chen and S. R. J. Brueck, J. Vac. Sci. Technol. B **17**, 921 (1999).
[15]T. M. Tridhavee, B. Santhanam, and S. R. J. Brueck, Proc. SPIE **5377**, 1544 (2004).

In situ real-time monitoring of profile evolution during plasma etching of mesoporous low-dielectric-constant SiO_2

Henry Gerung, C. Jeffrey Brinker, Steven R. J. Brueck, and Sang M. Han[a)]
University of New Mexico, Albuquerque, New Mexico 87131

(Received 21 June 2004; accepted 3 January 2005; published 1 March 2005)

We have employed attenuated total reflection Fourier transforms infrared spectroscopy (ATR-FTIRS) to monitor the profile evolution of patterned mesoporous, low-dielectric-constant SiO_2 films *in situ* and in real time during plasma etching. A stack of patterned photoresist, anti-reflective coating, and mesoporous SiO_2 is etched in an inductively coupled plasma reactor, using CHF_3 and Ar. During etching, the IR absorbance of Si–O–Si stretching modes near 1080 cm^{-1} decreases, and the rate of decrease in Si–O–Si absorbance translates to the SiO_2 removal rate. When corrected for the exponentially decaying evanescent electric field, the removal rate helps monitor the profile evolution and predict the final etch profile. The predicted profiles are in excellent agreement with the cross-sectional images taken by scanning electron microscopy. In a similar approach, we calculate the absolute total number of C–F bonds in the sidewall passivation and observe its formation rate as a function of time. Assuming that the thickness of the sidewall passivation tapers down towards the trench bottom, we deduce that C–F formation occurs mostly in the final stage of etching when the trench bottom meets the Ge ATR crystal and that a critical amount of C–F buildup is necessary to maintain the anisotropic etch profile. © *2005 American Vacuum Society.*
[DOI: 10.1116/1.1865154]

I. INTRODUCTION

The integrated circuit (IC) manufacturing has witnessed continuous device miniaturization giving rise to numerous engineering challenges. According to the 2003 International Technology Roadmap for Semiconductors (ITRS), the IC device dimension will reach an 18 nm node by 2018.[1] For advanced microprocessors and logic devices, in particular, such miniaturization requires a reduction in the resistance-capacitance (RC) time constant associated with metal interconnects and intermetal dielectrics.[2] The purpose is to increase the device operating speed despite the miniaturization. Two primary solutions exist today to reduce the RC constant. One is copper metallization, and the other is low-dielectric-constant materials insulating the metal interconnects. A variety of materials, such as polytetrafluoro ethylene (PTFE),[3,4] polyimides (PI),[5–7] silsesquioxanes,[8] and mesoporous SiO_2,[9] have been considered as viable candidates to replace the conventional vapor-deposition-based SiO_2. In this study, we focus on solgel-based mesoporous SiO_2 whose dielectric constant ranges from 1.5 to 2.0 in the 1–40 GHz range.[10] We demonstrate that the patterned etch profile of SiO_2 film, along with the sidewall passivation, can be monitored *in situ* and in real time, using attenuated total reflection Fourier transform infrared spectroscopy (ATR-FTIRS). The ATR technique has been previously used to study surface reactions during plasma enhanced processes[11–14] as well as "resetting" the wall condition of commercial plasma reactors for process reproducibility.[15] In all cases, the ATR technique provides sub-monolayer sensitivity to examine heterogeneous reactions occurring on semiconductor and insulator films.[16,17] We take advantage of the high sensitivity to probe the patterned surface of mesoporous SiO_2 films during etching. This nondestructive method eliminates the need for cross-sectional scanning electron microscopy (XSEM) to examine the etch profile. We expect that the ATR-FTIRS technique and our analytical approach can be applied to other dielectric films such as Si_3N_4 and SiO_xN_y during etching for the purpose of process development. The wafer-level use of the ATR technique in production tools, however, might be limiting, since the multilevel, intermetal dielectric architecture contains metal interconnects. These metal lines, whose pitch is smaller than the wavelength of IR, would prevent the evanescent wave from propagating through the dielectric film.

II. EXPERIMENT

Figure 1 illustrates a top cross-sectional view of our experimental setup. The stainless steel, tubular plasma chamber is 25 cm in diameter with multiple view ports. A turbomolecular pump (Alcatel 5900CP) with 1000 l/s pumping speed maintains the base pressure at 6.7×10^{-4} Pa. The turbomolecular pump is assisted by a 14 l/s double-stage rotary vane mechanical pump (Edwards E2M40). A gate valve controls the conductance to the turbomolecular pump and maintains the chamber pressure at 1.3 Pa during etching, independent of reactant gas flow rates. A mixture of CHF_3 and Ar is introduced to the chamber through a 10 cm diam circular gas distribution ring located 12.5 cm above the substrate. During etching, mass flow controllers (UNIT URS 100-5) maintain the flow rates of CHF_3 and Ar constant.

A "stove-top-coil" inductive plasma source is mounted on a 2.54 cm thick, 16 cm diam Pyrex window that separates the plasma source from vacuum. A radio frequency (rf) potential

a)Electronic mail: meister@unm.edu

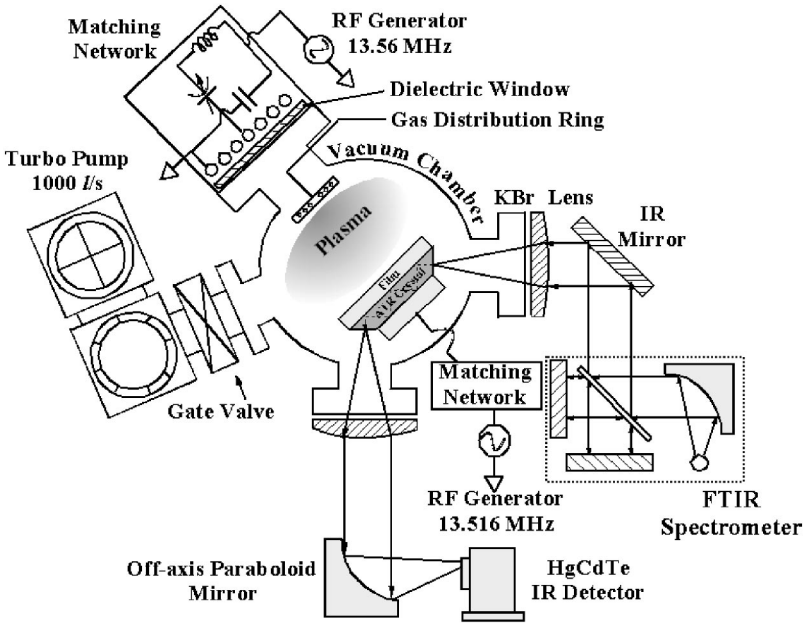


FIG. 1. Schematic diagram of plasma etch chamber equipped with ATR-FTIRS setup.

at 13.56 MHz is applied to the plasma source through an impedance matching network to maximize the power delivery to the plasma. An independent rf potential at 13.516 MHz is applied to the substrate holder to control the ion energy.[18–21] The slight frequency offset from 13.56 MHz is meant to avoid interference with the plasma source frequency. The rf power delivered to the plasma source is maintained at 400–800 W, while the rf power delivered to the substrate is maintained at 100–150 W. Although the reactor geometry departs from commercial reactors, we expect the plasma conditions (e.g., density of reactive neutrals and ions, ion energy distribution, and ion angular distribution) to be comparable to those of commercial reactors based on the aforementioned rf power consumption and independent bias.

For ATR-FTIRS, a trapezoidally shaped Ge(111) crystal is mounted on a metal substrate platen cooled by chilled water. The choice of the ATR substrate (e.g., GaAs and KRS-5) is immaterial, provided that the substrate is IR-transparent in the mid-IR region. We also expect the effect of the substrate on etch chemistry to be minimal prior to the endpoint, since reactive radicals and energetic ions impinge almost entirely on the film surface rather than on the underlying ATR substrate during etching. Figure 2 illustrates that the Ge crystal is coated with mesoporous SiO_2, anti-reflective coating (ARC), and fully patterned photoresist (PR). The patterned PR serves as an etch mask. The preparation steps for these films will be discussed in the following section.

The Ge crystal is 5 cm long, 2 cm wide, and 2 mm thick with two ends beveled at 45° with respect to the substrate normal. The IR beam from a FTIR spectrometer (Nicolet Nexus 670) is focused onto a beveled edge. Upon entering the crystal, the IR beam undergoes ~12 reflections from the crystal top surface and exits the opposite beveled edge. The above arrangement renders the Ge crystal transparent to IR above 650 cm^{-1}. The multiple internal reflections also significantly enhance the IR absorption that stems from the changes in the film on top of the ATR substrate, especially in comparison to the IR absorption that stems from the changes on the beveled edges. In addition, the flat crystal surface area is ~10 times greater than that of the two beveled edges. Note also that the beveled edges face away from the plasma source and that the oxide film, deposited on the longer side of the ATR crystal, faces the plasma source (Figs. 1 and 2). These factors ensure that the absorption signal from the film is at least two orders of magnitude greater than that from the beveled edges. The ATR technique is also robust against scratches inadvertently introduced during sample transfer to and from the plasma chamber. We have previously demonstrated that the surface corrugation, whether introduced natu-

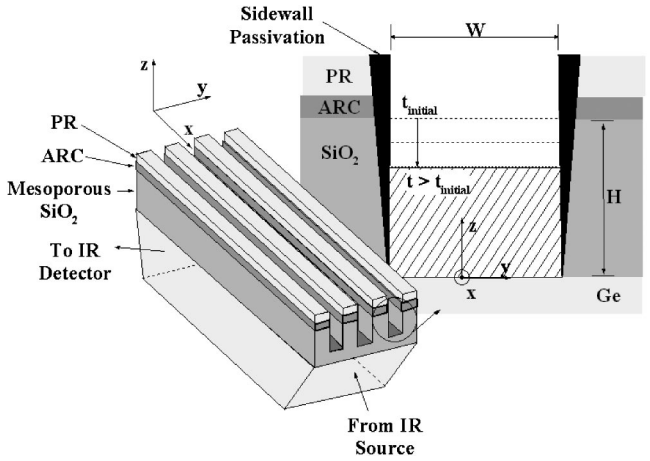


FIG. 2. Conceptual layout of a partially patterned SiO_2 film on a Ge ATR crystal. The ATR technique is used to measure the removal rate of SiO_2 as well as the formation rate of C–F in the sidewall passivation layer. Patterned photoresist and anti-reflective coating are stacked on top of mesoporous SiO_2 film. The magnified cross-sectional view outlines the etch front at the trench bottom.

rally or artificially, does not cause IR to scatter upon internal reflection, provided that the corrugation dimension is less than the IR wavelength.[22]

The internal reflections create an evanescent electromagnetic field whose strength decays exponentially away from the film–crystal interface. Infrared-active dipole moments in the etched mesoporous SiO_2 film (e.g., C–F and Si–O–Si stretching vibrational modes) absorb the evanescent field at their characteristic vibrational frequencies. For instance, the IR absorbance peaks of C–F_x ($x=1$, 2, and 3) and Si–O–Si stretching vibrational modes appear at 1200–1700 and 890–1195 cm^{-1}, respectively.[14,23,24] Because the evanescent field strength decays exponentially away from the film–crystal interface, the IR absorption by dipole moments near the film–crystal interface is more pronounced than that away from the interface. The mathematical correction for the exponential decay will be described in Result and Discussion. The exiting IR is collimated by a KBr lens and then focused onto a HgCdTe mid-range IR detector by an off-axis paraboloid mirror. Prior to etching, a background spectrum is taken with 4 cm^{-1} resolution averaged over 50 scans. During etching, a series of sample spectra are collected with 4 cm^{-1} resolution. Each absorbance spectrum is averaged over 20 scans to maintain the peak-to-peak noise level below 0.016, while achieving the time resolution of 12 s.

Figure 2 depicts partially etched parallel trenches created in the mesoporous SiO_2 film during etching. The magnified view of a trench shows how PR, ARC, and mesoporous SiO_2 film are stacked on top of the Ge ATR crystal. To create this stack of films, the Ge crystal is first cleaned in 100% ethanol (EtOH) to dissolve organic contaminants on the surface. The substrate is then treated in air plasma for 2 min to increase the adhesion of a sol that contains SiO_2 film precursors. The sol consists of EtOH, H_2O, tetraethyl orthosilicate (TEOS), polyoxyethylene(10) cetyl ether (Brij 56), and a trace amount of HCl. We prepare the sol by a 2-step process before spin-coating it on the Ge crystal at 2000 rpm for 20 s. The first step is to create a stock solution ($A2^{**}$) by mixing EtOH, H_2O, TEOS, and HCl at a molar ratio of 3.8: 1: 1: 5.1×10^{-5} and heating the solution at 333 K for 90 min. The purpose of heating is to expedite the hydrolysis of TEOS, replacing ethoxy groups with hydroxyl groups.[25] The second step is to dilute the prepared $A2^{**}$ stock solution with additional EtOH, H_2O, and HCl. Brij 56 is also added to set the molar composition of EtOH, H_2O, TEOS, Brij 56, and HCl at 20: 5: 1: 0.06: 4×10^{-3}. The HCl concentration is chosen to minimize premature condensation of TEOS while promoting the self-assembly of Brij 56 surfactant to form either a hexagonal or cubic phase template.[25] This template formation occurs concurrently with evaporation of EtOH and H_2O; hence, the template formation is named evaporation induced self-assembly (EISA).[26] The prepared sol is then spin-coated on the Ge crystal and allowed to dry in the ambient air (298 K and 30% relative humidity). The TEOS oligomers in the sol subsequently condense around the template upon calcining the spin-coated film in an oven filled with the ambient air. The oven temperature is ramped at 1 K/min to 673 K, maintained at this temperature for 180 min, and then cooled down to the room temperature (298 K) at the same rate. The removal of organic materials by calcination results in mesopores SiO_2 matrix. As control samples, we also prepare the mesoporous films on Si(100) under the same condition. These control samples are used to measure the etch rate and the post-etch profile, using XSEM. The SiO_2 film thickness is measured prior to coating ARC and PR, using variable-angle spectroscopic ellipsometry (Woolam WVASE32). The SiO_2 film thickness is ~2600 Å for all samples.

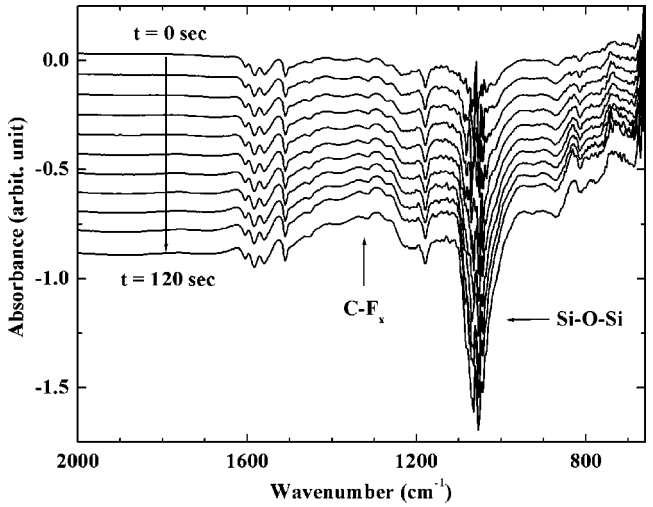


FIG. 3. Time-series of IR absorbance spectra taken during a 2 min etch process show both C–F and Si–O–Si stretching vibrational modes. Si–O–Si absorbance becomes increasingly negative, indicating its loss, whereas C–F absorbance becomes increasingly positive, indicating its gain during the etch.

The anti-reflective coating (Brewers Science XHRi-16) and photoresist (Shipley i300) are sequentially spin-coated on top of the SiO_2 film at 4000 rpm for 30 s. The thickness of ARC and PR is 1600 and 8000 Å, respectively. The photoresist is patterned by interferometric lithography (IL) described elsewhere.[27–30] The photoresist is exposed to the 355 nm line of a Nd:YaG laser operated at 55 mJ/pulse for 20 s, baked at 383 K, and then immersed in a developer solution (Shipley MF702) for 1 min. The pitch (P) and width (W) of the fully developed PR pattern are 0.3 and 0.15 μm, respectively. The underlying ARC is etched either in an O_2–Ar plasma or in a CHF_3–Ar plasma, depending on the desired etch profile. The O_2–Ar etch results in an anisotropic profile, whereas CHF_3–Ar plasma results in a profile where the midsection bows out. The rf power delivered to the plasma source and substrate platen, pressure, and flow rates of O_2 and Ar are 100 W, 50 W, 1.3 Pa, 4 sccm, and 25 sccm, respectively.

III. RESULTS AND DISCUSSION

Figure 3 shows a typical set of time-series absorbance spectra taken during a 2 min etch to completely remove the exposed 2600 Å thick SiO_2 under the etch mask. The absorbance by Si–O–Si stretching vibrational modes from 890 to 1195 cm^{-1} decreases continuously in the negative direction

(Fig. 3). This observation is consistent with the fact that SiO_2 is removed during etching. The pronounced noise level near 1050 cm^{-1} is due to the fact that Si–O–Si stretching mode is a strong IR absorber and that the IR intensity of the reference background in the said region is minimal with a 2600 Å thick SiO_2 layer.

In contrast to the Si–O–Si absorbance, the absorbance by $C–F_x$ vibrational modes near 1200–1700 cm^{-1} increases continuously, indicating that the total amount of fluorocarbon (FC) deposit increases throughout the etching process. This increase may stem from fluorocarbon accumulation on the sidewalls[31,32] and/or on the trench bottom. The work of Rueger *et al.*[33] and Standaert *et al.*,[34] using angle-resolved x-ray photoelectron spectroscopy (AR-XPS), demonstrates that a thin layer of FC exists on SiO_2 surfaces even under heavy ion bombardment. Thus, we expect the FC to exist both on the trench bottom as well as on the sidewalls. However, the heavy ion bombardment on an independently biased substrate is believed to maintain the FC accumulation at minimum.[33,35] The fluorocarbon would, therefore, accumulate mostly on the sidewalls.[32] In addition, the trench bottom area remains fairly constant throughout the etch, whereas the sidewall area continues to increase. Note also that we have discounted the FC deposit on trench tops from the possible cases discussed above. We have determined that the IR absorption contribution from the FC deposition on trench tops is minimal when the SiO_2 thickness is comparable to or significantly exceeds the depth of penetration of the evanescent wave. That is, when the FC deposit on trench top is placed outside or at a distance comparable to the probing depth of the evanescent wave, the observed IR absorption pertains to either the sidewalls or the trench bottom. Therefore, we deduce that the observed increase pertains mostly to the sidewalls.

Two mechanisms may contribute to the increase in fluorocarbon. (1) The fluorocarbon deposits in the sub-surface pores near the exposed SiO_2 sidewalls.[36] For porous SiO_2 with 2 nm wide pores and 30% porosity, similar to our mesoporous SiO_2, the sub-surface pore filling may reach a depth of ~3 to 4 nm.[36] As the etching exposes progressively increasing SiO_2 sidewall area with concurrent pore filling, the total amount of fluorocarbon increases; (2) the fluorocarbon deposits externally on top of the exposed SiO_2 sidewalls, and the total amount increases with increasing SiO_2 sidewall area. We assume that the rate of diffusion of fluorocarbon precursors through the mesopores (~2 nm) far exceeds the rate at which the exposed SiO_2 surface increases and that the precursor diffusion is self-limiting due to the fluorocarbon film sealing the mesopores.[32,36] Thus, we expect that the increase in $C–F_x$ absorbance is primarily due to the FC build-up on top of the exposed SiO_2 sidewalls.

For a quantitative analysis to obtain the absolute amount of SiO_2 removal and $C–F_x$ accumulation during etching, the IR absorbance spectra in Fig. 3 need to be corrected for the exponentially decaying evanescent field. A similar correction procedure has been previously described.[13,37] To summarize, we employ the effective thickness (d_e) approximation[38,39] where the IR absorbance (A) is normalized by d_e. The absorbance A corresponds to the SiO_2 film remaining at any given moment during etching. A is approximated by the sum of two absorbance spectra:

$$A = A_o + A_{\text{exp}}, \tag{1}$$

where A_o is the absorbance spectrum representing the pre-etch SiO_2 film taken with respect to the clean Ge ATR crystal, and A_{exp} is the experimentally measured absorbance spectrum taken with respect to a pre-etch background where the patterned PR and ARC are stacked on top of yet-to-be-etched SiO_2. A and d_e are related to the extinction coefficient (k) of the mesoporous SiO_2 film by

$$\tilde{k}(\tilde{\nu}) = \frac{A(\tilde{\nu})}{4\pi N_R d_e(\tilde{\nu})}, \tag{2}$$

where $\tilde{\nu}$ and N_R denote the wave number in cm^{-1} and the number of internal reflections from the SiO_2–Ge interface, respectively. N_R is 12 in this case. Since d_e is yet unknown, the pre-etch SiO_2 film thickness is used as the initial value of d_e. k is related to the real part (n) of the complex refractive index ($\tilde{n}=n+ik$) for the SiO_2 film by the Kramers–Kronig dispersion relation:

$$n_i = 1 + \frac{2}{\pi}\left[P\int_0^{\infty} \frac{\tilde{\nu}k(\tilde{\nu})}{\tilde{\nu}^2 - \tilde{\nu}_i^2} d\tilde{\nu} \right], \tag{3}$$

where n_i is the refractive index of the film at wave number $\tilde{\nu}_i$, and P denotes the Cauchy principle value of the integral from zero to infinity. However, the experimentally collected spectrum has a finite wave number range within the mid-IR region, and the integral cannot be evaluated from zero to infinity. This limitation can be circumvented by measuring a real part (n_r) of the complex refractive index at a known wave number $\tilde{\nu}_r$ and subtracting n_r from n_i. The inherent assumption is that the difference between n_i and n_r is close to zero in the domains outside the experimentally accessible wave number range: i.e., $0 \leqslant \tilde{\nu} < \tilde{\nu}_1$ and $\tilde{\nu}_2 < \tilde{\nu} \leqslant \infty$. This assumption leads to

$$n_i = n_r + \frac{2}{\pi}\left[P\int_{\tilde{\nu}_i}^{\tilde{\nu}_2} \frac{\tilde{\nu}k(\tilde{\nu})}{\tilde{\nu}^2 - \tilde{\nu}_i^2} d\tilde{\nu} - P\int_{\tilde{\nu}_i}^{\tilde{\nu}_2} \frac{\tilde{\nu}k(\tilde{\nu})}{\tilde{\nu}^2 - \tilde{\nu}_r^2} d\tilde{\nu} \right], \tag{4}$$

where n_r is measured at $\tilde{\nu}_r = 15\,800$ cm^{-1} by spectroscopic ellipsometry (SE).[13] The experimentally measured value of n_r is 1.26 for the mesoporous SiO_2 film. The limits ($\tilde{\nu}_1$ and $\tilde{\nu}_2$) of integration in the principle integral are 650 and 4000 cm^{-1}, respectively.

The value of n_i from Eq. (4) is then used to calculate the two components of the evanescent field: $E_{0\|}$ that is parallel to the substrate surface and $E_{0\perp}$ that is perpendicular to the substrate surface. The strength of these two components depends on n_i of the mesoporous SiO_2; $n_1(\tilde{\nu})$, the real part of the complex refractive index for Ge ATR crystal; and $n_3(\tilde{\nu})$, the real part of the complex refractive index for vacuum. $n_3(\tilde{\nu})$ approximates the refractive index of high vacuum in the etch chamber. Denoting n_i as $n_2(\tilde{\nu})$ for convenience, $E_{0\|}$ is expressed as

$$E_{0\parallel} = \frac{2\cos\theta[(1+n_{32}^4)\sin^2\theta - n_{31}^2]^{1/2}}{(1-n_{31}^2)^{1/2}[(1+n_{31}^2)\sin^2\theta - n_{31}^2]^{1/2}}, \quad (5)$$

where n_{lm} ($l=3$ and $m=1$ or 2) denotes n_l-to-n_m ratio. The perpendicular component, $E_{0\perp}$, is expressed as

$$E_{0\perp} = \frac{2\cos\theta}{(1-n_{31}^2)^{1/2}}. \quad (6)$$

The value of n_i from Eq. (4) is also used to calculate the characteristic penetration depth (d_p) of the evanescent wave by

$$d_p = \frac{\lambda/n_1}{2\pi(\sin^2\theta - n_{21}^2)^{1/2}}, \quad (7)$$

where λ is the IR wavelength in vacuum, and θ is the incident angle of the IR on the top surface inside the ATR crystal. The effective thickness (d_e) is then expressed as

$$d_e = \frac{n_{21}E_0^2}{2\cos\theta} d_p[1-\exp(-2\delta/d_p)], \quad (8)$$

where δ is the remaining SiO_2 film thickness. δ is calculated by

$$\delta = \delta_o - (ER \times t), \quad (9)$$

where δ_o, ER, and t represent pre-etch SiO_2 film thickness, etch rate measured on a blank SiO_2 film, and etch duration, respectively. Since the SiO_2 film is partially masked during the pattern etch, δ is an underestimation of remaining SiO_2. Equation (9) also assumes that the vertical etch rate remains constant throughout etching. To account for possible loading effect and time-dependent etch rate, *in situ* spectroscopic ellipsometry[14,40] or interferometry[41,42] can be used for more reliable estimate on δ. Note that Eq. (8) results in two effective thickness values ($d_{e\parallel}$ and $d_{e\perp}$), depending on whether $E_{0\parallel}$ or $E_{0\perp}$ is used in place of E_0. As an approximation, the arithmetic average of $d_{e\parallel}$ and $d_{e\perp}$ obtained from Eq. (8) is used for d_e in Eq. (2) to calculate k. The above procedure using Eqs. (2)–(8) is repeated until n_i and k converge. To test their convergence, the real part (ε_1) of the complex dielectric constant ($\bar{\varepsilon}$) is first calculated by

$$\varepsilon_1 = n^2 - k^2, \quad (10)$$

where the subscript i for n_i is omitted for convenience. The imaginary part (ε_2) is then calculated by

$$\varepsilon_2 = 2nk. \quad (11)$$

The magnitude of $\bar{\varepsilon}$ is then expressed as

$$\bar{\varepsilon} = \sqrt{\varepsilon_1^2 + \varepsilon_2^2}. \quad (12)$$

We employ an absolute convergence criterion where the root mean square (rms) value defined in Eq. (13) is less than 10^{-2}:

$$\text{rms} = \sqrt{\sum_k^N (\bar{\varepsilon}_{k,l+1} - \bar{\varepsilon}_{k,l})^2/N}, \quad (13)$$

where k is the kth wave number, l is the lth iteration, and N is the number of experimental data points in the absorbance spectrum.

After d_e is self-consistently calculated by the procedure described above, the absolute amount of SiO_2 remaining at each moment during etching is approximated by

$$N_{\text{Si–O–Si}} = \frac{2.60\times10^{12}\ \text{cm}^{-1}}{N_R}\frac{1}{f}\int_{\tilde{\nu}_1}^{\tilde{\nu}_2}\frac{A_{\text{Si–O–Si}}(\tilde{\nu})d\tilde{\nu}}{d_e(\tilde{\nu})}\times\delta \times A_{\text{top}}, \quad (14)$$

where $N_{\text{Si–O–Si}}$ is the total number of Si–O–Si removed, f is the oscillator strength of Si–O–Si stretching vibrational mode, and A_{top} is the top surface area of the Ge ATR crystal. The integration is evaluated from $\tilde{\nu}_1 = 984\ \text{cm}^{-1}$ to $\tilde{\nu}_2 = 1195\ \text{cm}^{-1}$. The top surface area remains constant at 5 cm^2, while δ decreases according to Eq. (9). The oscillator strength is 7.38×10^{-5} calculated from [43]

$$f = 1.29\times10^{17}\frac{\alpha W}{N_{\text{mesoporous}}}\frac{n_{\text{IR}}}{(n_{\text{IR}}^2+2)^2}, \quad (15)$$

where α, W, $N_{\text{mesoporous}}$, and n_{IR} denote the peak absorption coefficient (1.72 cm^{-2}), the full width at half maximum (100 cm^{-1}) for the Si–O–Si absorbance band, the concentration of SiO_2 units ($\sim1.92\times10^{22}\ \text{cm}^{-3}$) in the mesoporous film, and the real part (1.84) of the complex refractive index near the 1050 cm^{-1} region. We approximate $N_{\text{mesoporous}}$ by

$$N_{\text{mesoporous}} = \frac{n_{\text{mesoporous}}}{n_{\text{thermal}}}N_{\text{thermal}}, \quad (16)$$

where $n_{\text{mesoporous}}$ and n_{thermal} denote the real part of the complex refractive index in the visible range for mesoporous SiO_2 at 1.26 and that of thermal SiO_2 at 1.46, respectively. The experimentally measured $n_{\text{mesoporous}}$ is identical to a reported value.[36] N_{thermal} is $2.2\times10^{22}\ \text{cm}^{-3}$ based on the specific gravity and molecular weight of thermally grown SiO_2.[44] n_{IR} of mesoporous SiO_2 is similarly approximated to be 1.84 by Eq. (16), based on n_{IR} of thermal SiO_2 at 2.12.

For C–F, the total number of C–F ($N_{\text{C–F}}$) created during etching is calculated from [45]

$$N_{\text{C–F}} = \frac{1}{N_R}\frac{1}{\sigma_{\text{C–F}}}\int_{\tilde{\nu}_1}^{\tilde{\nu}_2}\frac{A_{C\text{–}F}(\tilde{\nu})d\tilde{\nu}}{\tilde{\nu}d_e(\tilde{\nu})}\times\delta_{\text{C–F}}\times(ER\times t)\times L, \quad (17)$$

where $\sigma_{\text{C–F}}$ is the IR cross-section of C–F stretching vibrational mode ($2.84\times10^{-20}\ \text{cm}^2$),[46] $\delta_{\text{C–F}}$ is the thickness of the fluorocarbon sidewall passivation, and L is the length of the etched trenches (5 cm). The integration is evaluated from $\tilde{\nu}_1 = 1200\ \text{cm}^{-1}$ to $\tilde{\nu}_2 = 1700\ \text{cm}^{-1}$. For comparison purpose only, we assume that $\delta_{\text{C–F}}$ remains constant[32] at 1 nm,[47] contrary to the reported observation that $\delta_{\text{C–F}}$ tapers down towards the bottom of the etched trenches.[31] This assumption helps visualize how the sidewall passivation thickness

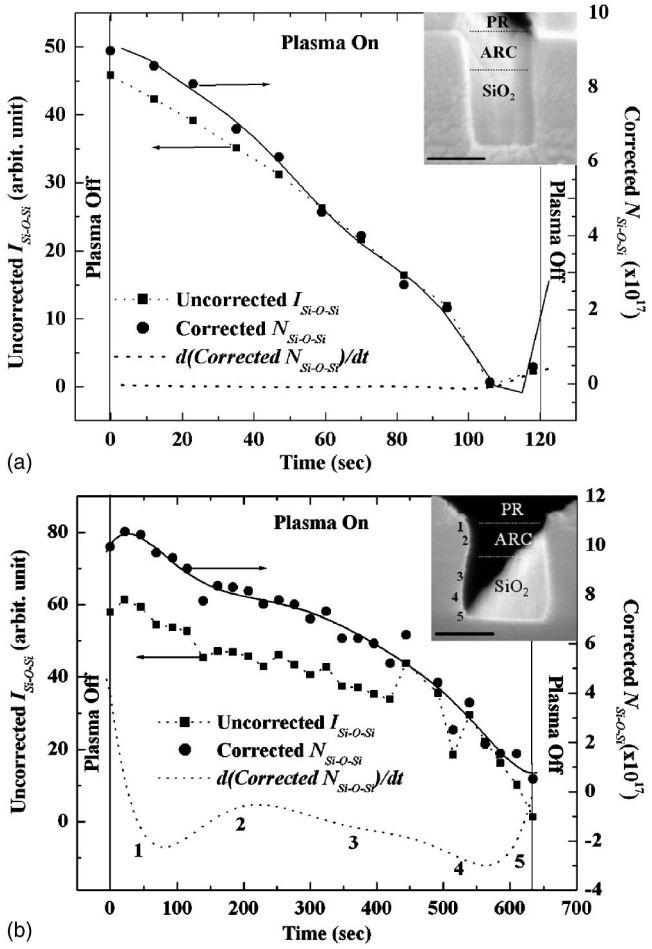


FIG. 4. Integrated absorbance of Si–O–Si is plotted vs time for (a) a sample with relatively anisotropic etch profile and (b) a sample with highly nonideal, bell-jar shaped etch profile. The corresponding cross-sectional SEM images are shown in the insets. The scale bar is 200 nm. The IR absorbance corrected for the exponentially decaying evanescent field is used to calculate the total loss of Si–O–Si bonds. A nonlinear slope of total Si–O–Si loss ($N_{\mathrm{Si–O–Si}}$) during plasma-on period foretells a deviation from perfect anisotropy. The nonideal profile is particularly pronounced for the profile shown in (b).

changes as a function of time. If $N_{\mathrm{C–F}}$ deviates from a linear line, then the slope of $N_{\mathrm{C–F}}$ describes a number of scenarios, as to whether the actual $\delta_{\mathrm{C–F}}$ tapers down towards the trench bottom, thickens towards the trench bottom, or thickens over the entire passivation layer while maintaining the general tapered down cross-section.

The SEM images in Fig. 4 illustrate two contrasting examples of profile evolution: (a) anisotropic profile and (b) highly nonideal, bell-jar shaped profile. The lateral opening of the SiO_2 trench is $\sim$0.15 μm for both profiles. The CHF_3 and Ar flow rates, rf power to the plasma source, and rf power to the substrate are maintained at 22 and 45 sccm, 800 W, and 150 W, respectively, for the anisotropic profile and at 11 and 50 sccm, 400 W, and 100 W, respectively, for the bell-jar shaped profile. Another difference is that the ARC is etched in an O_2–Ar plasma for 5 min for the anisotropic profile, whereas the ARC is etched under the identical conditions used in SiO_2 etch for the bell-jar shaped profile. The etch is stopped 30 s ($\sim$10% overetch) after the presumed endpoint based on the etch rates observed on identically patterned samples. Despite the overetch, we have determined that the underlying Ge etch is below detection limit.

Corresponding to the two etch profiles, Fig. 4 shows the integrated absorbance of Si–O–Si ($I_{\mathrm{Si–O–Si}}$) before correcting for the exponentially decaying evanescent wave. $I_{\mathrm{Si–O–Si}}$ qualitatively indicates that SiO_2 is removed approximately at a constant rate for the anisotropic profile and at a rate that varies throughout the etch for the bell-jar shaped profile. The detailed quantitative information on profile evolution is extracted from $N_{\mathrm{Si–O–Si}}$ based on Eq. (14). $N_{\mathrm{Si–O–Si}}$ accounts for the exponential decay, and its time derivative conveys the information on profile evolution. We employ a fifth-order polynomial function (—) to fit $N_{\mathrm{Si–O–Si}}$ (●). The smoothness of the polynomial fit reduces the noise in the calculated rate of decrease (---). The rate of decrease is equivalent to the removal rate of SiO_2. The removal rate accounts for both lateral and vertical losses; therefore, the removal rate is different from the blank film etch rate. For instance, the dotted line (---) in Fig. 4(a) indicates that the SiO_2 removal rate remains nearly constant until the very end of the etch process between 100 and 120 s. The reduced removal rate translates to narrowing profile, and the SEM image in Fig. 4(a) shows a slight footing near the bottom of the trench.

For the bell-jar shaped profile in Fig. 4(b), we demonstrate that the profile evolution of ARC and SiO_2 can be successively monitored. To this end, the IR absorbance of characteristic vibrational modes of ARC, which appear in the 1080 cm^{-1} region, is included in the calculation of $N_{\mathrm{Si–O–Si}}$. Thus, $N_{\mathrm{Si–O–Si}}$ partially reflects the removal rate of ARC in addition to that of SiO_2. The nonlinearity in the ARC and SiO_2 removal rate is consistent with the bulging profile of ARC and the lateral erosion of SiO_2. One can trace Points 1–5 on the removal rate curve (---) and match them with the corresponding points in the SEM image. Note that the removal rate is negative with its corresponding axis on the right hand side. Thus, the absolute removal rate increases with increasingly negative quantity. The observed removal rate suggests that the ARC etching lasts $\sim$350 s, after which SiO_2 etch begins at a comparatively accelerated pace. In Fig. 4(b), $I_{\mathrm{Si–O–Si}}$ at 350 s is $\sim$50, matching that of initial $I_{\mathrm{Si–O–Si}}$ in Fig. 4(a). Since the SiO_2 thickness of the sample in Fig. 4(a) is the same as that in Fig. 4(b), this equivalence validates the demarcation of when ARC etch ends, and SiO_2 etch begins, while supporting the positional accuracy of assigned numbers (Points 1–5) along the etch profile. In the future, we plan to compare the etch profiles, that are predicted and independently verified by XSEM, with experimentally measured ion energy and angular distribution functions in SiO_2, Si_3N_4, and SiO_xN_y cases where the etch is strongly driven by ion bombardment.

The absolute amount of C–F formation during etching, calculated from the absorbance of C–F stretching vibrational mode by Eq. (17), is shown in Fig. 5. Figure 5(a) corresponds to Fig. 4(a), and Fig. 5(b) corresponds to Fig. 4(b). In contrast to $N_{\mathrm{Si–O–Si}}$, $N_{\mathrm{C–F}}$ increases throughout the etch. The

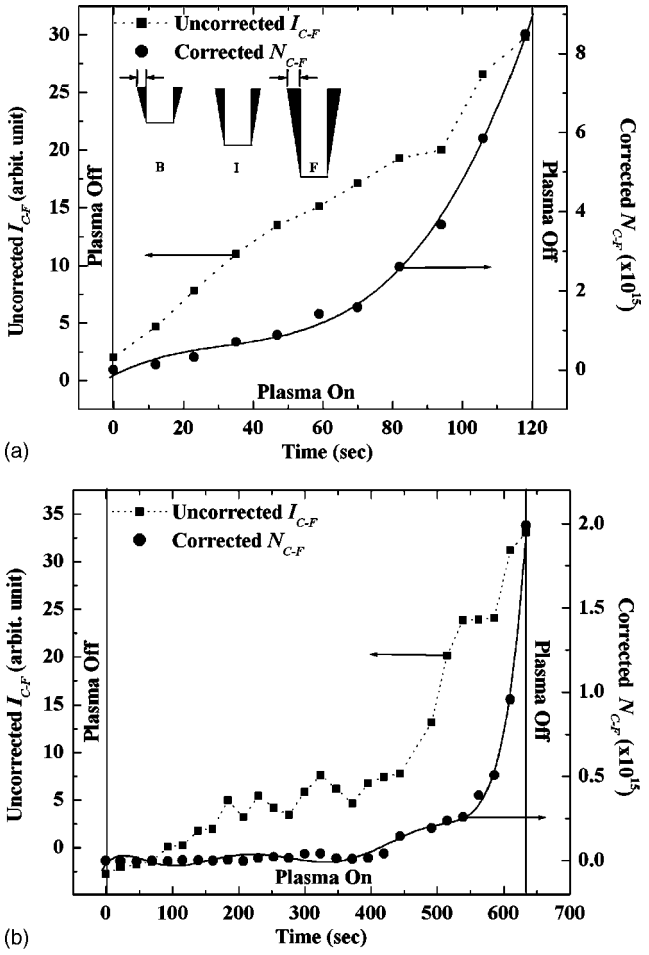


FIG. 5. IR absorbance of C–F stretching vibrational mode near 1300 cm^{-1} corrected for the exponentially decaying evanescent field strength is used to calculate the total number of C–F formed during etching for (a) a sample with relatively anisotropic etch profile and (b) a sample with highly non-ideal, bell-jar shaped etch profile. The inset in (a) conceptually describes the shape and thickness of sidewall passivation in the beginning (B), intermediate (I), and final (F) stage of etching. Note that the inset is not drawn to scale. The C–F buildup is pronounced towards the end point, and a critical buildup is necessary to maintain the anisotropic profile.

C–F accumulation rate is calculated from the slope of N_{C-F}, in the same way the SiO_2 removal rate is calculated. The accumulation rate, which rises sharply towards the end of the etching process, suggests that the most of C–F accumulation occurs towards the end, when the trench bottom meets the underlying Ge ATR crystal. If the sidewall passivation profile is assumed to maintain its wedge shape tapering down towards the trench bottom during etching, the sharp rise in N_{C-F} translates to a profile evolution schematically drawn in the inset of Fig. 5(a). B, I, and F in the inset represent beginning, intermediate, and final stage of etching, respectively. The profile maintains its downward wedge shape, but its overall thickness rises sharply near the endpoint. Both anisotropic and bell-jar shaped samples commonly share this trend. The only dramatic difference between the two samples is the absolute amount of C–F accumulation. The anisotropic profile exhibits a high level of C–F bond accumulation throughout the etch (8×10^{15} total when the etch is complete), whereas the bell-jar shaped profile exhibits a comparatively low level of C–F bond accumulation (2×10^{15} total at the end). Given the same SiO_2 thickness and trench top opening width, we infer that a critical amount of C–F exists, below which a substantial lateral loss in SiO_2 occurs. That is, the sample corresponding to Figs. 4(b) and 5(b) does not have enough fluorocarbon buildup on the sidewalls and, therefore, not enough passivation thickness to prevent the lateral loss.

IV. CONCLUSIONS

We have demonstrated the use of ATR-FTIRS to monitor the etch profile evolution of mesoporous, low-dielectric-constant SiO_2 *in situ* and in real time. The technique is non-intrusive as well as nondestructive, eliminating the necessity to cleave the etched samples for SEM imaging. We rely on the IR absorbance of Si–O–Si stretching vibrational mode during etching to calculate the SiO_2 loss. When the absorbance spectra are corrected for the exponentially decaying strength of the evanescent wave, using the algorithm described in Results and Discussion, the total amount of SiO_2 loss can be monitored in real time, and the removal rate is calculated from the slope of total SiO_2 loss. The calculated removal rate and the corresponding etch profile show good agreement, further validating that the ATR technique can be used to predict the profile evolution during etching. In a similar approach, we have investigated the fluorocarbon accumulation during etching. The total accumulation of C–F and the corresponding profile suggest that a critical amount of C–F buildup is needed to maintain an anisotropic profile and that the most of C–F buildup occurs towards the end of etching.

ACKNOWLEDGMENTS

The authors thank University of New Mexico SEED and the National Science Foundation CAREER (Award Number DMR-0094145) for generous financial support.

[1]International Technology Roadmap for Semiconductors, http://public.itrs.net/Files/2003ITRS/Home2003.htm (2003).
[2]J. Bremmer, Solid State Technol. Suppl. S3 (2001).
[3]S. Takeishi, H. Kudo, R. Shinohara, M. Hoshino, S. Fukuyama, J. Yamaguchi, and M. Yamada, J. Electrochem. Soc. **144**, 1797 (1997).
[4]T. E. F. M. Standaert, P. J. Matsuo, X. Li, G. S. Oehrlein, T. M. Lu, R. Gutmann, C. T. Rosenmayer, J. W. Bartz, J. G. Langan, and W. R. Entley, J. Vac. Sci. Technol. A **19**, 435 (2001).
[5]J. L. Hedrick, K. R. Carter, J. W. Labadie, R. D. Miller, W. Volksen, C. J. Hawker, D. Y. Yoon, T. P. Russell, J. E. McGrath, and R. M. Briber, Adv. Polym. Sci. **141**, 1 (1999).
[6]H. J. Cha, J. Hedrick, R. A. DiPietro, T. Blume, R. Beyers, and D. Y. Yoon, Appl. Phys. Lett. **68**, 1930 (1996).
[7]J. L. Hedrick, R. D. Miller, C. J. Hawker, K. R. Carter, W. Volksen, D. Y. Yoon, and M. Trollsas, Adv. Mater. (Weinheim, Ger.) **10**, 1049 (1998).
[8]C. V. Nguyen, K. R. Carter, C. J. Hawker, J. L. Hedrick, R. L. Jaffe, R. D. Miller, J. F. Remenar, H. W. Rhee, P. M. Rice, M. F. Toney, M. Trollsas, and D. Y. Yoon, Chem. Mater. **11**, 3080 (1999).
[9]C. J. Brinker, Curr. Opin. Solid State Mater. Sci. **1**, 798 (1996).
[10]L. W. Hrubesh, L. E. Keene, and V. R. Latorre, J. Mater. Res. **8**, 1736 (1996).
[11]S. J. Ullal, T. W. Kim, V. Vahedi, and E. S. Aydil, J. Vac. Sci. Technol. A **21**, 589 (2003).
[12]S. M. Han and E. S. Aydil, J. Vac. Sci. Technol. A **14**, 2062 (1996).

[13]S. M. Han and E. S. Aydil, J. Appl. Phys. **83**, 2172 (1997).
[14]D. C. Marra and E. S. Aydil, J. Vac. Sci. Technol. A **15**, 2508 (1997).
[15]S. J. Ullal, H. Singh, J. Daugherty, V. Vahedi, and E. S. Aydil, J. Vac. Sci. Technol. A **20**, 1195 (2002).
[16]Z. Zhou, E. S. Aydil, R. A. Gottscho, Y. J. Chabal, and R. Reif, J. Electrochem. Soc. **140**, 3316 (1993).
[17]S. M. Han and E. S. Aydil, J. Vac. Sci. Technol. A **14**, 2062 (1996).
[18]J. R. Woodworth, I. C. Abraham, M. E. Riley, P. A. Miller, T. W. Hamilton, B. P. Aragon, and C. G. Willison, J. Vac. Sci. Technol. A **20**, 873 (2002).
[19]R. J. Hoekstra and M. J. Kushner, J. Appl. Phys. **79**, 2275 (1996).
[20]J. R. Woodworth, M. E. Riley, D. C. Meister, B. P. Aragon, M. S. Le, and H. H. Sawin, J. Appl. Phys. **80**, 1304 (1996).
[21]R. A. Gottscho, J. Vac. Sci. Technol. B **11**, 1884 (1993).
[22]H. Gerung, D. Doshi, C. J. Brinker, and S. M. Han, AVS 49th International Symposium, Denver, CO (2002).
[23]K. Ishikawa and M. Sekine, J. Appl. Phys. **91**, 1661 (2002).
[24]C. T Kirk, Phys. Rev. B **38**, 1255 (1988).
[25]C. J. Brinker and G. W. Scherer, *Sol-Gel Science*, 1st ed. (Academic, San Diego, 1990), pp. 108–216.
[26]C. J. Brinker, Y. F. Lu, A. Sellinger, and H. Y. Fan, Adv. Mater. (Weinheim, Ger.) **11**, 579 (1999).
[27]X. L. Chen and S. R. J. Brueck, Opt. Lett. **24**, 124 (1999).
[28]S. H. Zaidi and S. R. J. Brueck, J. Vac. Sci. Technol. B **11**, 658 (1993).
[29]E. H. Anderson, C. M. Horwitz, and H. I. Smith, Appl. Phys. Lett. **43**, 874 (1983).
[30]L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll, Appl. Opt. **17**, 1165 (1978).
[31]G. S. Oehrlein and Y. Kurogi, Mater. Sci. Eng., R. **24**, 153 (1998).
[32]A. Sankaran and M. J. Kushner, Appl. Phys. Lett. **82**, 1824 (2003).
[33]N. R. Rueger, J. J. Beulens, M. Schaepkens, M. F. Doemling, J. M. Mirza, T. E. F. M. Standaert, and G. S. Oehrlein, J. Vac. Sci. Technol. A **15**, 1881 (1997).
[34]T. E. F. M. Standaert, C. Hedlund, E. A. Joseph, G. S. Oehrlein, and T. J. Dalton, J. Vac. Sci. Technol. A **22**(1), 53 (2004).
[35]E. Gogolides, P. Vauvert, G. Kokkoris, G. Turban, and A. G. Boudouvis, J. Appl. Phys. **88**, 5570 (2000).
[36]T. E. F. M. Standaert, E. A. Joseph, G. S. Oehrlein, A. Jain, W. N. Gill, P. C. Wayner, Jr., and J. L. Plawsky, J. Vac. Sci. Technol. A **18**, 2742 (2000).
[37]S. M. Han and E. S. Aydil, J. Electrochem. Soc. **144**, 3963 (1997).
[38]N. J. Harrick, *Internal Reflection Spectroscopy*, 1st ed. (Wiley, New York, 1967), p. 42.
[39]F. M. Mirabella, Jr., in *Internal Reflection Spectroscopy*, 1st ed. (Marcel Dekker, New York, 1993), p. 17.
[40]W. M. Duncan and S. A. Henck, Appl. Surf. Sci. **63**, 9 (1993).
[41]T. J. Dalton, W. T. Conner, and H. H. Sawin, J. Electrochem. Soc. **141**, 1893 (1994).
[42]D. Economou, E. S. Aydil, and G. Barna, Solid State Technol. **34**, 107 (1991).
[43]K. Awazu, J. Non-Cryst. Solids **260**, 242 (1999).
[44]W. R. Runyan and K. E. Bean, *Semiconductor Integrated Circuit Processing Technology* (Addison-Wesley, New York, 1990), p. 58.
[45]M. H. Brodsky, M. Cardona, and J. J. Cuomo, Phys. Rev. B **16**, 3556 (1977).
[46]X. Wang, H. Harris, H. Temkin, S. Gangopadhyay, M. D. Strathman, and M. West, Appl. Phys. Lett. **78**, 3079 (2001).
[47]T. E. F. M. Standaert, M. Schaepkens, N. R. Rueger, P. G. M. Sebel, G. S. Oehrlein, and J. M. Cook, J. Vac. Sci. Technol. A **16**, 239 (1998).

HTML ABSTRACT + LINKS

Quantitative determination of tensile stress creation during island coalescence using selective-area growth

S. J. Hearne, S. C. Seel, J. A. Floro, and C. W. Dyck
Sandia National Laboratories, Albuquerque, New Mexico 87185-1415

W. Fan and S. R. J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

(Received 15 October 2004; accepted 21 January 2005; published online 7 April 2005)

Island coalescence during Volmer–Weber thin film growth is generally accepted to be a source of tensile stress. However, the stochastic nature of unpatterned film nucleation and growth prevents meaningful comparison of stress measurements taken during growth to that predicted by theoretical models. We have overcome this by systematically controlling island geometry using selective growth of Ni films on patterned substrates via electrodeposition. It was determined that the measured power-law dependence of mean stress on island size agreed well with theory. However, our data clearly demonstrates that the majority of the tensile stress associated with coalescence actually occurred after the initial contact of neighboring islands as the film planarizes with additional deposition. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1870109]

I. INTRODUCTION

The effective mechanical properties of a material are highly dependent upon the residual stress state introduced during fabrication. This is particularly true for thin films where the residual stress can be a significant portion of the yield strength. Unfortunately, the evolution of intrinsic stress during deposition of Volmer–Weber thin films is quite complex and still highly debated.[1–3] The only region where there is a consensus as to the mechanism is during island coalescence, where tensile stresses have been observed to exceed one gigapascal.[4] Conceptually, tensile stress is generated when adjacent islands initially touch one another, and then elastically deform to contact over a finite area in order to reduce the overall surface energy. Existing theoretical models for coalescence stress all determine the mean tensile stress in mechanical equilibrium, as a function of island size and geometry at the moment of coalescence.[1,4–8] However, computational necessity requires use of highly simplified island geometries with uniform sizes and simultaneous coalescence. In real films, coalescence events are stochastically distributed in time and occur among islands with a broad range of sizes and shapes. Furthermore, multiple mechanisms for stress generation can be operating at the same time. As a result, it is not possible to quantitatively equate theoretical predictions for the tensile coalescence stress with existing measurements in stochastically nucleated films and a meaningful comparison is only obtainable when island geometries are systematically controlled.

We obtain both the functional dependence of mean tensile stress on island radius, and the absolute magnitude of the stress, by measuring stress changes during electrodeposition of Ni islands where the coalescence process was constrained via lithographically-defined island nucleation sites and selective-area growth (Fig. 1). We demonstrate that the experimentally measured coalescence stress is in good agreement with the predictions from the Hertzian contact model of Freund and Chason[7] and with two-dimensional finite element (FE) analysis. We have also determined that the initial coalescence stress was the minority component of the total stress created during the coalescence and planarization of the film. This determination lays the groundwork for future analysis of the postcontact stress generation process, which is considerably more complex (and beyond the scope of this paper).

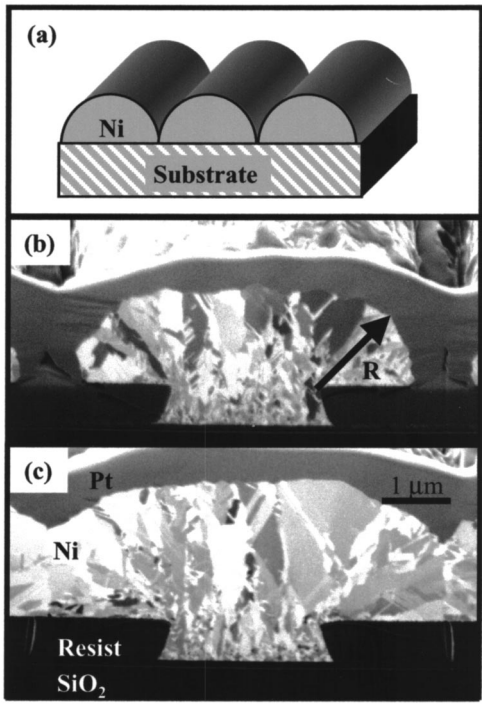


FIG. 1. (a) Schematic of an idealized cylindrical structure, (b) ion channeling focused ion beam (ICFIB) cross-section image of actual structure prior to coalescence, and (c) ICFIB image of the actual structure after coalescence. Note, the islands were coated after electrodeposition with Pt for processing in the FIB.

 © 2005 American Institute of Physics

Downloaded 18 Apr 2005 to 134.253.26.12. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

II. EXPERIMENTAL TECHNIQUES

Motivated by recent models for coalescence stress by Freund and Chason[7] and by Nix and Clemens,[6] we chose to experimentally evaluate the "two-dimensional" coalescence geometry, where the islands were idealized as an array of parallel half-cylinders, as shown in Fig. 1(a). In order to accomplish this, periodic trench arrays were patterned into a photoresist layer to expose a 1500 Å thick Au nucleation layer. This Au film was grown on a Ti adhesion layer on a thermally oxidized Si (001) substrate. Island size was controlled by varying the spacing of the trenches, while keeping the nominal trench pitch-to-width ratio $(d_{\text{pitch}}/d_{\text{trench}})=2.65$. The effective cylinder radii, R, ranged from 0.3 μm to 26 μm. Ni films were potentiostatically electroplated from an additive-free Ni–sulfamate bath at 40 °C. Prior to electrodeposition of the coalescence samples, the bath was conditioned at 10 mA/cm^2 for 4 h to remove trace ionic contaminants such as Fe and Co. Additionally, ultrahigh purity N_2 was bubbled through the bath for a minimum of 24 h, which increased run-to-run reproducibility. The plating efficiency was determined by comparing the Faradic current to thicknesses measured using Rutherford backscattering spectrometry. During deposition, the substrate curvature induced by thin-film stress was measured using laser deflectometry,[3] with the wafer clamped in a cantilever configuration. Because film thicknesses were often an appreciable fraction of the substrate thickness, the usual Stoney's equation did not hold, and the following relation for curvature (κ) was used:[9]

$$\kappa = \frac{6h_f\sigma}{h_s^2}\left[\frac{1+H}{1+HY(4+6H+4H^2)+H^4Y^2}\right], \tag{1}$$

where H is the ratio of the film thickness (h_f), to the substrate thickness (h_s), σ is the mean film stress, and Y is the ratio of the film modulus to the substrate modulus. Young's modulus was used rather than the biaxial modulus due to the uniaxial loading geometry used here.

III. RESULTS

Figures 1(b) and 1(c) demonstrate the selective lateral growth that can be obtained using electrodeposition. At coalescence, the islands were not true half-cylinders, as shown in Figs. 1(b) and 1(c), and we will account for this in our analysis. Figure 2 shows the measured evolution of "stress-thickness," σh_f, calculated from curvature using Eq. (1). In the figure, Ni was electrodeposited into an array of trenches that yielded nominally 4.2 μm radius cylinders [with the radius as defined in Fig. 1(b)]. The first rise in the tensile stress was the result of random Ni nucleation and coalescence within the trenches themselves to form the Ni lines. The second tensile rise was due to the coalescence of adjacent cylindrical islands, and was the stress change that we analyzed. In order to do this, we found it necessary to differentiate between the stress created during the initial coalescence event and the subsequent stress created as the film evolved from cylinders towards a planar surface.

During potentiostatic deposition plating current is linearly proportional to the surface area. Therefore, since the surface roughness of a film is maximum at the moment of coalescence there must be a corresponding maximum in the plating current at the same moment. To quantify this we developed a geometric model that predicts the evolution of the surface area and in turn, the deposition current of a cylindrical geometry as shown in Fig. 3. Prior to coalescence, the surface is given by the sum of two quarter cylinders and the trench that grow radially outwards due to the growth direction being normal to the growth surface during electrodeposition [Fig. 3(a)]. This results in the total current (I_{tot}) of the wafer prior to coalescence being

$$I_{\text{tot}} = (\pi r + w_{\text{tr}})\frac{dr}{dt}\frac{dq}{dV}n_{\text{tr}}w_s, \tag{2}$$

where, r is the island radius, n_{tr} is the number of trench across the samples, w_s width of the sample, w_{tr} width of the trench, dr/dt is the deposition rate, and dq/dV is the plating efficiency measured to be 3.7×10^7 mC/cm^3 from RBS

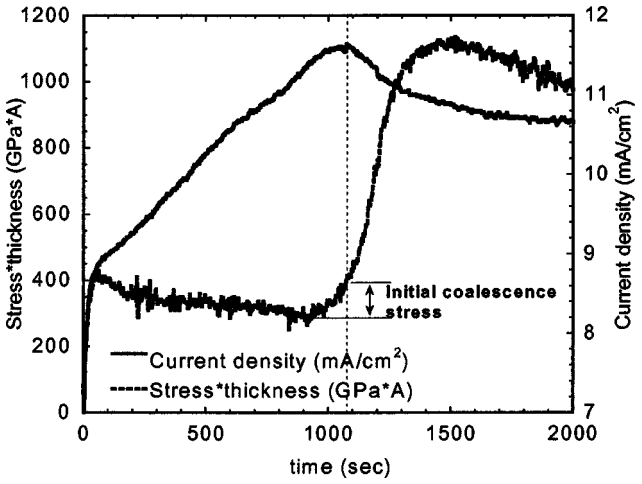


FIG. 2. Stress-thickness vs time plot for a 4.2 μm radius coalescence sample showing a measured change in stress*thickness during coalescence of 96 GPa Å.

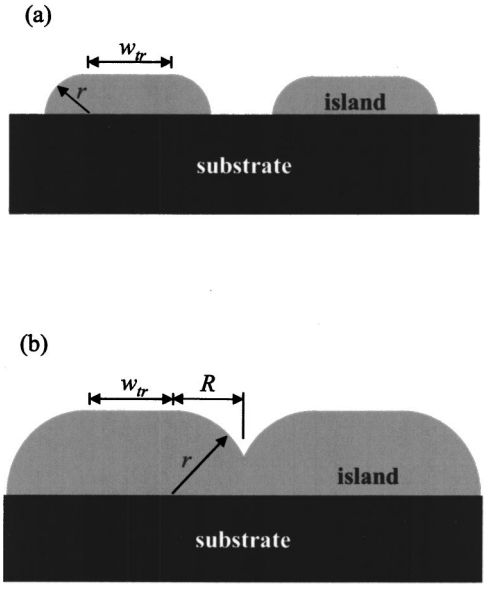


FIG. 3. Idealized cross-sectional geometry used in current evolution model, (a) prior to coalescence, (b) postcoalescence.

Downloaded 18 Apr 2005 to 134.253.26.12. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

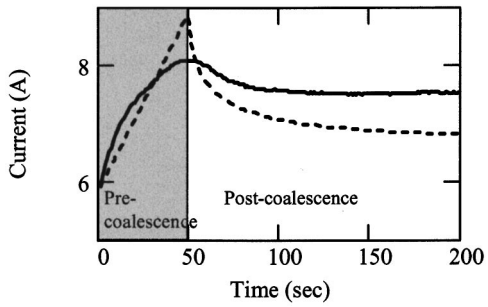


FIG. 4. Overlay of current predicted from geometric model (dashed line) with the current measured during deposition (solid line).

thickness measurement of unpatterned films. After contact, the islands asymptotically approached a planar surface due to the overlapping of the cylinder [Fig. 3(b)] with the total current being equal to

$$I_{\mathrm{tot}} = \left[2r \cdot \arcsin\left(\frac{R}{r}\right) + w_{\mathrm{tr}} \right] \frac{dr}{dt}\frac{dq}{dV} n_{\mathrm{tr}} w_s, \tag{3}$$

where R is the radius of the islands when they coalesced.

Figure 4 is an overlay of the predicted (dashed line) and measured (solid line) currents for a single representative experiment. The only fitting parameter was the position of the maximum current, which was required because the plating rate is larger than the steady-state value during the formation of the diffusion zone. This initial increase in the deposition rate causes the peak to occur earlier than was predicted by a constant deposition rate. In comparing, the modeled and measured current two main features are apparent, rounding of the measured peak current and a higher asymptotic current. Rounding of the peak in the measured current was likely the results of variation in the thickness of the photoresist, variation in the trench width across the samples, and surface roughness, where as the difference between the measured and predicted asymptotic currents is likely the result of the increased surface area due to surface roughness. For the sample used in Fig. 4 the difference between the asymptotic currents was ~12%, where as AFM measurements of the surface indicated the surface area was ~6% larger than a planar film, which is reasonable considering the reduction in the measured surface roughness due to the finite radius of curvature of the AFM tip.

IV. DISCUSSION

Figure 4 demonstrates that the moment of initial coalescence corresponds to the maximum surface area of the film. Therefore, the change in stress-thickness due to the initial coalescence of the islands was determined as the deviation in stress-thickness from the background level to the point of maximum current, as shown in Fig. 2. The remainder of the tensile rise was the result of continual coalescence of the islands during subsequent deposition and planarization. Our data demonstrates that the majority of the tensile stress was actually the result of postcontact coalescence processes. Similar behavior has been observed during chemical vapor deposition of unpatterned diamond films and was attributed to the continual coalescence of faceted grains.[10] Our results

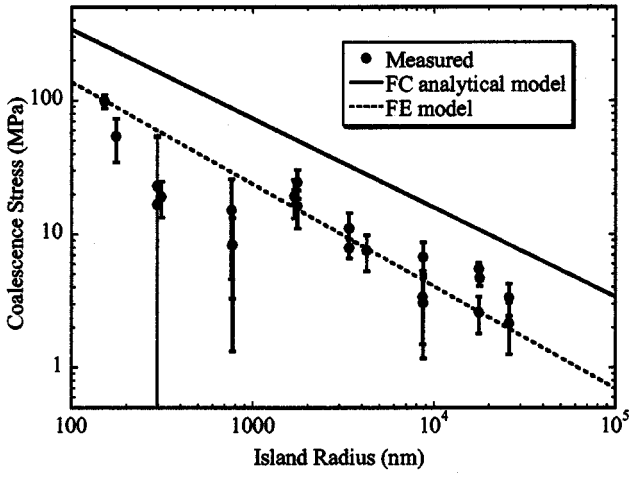


FIG. 5. RMS roughness surface roughness of planar films measured using *ex situ* AFM. The % RMS roughness decreased significantly and stabilized at ~1 micron of thickness.

support the generality of that observation. The detailed understanding of continual coalescence requires significant additional experimentation and analysis, which will be reported in future work.

Figure 5 summarizes our results for the measured stress associated with initial coalescence as a function of the island contact radius. This volume-average stress was obtained by dividing the total change in stress-thickness due to coalescence, as defined above and in Fig. 2, by the mean thickness determined from the plating current.[8,11] On a log–log plot the data exhibits a roughly linear increase in mean stress with decreasing island radius. There is some negative deviation from linear behavior at smaller radii, which was typically associated with depositions where the very poor signal-to-noise ratio created unacceptably large error in the measured stress. Additionally, AFM measurements indicate that the *relative* roughness on surfaces of the half-cylinders was significantly larger for cylinder radii below 1 μm (Fig. 6). Surface roughness decreases the contact area between islands, which reduces the measured stress. Therefore, in the follow-

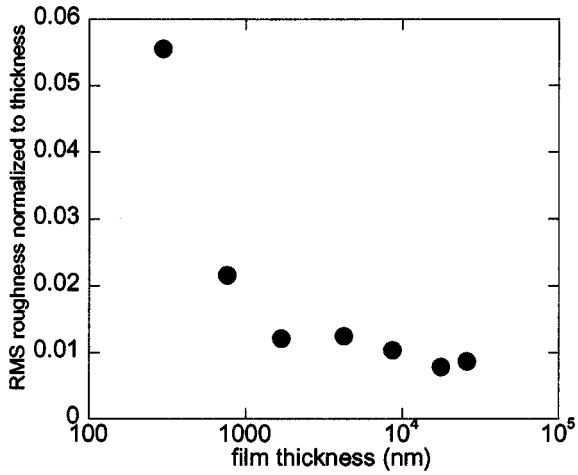


FIG. 6. Plot of measured, FC, and FE calculated island coalescence stress as a function of radius. Both models assume perfect cylinders coalescing and therefore should overestimate the coalescence stress.

Downloaded 18 Apr 2005 to 134.253.26.12. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

ing analysis we shall only consider samples with radii larger than 1 μm for quantitative comparison with the analytical and FE models.

The Freund and Chason model (FC) (Ref. 7) uses Hertzian contact theory to determine the volume average stress (σ_{avg}) created within the islands as a function of the initial coalescence geometry of the island, e.g., blocks, cylinders, or hemispheres. Our experiments used cylinders with their axes parallel to the substrate, whereas FC used cylinders with their axis of symmetry normal to the substrate. This required a slight modification to the FC model as follows.

According to FC the total contact force per unit length (P) for the general case for cylinder contacting under plane strain along their axis of symmetry is given by

$$P = p(\pi R\overline{E})(\zeta/R)^{2/3}, \tag{4}$$

where p the normalized contact force for cylinders and was determined by FC to equal 0.3, $\overline{E}$ is the plane strain modulus, $\zeta \equiv 2\gamma/\pi\overline{E}$, and γ is the difference between the surface energy and half the grain boundary energy, i.e., the energy of the interface between the coalesced islands. The volume average stress is found by dividing the contact force per unit area by the mean film height for the modified geometry, which for a half cylinder is given by $\pi R/4$. Therefore, the volume average stress for a cylinder with its axis parallel to the substrate is

$$\sigma_{avg} = 1.2\overline{E}\left(\frac{2\gamma}{\pi\overline{E}R}\right)^{2/3} = 0.888\overline{E}^{1/3}\left(\frac{\gamma}{R}\right)^{2/3}. \tag{5}$$

This volume average stress, for half cylinders coalescing with their axis of symmetry parallel to the substrate, is twice that of cylinders with their axis perpendicular to the substrate. However, the power-law dependence of the predicted stress with island radius is the same for both geometries and equal to $-2/3$.

As shown in Fig. 1, the true geometry of the coalescing islands used in these experiments was more complex than the idealized analytical model due to the additional block of material in and above the trench. This had two effects on the volume average stress: First, it decreased the number of boundaries formed per unit area causing an overall decrease in the magnitude of the volume average stress. Second, the material above the trench acted as a "shear-lag zone" that transmitted the stress to the substrate, which acted to increase the volume average stress. 2D FE modeling was used to quantify the effect of the material in and above the trench on the functional dependence of the volume average stress on island radius. The details of using FE to model island coalescence have been presented elsewhere.[12] Two geometries were compared using FE modeling: ideal half cylinders [Fig. 1(a)] and the actual island structure (Fig. 7). We found no significant variation in the functional dependence of stress on island radius (slope of -0.76 to -0.74) as a result of the different geometries and only a slight decrease in the magnitude of the stress with the addition of the trench and photoresist. This result supports the use of a simplified geometry (pure cylinders) in the FC model. In calculating the theoretical stress shown in Fig. 5 no fitting parameters were used

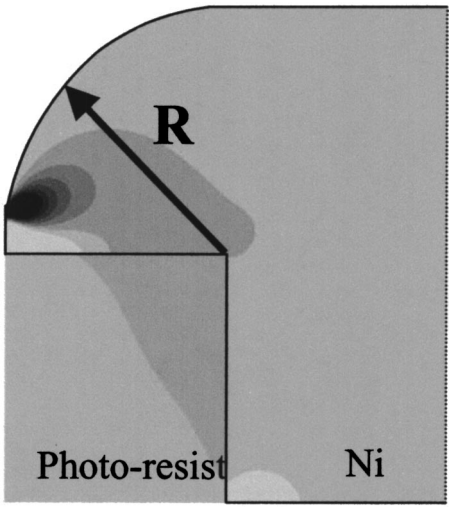


FIG. 7. In-plane stress (σ_{yy}) fields in the full structure as calculated from FE model. Darker color indicates stronger tensile stress.

and both the FE and FC models were calculated assuming $\gamma = 1.85$ J/m^2,[13] and $\overline{E} = 200$ GPa (measured in our films using nanoindentation). It should be noted that both the FE and the analytical models predict the maximum possible stress created during initial coalescence and any error in γ or $\overline{E}$ would result in a shift in the magnitude of both curves and cannot be used to explain the difference between the results. Mechanisms such as incomplete coalescence due to surface roughness or localized yielding will always act to reduce the observed stress. Additionally, the slight difference in the exponents between the FE ($-3/4$) and the Hertzian models ($-2/3$) was likely the result of the lack of an exact solution for the force distribution in contacting cylinders used in the Hertzian model.[7,14]

Figure 5 compares the calculated coalescence stress for the FE (dashed line) and the FC (solid line) models with the experimentally measured initial coalescence stress. In comparing theory vs data, we consider the power-law behavior and the absolute stress magnitudes independently, because the magnitude depends on materials properties whereas the power-law behavior is solely a function of geometry. The observation that the stress magnitudes of the measured data and FE results are similar might appear fortuitous, because there is considerable uncertainty in the surface and grain-boundary energies used in the models and because of potential surface roughness effects. However, since the observed stress was of a similar magnitude to the predicted stress, which is an upper bound, it is reasonable to conclude that any reduction in the initial coalescence area by surface roughness was minimal. More definitively, Fig. 5 shows that the power-law behavior predicted by both the FE and the FC models is within the experimental variation of the measured coalescence stress versus island radius for samples larger than

1 μm (-0.70 ± 0.25). Furthermore, the data lie outside the predicted behavior for both slope and magnitude for the 1D and 3D coalescence geometries in the FC model. *Therefore, the analytical models capture the dominant physical mecha-*

Downloaded 18 Apr 2005 to 134.253.26.12. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

nisms underlying the generation of tensile stress during island coalescence.

V. CONCLUSION

In conclusion, using lithographically defined nucleation and selective area growth via electrodeposition, we have obtained the first meaningful comparison of experimentally measured stresses due to island coalescence with the stresses predicted by theory. These experiments yielded two main results: First, the measured initial coalescence stress had a functional behavior similar to that of the FC and FE models for coalescence stress. From this we conclude that both models correctly account for all of the dominant physical mechanisms active during the initial coalescence of islands. Second, the majority of the stress associated with island coalescence occurred during planarization of the films, not during the initial contact of the islands. This striking result had not been previously recognized in unpatterned metal films due to the inability to differentiate the stress created at the initial contact from that created during planarization.

ACKNOWLEDGMENTS

The authors would like to acknowledge fruitful discussions with Ben Freund, Thomas Buchheit, Brian Sheldon, Eric Chason, and Harley Johnson. This work was supported by the DOE office of Basic Energy Science and by the ARO/MURI in Deep Subwavelength Optical Nanolithography at the University of New Mexico. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under Contract No. DE-AC04-94AL85000.

[1] R. Koch, J. Phys.: Condens. Matter **6**, 9519 (1994).
[2] J. A. Floro, S. J. Hearne, J. A. Hunter, P. Kotula, E. Chason, S. C. Seel, and C. V. Thompson, J. Appl. Phys. **89**, 4886 (2001).
[3] J. A. Floro, E. Chason, and S. R. Lee, Mater. Res. Soc. Symp. Proc. **405**, (1996) (381).
[4] R. Abermann and P. Martinz, Thin Solid Films **115**, 185 (1984).
[5] R. W. Hoffman, Thin Solid Films **34**, 185 (1976).
[6] W. D. Nix and B. M. Clemens, J. Mater. Res. **14**, 3467 (1999).
[7] L. B. Freund and E. Chason, J. Appl. Phys. **89**, 4866 (2001).
[8] Allison Suh, Ning Yu, Ki Myung Lee, Andreas A. Polycarpou, and H. T. Johnson, J. Appl. Phys. **96**, 1 (2004).
[9] L. B. Freund, J. A. Floro, and E. Chason, Appl. Phys. Lett. **74**, 1987 (1999).
[10] B. W. Sheldon, K. H. A. Lau, and A. Rajamani, J. Appl. Phys. **90**, 5097 (2001).
[11] L. B. Freund and S. Suresh, *Thin Film Materials* (Cambridge University Press, Cambridge, 2003), Vol. 1, p. 206.
[12] S. C. Seel, C. V. C. V. Thompson, S. J. Hearne, and J. A. Floro, J. Appl. Phys. **88**, 7079 (2000).
[13] J. P. Hirth and J. Lothe, *Theory of Dislocations*, 2nd ed. (Wiley, New York, 1975), p. 839.
[14] K. L. Johnson, *Contact Mechanics*, 1st ed. (Cambridge University Press, Cambridge, 1985), Chap. 4.2, p. 99.

Downloaded 18 Apr 2005 to 134.253.26.12. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

Demonstration of imaging interferometric microscopy

Christian J. Schwarz, Yuliya Kuznetsova and S. R. J. Brueck
Center for High Technology Materials, University of New Mexico, MSC04 2710, 1313 Goddard SE, Albuquerque, NM, USA 87106

ABSTRACT

We demonstrate imaging interferometric microscopy (IIM) for binary objects in two dimensions. Combining multiple exposures with different off-axis illumination configurations together with interferometric restoration of the zero-order beam during dark-field conditions, IIM provides high lateral resolution at low numerical apertures (*NA*). It retains the large field-of-view, long working distance and large depth-of-field of a low-*NA* imaging system. Also we include a first demonstration of imaging of a phase mask. All these properties are increasingly important for in semiconductor mask metrology. IIM relies on image processing to reconstruct the image; we present the processes necessary to obtain the combined image. Finally we compare the experiment with a simple Fourier optics model.

Keywords: Semiconductor metrology, optical high resolution microscopy, phase sensitive metrology, defect analysis, holography, synthetic aperture radar, Fourier optics, frequency space synthesis, interferometry, off-axis illumination

1. INTRODUCTION

Sub-wavelength resolution in the optical regime is of increasing importance, both for lithographic printing of features in integrated circuits[1] and for microscopy, particularly of biological specimens.[2] The specific technological requirements of mask inspection for microlithography include: a large field-of-view (FOV) in order to cover as large of an area of mask in one image as possible, a long working distance (WD) in order to be able to inspect masks that are protected by a pellicle mounted several millimeters away from the mask surface, and high lateral image resolution, on the order of scaled critical dimensions (CD) or smaller, in order to detect defects. Further a large depth-of-focus (DOF) and phase sensitivity are advantageous for a general inspection tool. Imaging interferometric microscopy (IIM) offers all these properties at the same time. As an optical technology IIM allows for high throughput, standard environmental conditions and is both noncontact and nondestructive, which makes it attractive not only for mask inspection, but also to other fields of application, such as *in vivo* investigations of cells.

IIM takes advantage of the fact that information about an object is contained in diffraction directions that are not collected using conventional illumination, but can be recovered by extreme off-axis illumination (so that the zero-order of the mask is outside of the lens collection cone) along with interferometric reintroduction of a zero-order reference beam on the low-*NA* side of the microscope. IIM developed out of an effort aimed at improved high-resolution printing, namely imaging interferometric lithography (IIL).[3] Both IIL and IIM approach the problem of high resolution imaging from the point of view of linear systems response in which a lens or any optical system is characterized by its linear response function in frequency space,[4] and a lens is characterized as a low pass filter. The question of high-resolution imaging then becomes, how to transmit as many spatial frequencies as possible through the lens using multiple exposures and different illumination conditions, and can be associated with the problem of frequency multiplexing in telecommunication. The difference of IIM and IIL is in the recording technique. IIM, with electronic image capture, allows for image subtraction, IIL, with chemical image capture in the resist latent image, does not. Also the underlying mathematical problem is different due to the difference in applications. In IIL we know what we want to print and the question becomes (how do we print it the best, most efficient way) leading to an optimization problem. In IIM, we ultimately don't know what we are looking at leading us to a difficult inverse reconstruction problem. For the mask inspection application, the mask database is known, greatly simplifying the inverse problem.

In this paper we first review the principle of IIM and explain our experimental arrangement, then present experimental data for two-dimensional objects, introduce a Fourier optics computational model, compare our experiment with the

model, and with data from a conventional high-*NA* microscope image. Finally we discuss the presented material and its future possibilities and implications.

2. PRINCIPLE OF IIM

Any given two-dimensional object can be described by the Fourier transform representing its spatial frequency content. The linear systems analysis of the impact of the optical system on that spectral content provides insight into IIM. The spatial frequency response of an optical system depends on both the illumination and imaging optics. The simplest case is coherent illumination, e.g. plane wave illumination from a specific direction. Scattering by the mask generates a range of spatial frequencies, corresponding to a band-limited Fourier transform of the object. The band limit is the result of the finite range of propagating wavevectors. Scattering from higher spatial frequency components of the object that results in non-propagating evanescent waves cannot be captured by any optical system relying on propagating waves. The highest frequency that can be scattered corresponds to $1/\lambda$ for normal incidence illumination (e.g. to light propagating along the plane of the object). For grazing incidence illumination, the highest spatial frequency scattering corresponds to back reflection, or to a spatial wavelength of $2/\lambda$. This simple example serves to illustrate the higher spatial frequencies accessible with off-axis illumination.

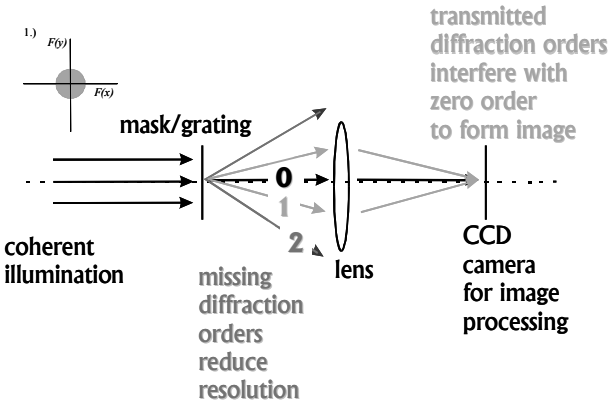


Figure 1: For normal incidence illumination, the imaging optical system captures frequencies up to *NA*/λ.

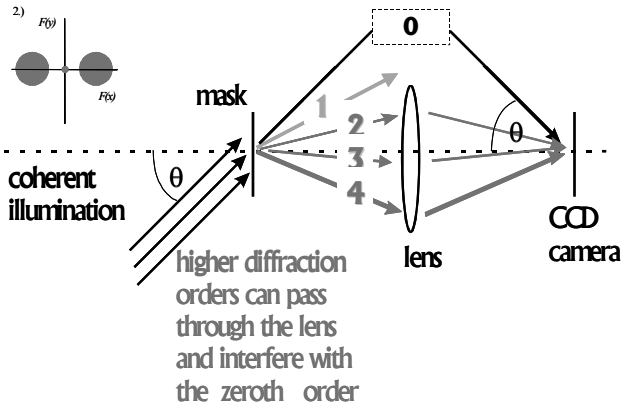


Figure 2: For extreme off-axis illumination, the optical system captures higher frequencies. Reintroduction of the zero-order on the low *NA* side of the optical system restores the image components.

Introducing a lens after the object further restricts the Fourier coefficients that will be transmitted through the lens (see Figs. 1 and 2). Optical systems can be characterized for imaging purposes as linear low-pass filters as mentioned in the introduction.[4] The restricted spectral bandpass leads to a larger depth-of-field (DOF) as compared with a hypothetical lens with an *NA* of unity. In addition, because of the aberrations associated with high-*NA* geometric optics, a lower *NA* lens typically has a longer working-distance (WD) from the lens to the object and images a larger field-of-view (FOV) than a higher-*NA* lens. The strategy of IIM is to use a low-*NA* objective with desirable DOF, WD and FOV and to extend the range of captured spatial frequencies by using extreme off-axis illumination to capture additional parts of the spatial frequency spectrum as illustrated in Fig. 2. The "extreme" is to designate that the off-axis illumination angle is so steep that the zero-order transmission of the mask is not captured by the lens, but is instead interferometrically reintroduced around the lens. The recording medium, a CCD camera or film, is a square law device. The recorded spatial frequencies correspond both to the desired image information (the interference between the zero-order and each of the scattered orders) and to dark-field components (the autocorrelation of the scattered fields) that can be subtracted from the final image. In principle IIM allows us to capture the entire available frequency space up to the propagation limit of $2/\lambda$ noted above.

For the case of a thin one-dimensional grating, illustrated in Figs. 1 and 2, where we have two positive and two negative orders within the optical system limit, we extract the linear image terms by subtracting the dark field image (no zero

order) from the total intensity image. First, for an optical system that captures all of the spectral components in a single image:

$$\begin{aligned} I_{total,lin} &= I_{total} - I_{total,dark} \\ &= \left|E_0 + 2E_1\cos(2\pi x/d) + 2E_2\cos(4\pi x/d)\right|^2 - \left|2E_1\cos(2\pi x/d) + 2E_2\cos(4\pi x/d)\right|^2 \\ &= \left|E_0\right|^2 + 4\left|E_0E_1\right|\cos(2\pi\, x/d) + 4\left|E_0E_1\right|\cos(4\pi\, x/d) \end{aligned} \tag{1}$$

These linear image terms can also be recovered by taking a low-frequency image (Fig. 1) that only captures the zero and first orders in our example and a high frequency image that captures the second order. This gives us:

$$\begin{aligned} I_{low,lin} &= I_{low} - I_{low,dark} \\ &= \left|E_0 + 2E_1\cos(2\pi\, x/d)\right|^2 - \left|2E_1\cos(2\pi\, x/d)\right|^2 \\ &= \left|E_0\right|^2 + 4\left|E_0E_1\right|\cos(2\pi\, x/d) \end{aligned} \tag{2}$$

and

$$\begin{aligned} I_{high.lin} &= I_{high} - I_{high,dark} - I_{ref} \\ &= \left|2E_0 + E_2e^{i4\pi x/d}\right|^2 - \left|E_2\right|^2 - 4\left|E_0\right|^2 \\ &= 4\left|E_0E_2\right|\cos(4\pi x/d) \end{aligned} \tag{3}$$

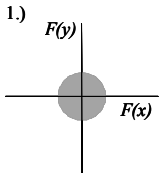
where the reference beam has been increased by a factor of two to compensate for the single-sideband detection implicit in off-axis illumination. Together we have:

$$I_{total,lin} = I_{Ilow,lin} + I_{high,lin} = \left|E_0\right|^2 + 4\left|E_0E_1\right|\cos(2\pi\, x/d) + 4\left|E_0E_1\right|\cos(4\pi\, x/d) \tag{4}$$

The same principle, works for the two dimensional continuous case, *i. e.* for arbitrary images.

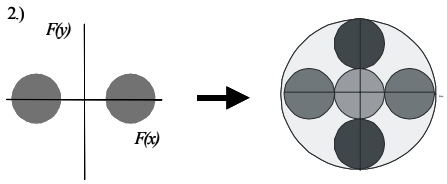
$$I_{low,lin} = I_{low} - I_{low,dark} =$$

$$\left|E_0\right|^2 + 4\left|E_0E_1\right|\cos(2\pi\, x/d)$$



$$I_{high,x,lin} = I_{high,x} - I_{high,x,dark} - I_{high,x,ref} =$$

$$4\left|E_0E_2\right|\cos(4\pi x/d)$$



$$I_{high,y,lin} = I_{high,y} - I_{high,y,dark} - I_{high,y,ref} =$$

$$4\left|E_0E_2\right|\cos(4\pi y/d)$$

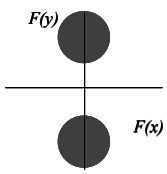


Fig. 3: Illustration of the addition of spectral components from different parts of frequency space, corresponding to different extreme off-axis illumination conditions.

3. THE EXPERIMENTAL SETUP

In order to demonstrate IIM, we built an experimental setup. As an illumination source we use a He-Ne laser at a 633-nm wavelength. Beamsplitters allow for an on-axis illumination and two off-axis illuminations, in the *x*- and *y*- directions, respectively, and two reference beams for the two extreme off-axis images. The angles of the off-axis illumination and reference beams, their polarizations and intensities can be adjusted individually. The shape of the beam is Gaussian and its diameter is about ~2 mm compared to a FOV of 32×24 μm^2 giving essentially uniform illumination across the field. After the mask, the diffracted light from the mask is captured by a 20×, 0.4 *NA* microscope objective. Since the pupil plane of the objective is not accessible, we relay out a conjugate pupil plane, using two 20-cm focal length achromats in a confocal configuration. Also we use an additional 10×, 0.1 *NA* microscope objective in order to magnify the image onto to our CCD sensor, which is a ½" diameter sized 640×480 pixel sensor where each pixel is about 10×10 μm^2.

The reference beam is passed through a second, matched 20×, 0.4 *NA* microscope objective, in order to mode-match the reference beam to the scattered beams. After passing through this objective, the beam is split in two, one for the *x*- and a second for the *y*-direction reference beams. They are reintroduced in the first image plane of the imaging path (after the first microscope objective, see Fig. 3 using two beam splitters and a half-wave-plate in order to match the polarization of the vertical beam. The illumination angle is chosen to be $2\sin^{-1}(NA)$ with $NA = 0.4$, the angles of the reference beam have to be matched to $2\sin^{-1}(NA)/20$, where 20 is the magnification of the first microscope objective. This gives illumination angles of 53° and reference beam angles of 2.3°. In order to control the relative phase between the diffracted light from the mask and the respective reference beam, we installed piezo-actuators on two of the steering mirrors. The images were acquired with an 8-bit video camera and an analog video card and processed using LabView™.

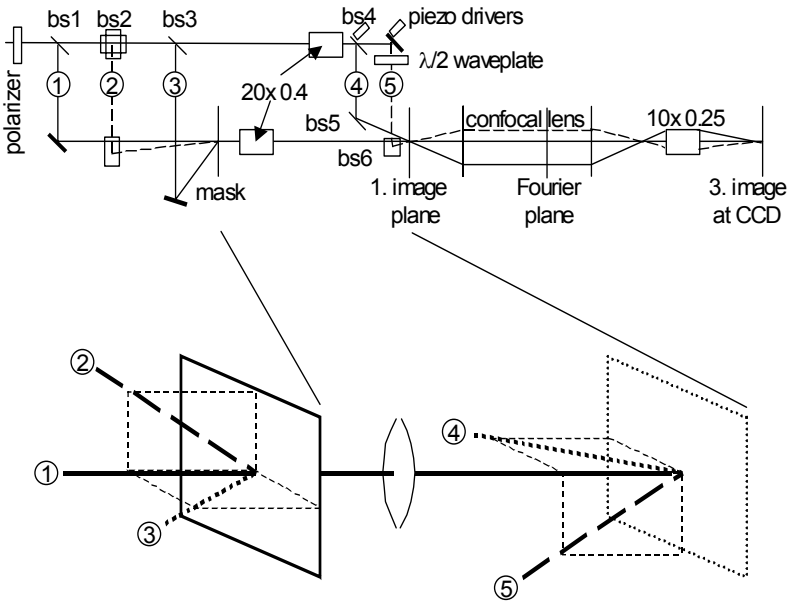


Figure 4: Experimental layout. The lower part of the figure shows the five input beams, 1-on-axis, 2- and 3- extreme off-axis in y- and x-, respectively, and 4- and 5- the two reference beams reintroduced after the first microscope objective.

The illumination angle for the extreme off-axis illumination beams is adjusted by using the back reflection of the mask with the steering mirror mounted on a rotary stage. This allows angle adjustments of about 0.5° on the illumination side. Observation of the interference frequency between the on-axis illumination beam and each of the reference beams is used to set the angles for the reference beams. The focus of the images is mainly controlled by the distance of the mask/object from the first microscope objective and is checked best with an object containing as many spatial frequencies as possible. We typically use one of our Manhattan structures. The phase of the two reference beams has to be calibrated on a known object; here again we use one of our Manhattan structures. The intensities of the different beams are

controlled by variable density filters and are set most precisely by using a global average function over the whole CCD array.

4. EXPERIMENTAL DATA

After all these parameters are adjusted we take a complete set of images as described above, which results in the nine intensity distributions listed in Table 1, where the numbering of the beams is as in Fig. 3.

Table 1. Overview of the different image acquisition configurations

Open	Blocked	Resulting image	Name
1	2,3,4,5	Conventional on-axis	I_{low}
1– 0 order	2,3,4,5	Dark-field of on-axis	$I_{low,dark}$
1: clear	2,3,4,5	Background on-axis	$I_{low,bgd}$
2,5	1,3,4	High-frequency H (horizontal structures)	$I_{high,y}$
2	1,3,4,5	Dark-field of H	$I_{high,y,dark}$
5	1,2,3,4	Reference beam for H	$I_{high,y,ref}$
3,4	1,2,5	High-frequency V (vertical structures)	$I_{high,x}$
3	1,2,4,5	Dark-field of V	$I_{high,x,dark}$
4	1,2,3,5	Reference beam for V	$I_{high,x,ref}$

The major experimental results are shown in Figs. 5 to 7 for a Manhattan geometry (*x*-*y* lines) object with a dense critical dimension of 0.5 μm. Figure 5 shows the sequence of images used to obtain the low frequency image. The left panel shows the as-observed low frequency image, the middle panels show the dark-field (not normalized) and the background obtained by a slightly shifted mask so that there are no objects in the FOV. The background speckle pattern arises from multiple reflections between the optical components.

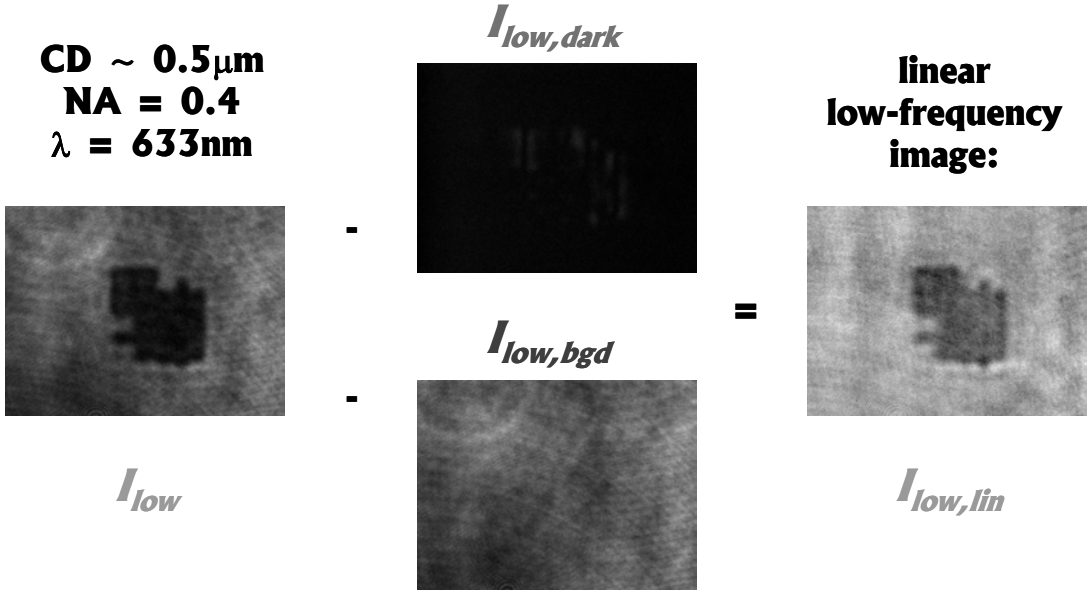


Figure 5: Low frequency image. The left panel is the measured image. The middle panels are the measured dark-field (obtained by blocking the zero-order in the relayed pupil plane) and the background with no object. The interference fringes are the result of speckle from multiple reflections among the optical surfaces. Subtracting the dark-field and background images from the measured image obtain the final low-frequency image.

Figure 6 shows a similar sequence for the *y*-directed extreme off-axis exposure. Finally, figure 7 shows the result of combining all of these images to get a composite image with all of the spatial frequencies included. The reconstructed image was subject to a linear Fourier filter in order to eliminate the spatial frequencies that don't make it through the system and therefore must be noise.

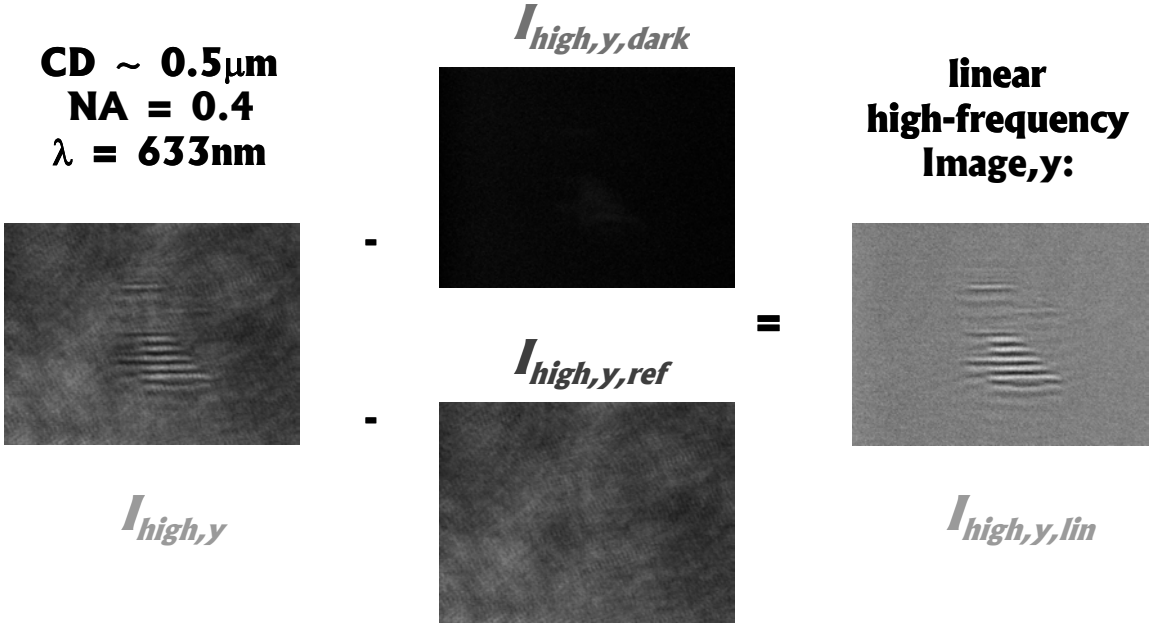


Figure 6: High frequency images for the *y*-directed extreme off-axis case (see table 1).

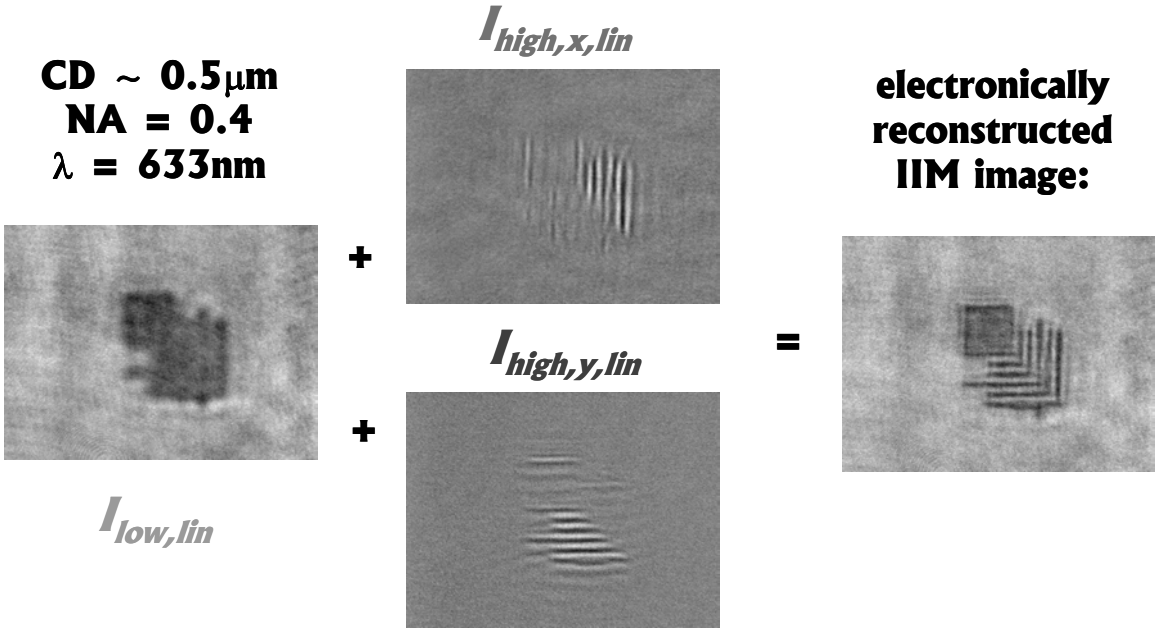


Figure 7: Reconstruction of an image using IIM. The left panel shows the on-axis, low-frequency image. Note that the individual lines are not resolved. The two middle images show the x- and y- directed high frequency images, respectively. Here only the high frequency edges of the large box are observed. Finally, the right panel shows the combined, reconstructed image. The defect on the lowest horizontal line is a true representation of the object as shown below.

Also, we acquired data using a standard white-light, incoherent-illumination microscope with a 100× *NA* 0.9 objective. The image was taken in reflection so the distortion on the bottom line, which is due to corrosion of the mask Cr layer stands out brightly relative to the rest of the image. More discussion of this image along with a comparison of the intensity crosscuts for both IIM and conventional imaging will be presented below after the Fourier optics model is discussed.

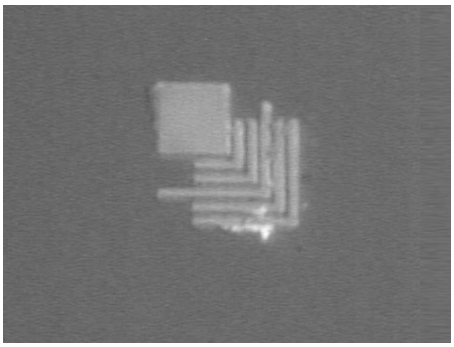


Figure 8: White-light, incoherent-illumination, reflection image of the same object taken with a 100×, *NA* = 0.9 microscope objective.

Also we obtained an image from a phase object, in this case a simple grating. One of the advantages of IIM is that it is sensitive to phase objects for the off-axis exposures since the zero-order reference beam does not experience the phase shifts that the diffracted beam does. As expected, the image is simply a grating, since one diffracted order was collected and interfered with the reference beam to give a simple grating. More work on phase-shift masks has to be undertaken to demonstrate the full advantages of IIM.

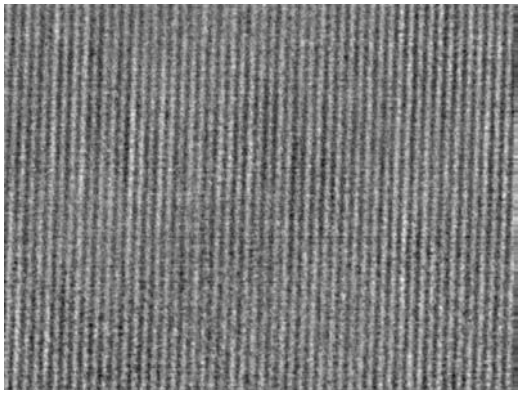


Figure 9: Image of a phase grating.

5. SIMULATION

In order to gain a better understanding of the imaging process of IIM we implemented a simple, scalar Fourier-optical simulation of the optical system. Starting from the Manhattan structure we are imaging, we compute its Fourier transform. Then linear filters are applied specific to each exposure. After application of these filters the absolute square is taken in order to simulate the intensity response of the detector. This model allows us to simulate the low- and high- frequency images and their dark fields for different aperture sizes and off-axis angles in both directions. Fig. 10 shows the results.

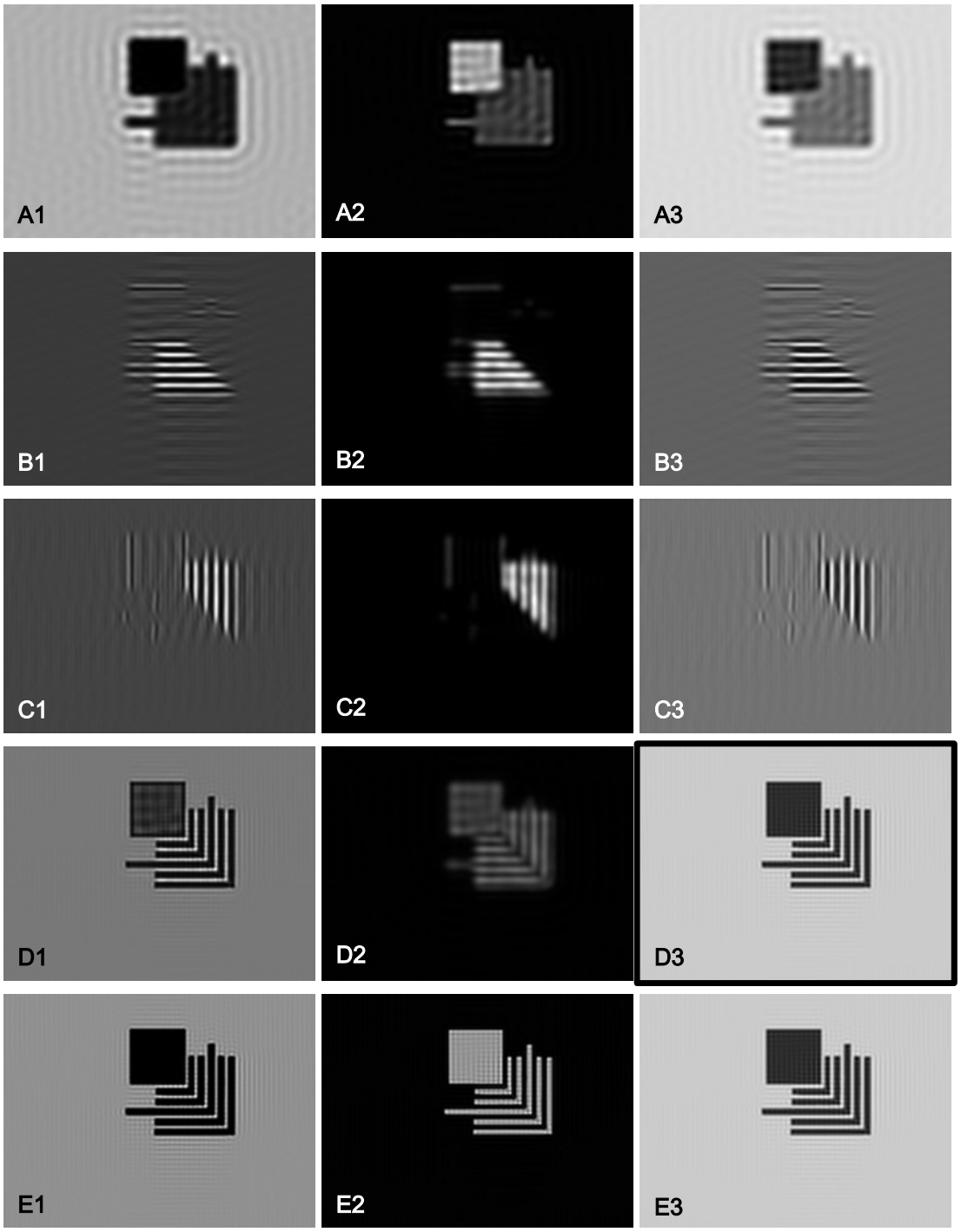


Figure 10: Fourier-optics simulation of IIM. RowA) Low-frequency images, RowB) High-frequency images with *y*-offset illumination, RowC) High-frequency images with *x*-offset illumination, Col.1) Raw image including the autocorrelation terms, Col.2) Dark field images (autocorrelation), Col.3) linear images (after subtraction of dark field), RowD) Linear addition of images A,B,C for each column, respectively, leading to the simulated image of IIM (D3) that should be compared with the experimental result. Finally, rowE) provides images in the respective columns assuming coherent, normal-incidence illumination with a full aperture of *NA*=1.2, covering the same spectral range as the IIM experiment.

6. COMPARISON OF IIM, STANDARD MICROSCOPY AND THE FOURIER MODEL

We can now compare these simulations with the experiment. Figure 11 shows the simulations and the experiment for both IIM (*NA* = 0.4) and conventional, incoherent-illumination imaging (*NA* = 0.9). A more quantitative comparison can be made by comparing intensity cross sections (taken along a horizontal line that cuts across the large square and all of the individual fingers as shown in Fig. 12. Note that the high-frequency contrast (amplitude of the intensity swing across the fingers) in the IIM image is better than in conventional imaging. This is a result of the effective unity transfer function of IIM out to the edges of its frequency coverage compared with the decreasing contrast of incoherent image as the spatial frequency is increased. This effect becomes more pronounced as the frequency content of the image is increased.

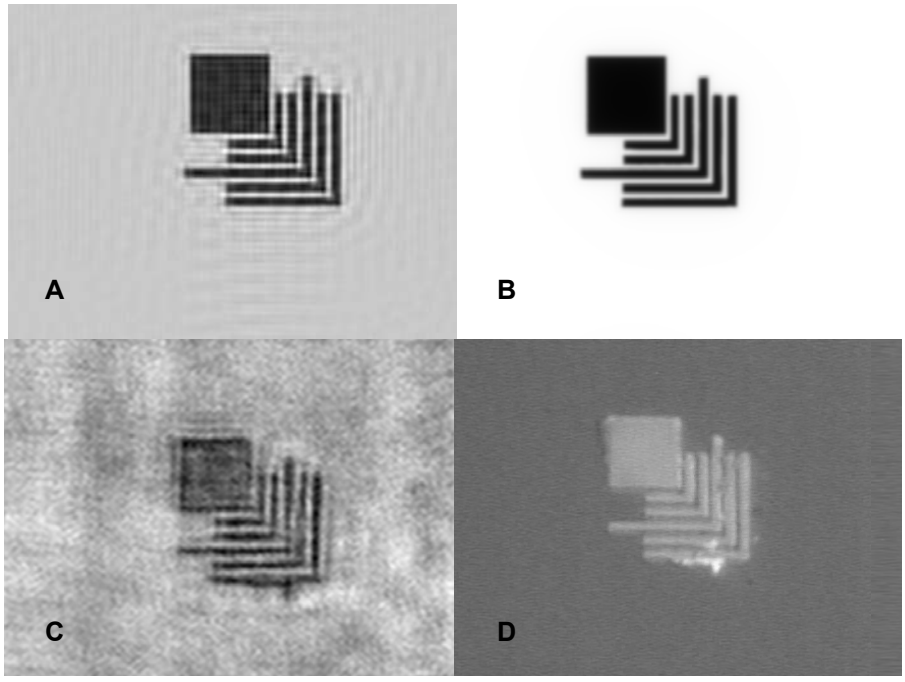


Figure 11: Comparison of the Fourier optics simulation (top) and the measurement (bottom) for IIM (*NA* = 0.4, left) and conventional incoherent-illumination microscopy (*NA* = 0.9, right).

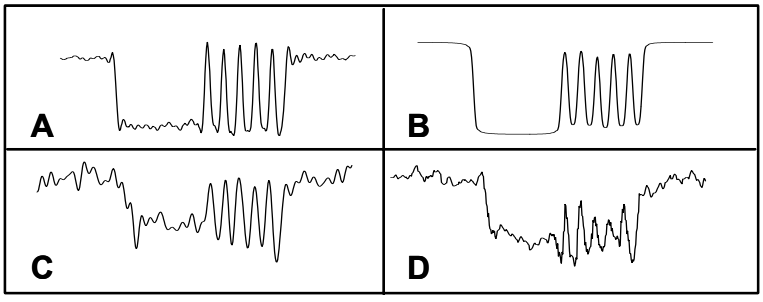


Figure 12: Comparison of the cross-sections of *NA* = 0.4 IIM (A: simulation, C: reconstructed experiment) and the *NA* = 0.9 incoherent-illumination microscope image (B: simulation, D: experiment).

7. DISCUSSION

These initial experiments serve to illustrate the potential of IIM. The mask test structure, at a dense CD of 0.5 μm, is just small enough to demonstrate the possibilities of IIM, the dense CD resolution limit of the present optical arrangement is ~ 260 nm. This is illustrated in Fig. 13, which shows the frequency coverage of the present experiment.

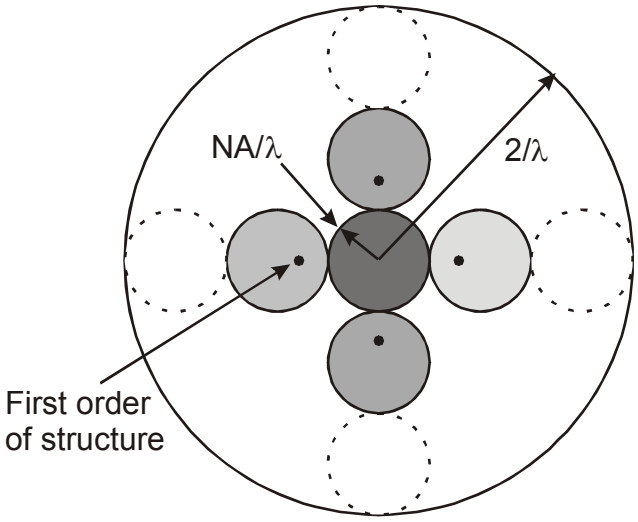


Figure 13: Frequency coverage of present experiment.

The large circle with radius 2/λ is the frequency space limit imposed by optical propagation. The small circle in the center of radius NA/λ with $NA = 0.4$ is the frequency space coverage of the normal incidence coherent exposure. The first order diffraction of the fingers is located at the black dots inside the offset circles that correspond to the two extreme off-axis exposures offset by $2NA/\lambda$. Smaller structures could be imaged with the existing system. The limit is to have the first order spectral frequencies just at the extreme edges of the offset exposures, for this 633-nm wavelength this corresponds to a CD of 264 nm ($CD \sim \lambda/6NA$).

The dotted circles in Fig. 13 indicate that additional frequency space coverage is available. Accessing this frequency coverage requires both additional off-axis tilt, and tilt of the mask relative to the optical axis of the microscope. This will introduce astigmatism and require a much more sophisticated optical system. Just increasing the tilt of the illumination beams to grazing incidence extends the resolution of this system to a CD of ~226 nm [$\sim \lambda/2(1+NA)$] compared with the ultimate resolution of $\sim \lambda/4 \sim 158$ nm. These minimum CD's all scale with wavelength.

8. CONCLUSIONS AND FUTURE DIRECTIONS

We have introduced and demonstrated the principle of imaging interferometric microscopy (IIM). The reconstructed image shows the potential of higher contrast at high spatial frequencies as compared to standard microscopy. Also we achieve the same resolution, with higher contrast at a higher magnification factor and with a larger FOV, a longer WD, and a larger DOF using the IIM approach. We used a 20× $NA = 0.4$ objective with coherent illumination and compared it with a 100× $NA = 0.9$ objective with incoherent, white-light illumination. The FOV, DOF and WD of the lower-*NA* objective were retained in these experiments, but at a resolution comparable to that of the higher *NA* objective. Future work will explore the resolution limits as well as explore a shift to shorter wavelengths. In order to achieve the full potential of IIM, significant improvements in both the optics and the electronics are necessary. Work is underway to provide a more quantitative assessment of IIM.

ACKNOWLEDGEMENTS

Support for this work was provided as part of the ARO/MURI in "Deep Subwavelength Optical Nanolithography." We thank Andrew Frauenglass for important contributions to the optical arrangement and Alexander Neumann for assistance with the Fourier optics simulation.

REFERENCES

[1] See, for example, the International Technology Roadmap for Semiconductors at www.public.itrs.net.
[2] M. G. L. Gustafsson, Current Opinion in Structural Biology **9**, 627 (1999) and references therein.
[3] Xiaolan Chen and S. R. J. Brueck, *Imaging Interferometric Lithography - Approaching the Resolution Limits of Optics,* Opt. Lett. **24**, 124-126 (1999).
[4] J. W. Goodman, Introduction to Fourier Optics, 2nd Ed. (John Wiley and Sons, 1998).

Imaging interferometric microscopy

Christian J. Schwarz, Yuliya Kuznetsova, and S. R. J. Brueck

Center for High Technology Materials and Department of Physics and Astronomy, University of New Mexico, MSC04 2710, 1313 Goddard SE, Albuquerque, New Mexico 87106

Received March 18, 2003

We introduce and demonstrate a new microscopy concept: imaging interferometric microscopy (IIM), which is related to holography, synthetic-aperture imaging, and off-axis–dark-field illumination techniques. IIM is a wavelength-division multiplex approach to image formation that combines multiple images covering different spatial-frequency regions to form a composite image with a resolution much greater than that permitted by the same optical system using conventional techniques. This new type of microscopy involves both off-axis coherent illumination and reinjection of appropriate zero-order reference beams. Images demonstrate high resolution, comparable with that of a high-numerical-aperture (NA) objective, while they retain the long working distance, the large depth of field, and the large field of view of a low-NA objective. A Fourier-optics model of IIM is in good agreement with the experiment. © 2003 Optical Society of America

OCIS codes: 180.0180, 180.3170, 110.0180, 100.2000, 090.2880, 100.6640.

Microscopy is among the oldest applications of optical science[1] and remains one of the most widely used optical technologies. In recent years there have been several efforts to exceed the diffraction limit of optical microscopy. Especially in fluorescence microscopy, approaches have been reported that permit better axial as well as lateral resolution.[2] In particular, these techniques take advantage of a larger set of angles about the sample to collect the light[3] and explore different illumination schemes[4] as well as nonlinear processes[5] such as multiphoton absorption.[6] These techniques are enhanced further by growing capabilities in computational image processing.[7]

Improvements in microscopy will have significant effects in many fields. Subwavelength resolution imaging has been of intense interest, particularly for lithographic application in semiconductor manufacturing.[8] The concept of imaging interferometric microscopy (IIM) grew out of imaging interferometric lithography, a recently introduced resolution-enhancement technique.[9]

Many applications present conflicting requirements for high-resolution imaging over large fields at a large working distance (WD) from the specimen. For traditional optical designs, high-numerical-aperture (NA) and hence high resolution are associated with a short working distance, a small field of view (FOV), and a small depth of focus (DOF). Mask inspection for the semiconductor industry provides one example of these requirements. The dimensions of mask features are decreasing apace with Moore's law and are currently smaller than 360 nm (and even smaller for subresolution assist features), assuming a 4× reduction lithographic system. High-speed inspection demands an imaging solution in place of the traditional serial, raster-scan approach. A lithographic mask has a protective pellicle mounted several millimeters above it; the imaging optics must stand off from the mask, requiring long-WD optics.[8]

IIM approaches this problem by using a low-NA objective that inherently provides a long WD, a large DOF, and a large FOV. To overcome the resolution limits associated with the lower optical frequency cutoff of this optical system, multiple coherent images with different off-axis illumination directions are recorded. For illumination directions beyond the NA of the imaging lens, a zero-order illumination beam is reinjected into the optical system on the low-NA side of the lens. Effectively, the off-axis illumination downshifts the high-spatial-frequency content of the image such that the image can pass through the lens, and interference with the reinjected reference beam upshifts the image content back to the proper spatial frequencies. The result is a spatial wavelength-division multiplex approach that captures all the available spatial-frequency content. Signal processing is then used to reconstruct the complete image. IIM is related to synthetic-aperture imaging, in which one achieves different channels by sending radar pulses from an airplane as it flies over the target. Each radar subimage, corresponding to a dark-field image with a relatively low NA appropriate to the antenna size, is stored, and the full image is reconstructed by combination of these subimages.[10,11]

In IIM, spatially coherent light is used to illuminate the object. For normal incidence illumination, only the beams diffracted from the mask that correspond to spatial frequencies of $\leq \mathrm{NA}/\lambda$, including the zero-order transmitted beam, are transmitted through the optical system. At the image plane the interference between the diffracted beams and the zero-order beam reconstructs a band-limited image. If the illumination direction is shifted off axis, the lens accepts higher spatial frequencies in the direction opposite the offset angle. For illumination at an angle corresponding to 2NA, only the diffracted beams that correspond to a circle of radius NA/λ centered on $2\mathrm{NA}/\lambda$ are accepted by the optical system (Fig. 1). If these diffracted beams alone are incident upon the detector, the result is a dark-field image with frequencies up to $2\mathrm{NA}/\lambda$. The original frequencies can be restored by the introduction of a reference beam at the appropriate angle on the low-NA side of the lens, much as in holography, that interferes with the image

0146-9592/03/161424-03$15.00/0 © 2003 Optical Society of America

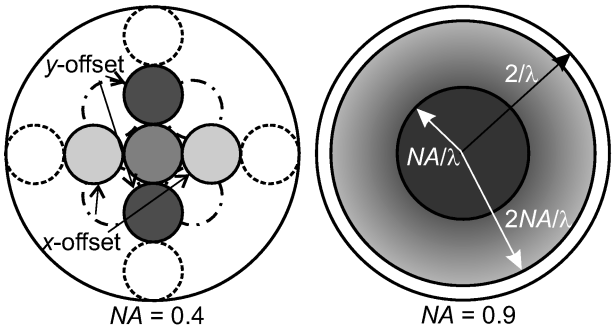


Fig. 1. Frequency space coverage of three-exposure IIM with a NA of 0.4 and of conventional incoherent illumination imaging with a NA of 0.9. The experiment is illustrated by the figure at the left with one offset illumination in each of the x and y directions. Additional exposures can enhance the resolution in the $x-y$ direction, as indicated by the short-dashed circles, or oriented in other directions (e.g., 45° lines) as indicated by the dashed-dotted circles.

diffraction beams to reset the spatial frequencies. The square law (power) detection process results in recording the real part of the interference between the reference beam and the image beams, i.e., to the two x-offset circles shown in Fig. 1. As in single-sideband communications, the full spectrum is re-created by the nonlinearity of the detection process even though only one sideband is transmitted through the system. This process is repeated twice, once offset in the x direction once in the y direction. Since the images are stored electronically, the dark-field images can be separately measured and subtracted from the off-axis images. Combining these off-axis images along with the on-axis image reconstructs a high-resolution image with spatial frequencies extending to $\sim 3\mathrm{NA}/\lambda$.[12] As is clear from Fig. 1, additional parts of frequency space can be covered with additional exposures. Pupil plane filters can be used to ensure that every region of frequency space is weighted uniformly.

For illumination we use a 633-nm wavelength He–Ne laser source. The object is a binary chrome-on-glass mask with dense 0.5-μm critical dimension features. In this Letter we demonstrate the imaging of a Manhattan-geometry structure composed of rectangles oriented along the orthogonal x and y axes. The mask pattern consists of a large square (5 μm × 5 μm) and 5 L-shaped nested lines (0.5 μm wide on 1-μm centers) with various lengths.

For imaging we use a 20×, 0.4-NA objective. Following the image plane of this lens, a pair of confocal achromatic lenses allows access to the image Fourier plane. The relayed image is further magnified with a 10×, 0.25-NA objective and detected with a CCD camera. The overall system has a magnification of 200×, and the 640 × 480 pixels on the CCD array correspond to a 32 μm × 24 μm FOV. The same object on the mask can be illuminated from three directions: normal incidence, off axis from the x direction, and off axis from the y direction. The off-axis illumination angles are $\sin^{-1}(2\mathrm{NA})$, or 53.1° from the optical axis. In this case the off-axis zero-order transmitted beams are not collected by the objective lens, and reference beams are reintroduced at the image plane at an angle matched to $\sin^{-1}(2\mathrm{NA}/20) = 2.3°$ from the optical axis (see Fig. 2). Piezo drivers are used to adjust the phases of the reference beams. A collimated beam is used for illumination, with the result that the zero order is divergent in the first image plane in this nontelecentric optical system. Matched 20×, 0.4-NA objectives are used to mode match the reference beams.

The schematic experimental setup for the x-directed, high-frequency image is shown in Fig. 2. A second pair of beams, not shown here, permits illumination of the y-directed high-frequency image. By sequentially blocking two of the three illumination beams and the respective reference beams, we record different parts of the spatial-frequency space of our image, as shown in Fig. 1. Off-axis illumination without any reference beam gives us the dark-field images. Moving the mask to a clear spot away from any chrome features provides the background for the low-frequency image, and transmitting a single reference beam without any illumination serves as the reference for the high-frequency images. Using an 8-bit camera and analog-to-digital image-acquisition board, we digitally recorded the images.

Figures 3 and 4 show the experimental results. Figure 3A shows the on-axis image, which contains only the low-frequency mask information. The square and the nested letters L are merged into a single unresolved structure. Only the extended lines are visible. The background of the structure is clear on the mask but appears nonuniform in the image.

The high-frequency components are imaged separately. One image captures the high-frequency components of the horizontal structures (y direction, Fig. 3B); another, the high-frequency components of the vertical structures (x direction, Fig. 3C). The background of the image is rather homogeneous but not completely uniform; the spaces between the lines show higher intensity than the background. The line structures and the high-frequency edges of the square are seen in their respective images.

Electronic addition of the low- and high-frequency images gives the reconstructed image (Fig. 3D). A digital-image-processing low-pass Fourier filter was applied after the reconstruction simply to eliminate any electronically induced noise at spatial frequencies

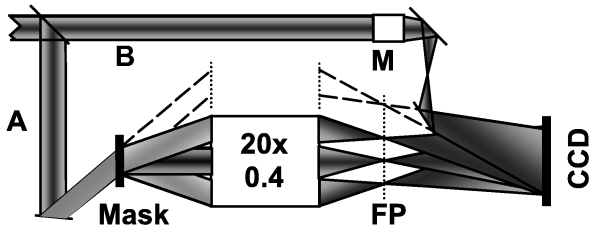


Fig. 2. Schematic of the high-frequency imaging setup. The incoming beam is split into illumination beam A and reference beam B; beam B is reintroduced to the image plane at the CCD, replacing the zero order that did not pass through the optical system (dashed lines). The three beams impinging upon the optical system represent the range of image spatial frequencies collected. Beam B is mode matched with a second microscope objective (M). FP is the relayed Fourier plane.

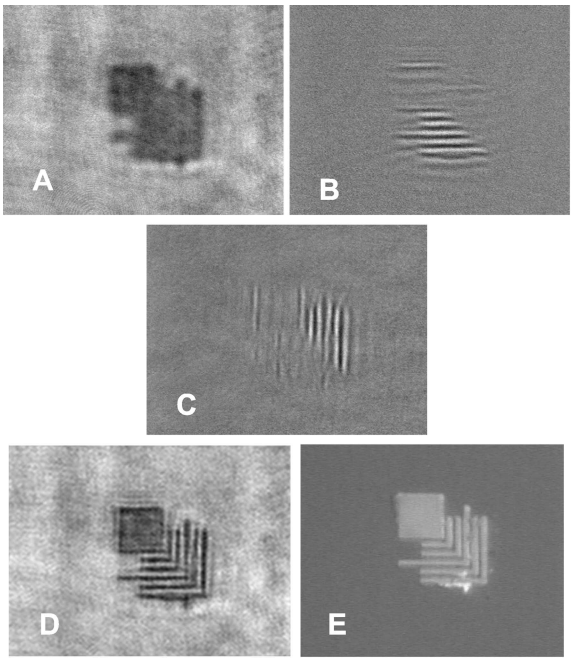


Fig. 3. A, On-axis image after dark-field and background subtraction and renormalization. Only the low-frequency components of the mask are imaged. B, High-frequency image of horizontal structures after dark-field and reference image subtraction and renormalization. C, High-frequency image of vertical structures after dark-field and reference image subtraction and renormalization. D, Reconstructed image. E, Incoherent illumination; white-light image from a 100×, 0.9-NA lens.

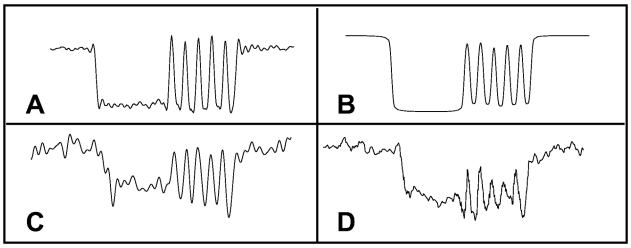


Fig. 4. Comparison of the cross sections of three-exposure, 0.4-NA IIM (A, simulation; C, reconstructed experiment) and the 0.9-NA incoherent illumination aerial images (B, simulation; D, experiment).

outside the optical system capabilities. The individual lines are now separated from the square and have their appropriate lengths. The edges of the lines are clearly visible. Irregularities are evident at the left side of the square as well at the lowest horizontal line. Finally, Fig. 3E shows an image of the same structure taken in reflection with a standard microscope using a 100×, 0.9-NA objective with incoherent, white-light illumination ($\lambda \sim 500$ nm). The edges of the structure are not as sharp as in the reconstructed image. The FOV was adjusted to compensate for the different magnifications employed in the two images. Some corrosion of the mask metallization is evident on the lowest horizontal finger. This defect was also resolved in the IIM image.

To compare the pictures more quantitatively, we show the cross sections of (Fig. 4A) a Fourier-optics simulation of IIM (NA of 0.4),[13] (Fig. 4B) an incoherently illuminated image (NA of 0.9) and the experimental data, (Fig. 4C) the reconstructed IIM image, and (Fig. 4D) the conventional microscope image. The larger contrast of IIM at high spatial frequencies than for the standard image can be seen. This difference is related to the low modulation transfer function of incoherent imaging at high spatial frequencies, whereas the effective modulation transfer function of IIM is unity.

We have introduced and demonstrated the principle of imaging interferometric microscopy. The reconstructed image shows the potential of higher contrast than that of standard microscopy at high spatial frequencies. Also, we achieved the same resolution, with higher contrast at a higher magnification factor and with a larger FOV, a longer WD, and a larger DOF, by using the IIM approach. We used a 20×, 0.4-NA objective with coherent illumination and compared it to a 100×, 0.9-NA objective with incoherent, white-light illumination. The FOV, the DOF, and the WD of the lower-NA objective were retained in these experiments, because addition of the reference beam does not affect any of these aspects of the low-NA image, but at resolution comparable to that of the higher-NA objective. The mask test structure, at a critical dimension of 0.5 μm, is just small enough to demonstrate the possibilities of IIM. The dense critical dimension resolution limit of the present optical arrangement is ~0.26 μm. We intend to explore these resolution limits.

The expert assistance of Andrew Frauenglass in these experiments is gratefully acknowledged. Support for this work was provided by the U.S. Army Research Office/Multidisciplinary University Research Initiative. S. R. J. Brueck's e-mail address is brueck@chtm.unm.edu.

References

1. E. Abbe, Ark. Mikrosk. Anat. **9,** 413 (1873).
2. M. G. L. Gustafsson, Curr. Opin. Struct. Biol. **9,** 627 (1999), and references therein.
3. S. W. Hell and E. H. K. Stelzer, Opt. Commun. **93,** 277 (1992).
4. A. V. Failla, A. Cavallo, and C. Cremer, Appl. Opt. **41,** 6651 (2002).
5. S. W. Hell and J. Wichmann, Opt. Lett. **19,** 780 (1994).
6. W. Denk, J. H. Strickler, and W. W. Webb, Science **248,** 73 (1990).
7. E. R. Dowski and W. T. Cathey, Appl. Opt. **34,** 1859 (1995).
8. International Technology Roadmap for Semiconductors at www.public.itrs.net.
9. X. Chen and S. R. J. Brueck, Opt. Lett. **24,** 124 (1999).
10. C. J. Oliver, J. Phys. D **22,** 871 (1989).
11. F. Le Clerc, M. Gross, and L. Collot, Opt. Lett. **26,** 1550 (2001).
12. C. J. Schwarz, Y. Kuznetsova, and S. R. J. Brueck, Proc. SPIE **4689,** 802 (2002).
13. S. R. J. Brueck and X. Chen, J. Vac. Sci. Technol. B **17,** 908 (1999).

SELF-CALIBRATION FOR 2-DIMENSIONAL PRECISION STAGES BASED ON CIRCLE CLOSURE PRINCIPLE

Xiaoming Lu & Hy D. Tran *
Department of Mechanical Engineering
University of New Mexico
Albuquerque, NM 87131

Abstract

Self-calibration is one of the main approaches to obtain greater accuracy in precision engineering. By measuring a rigid artifact placed at different locations and orientations on a stage, it is possible to separate the stage error map from artifact errors by comparing the different views. However, it is impossible in practice to realize an exact 90° rotation about an exact pivot point, and achieve an exactly known translation with absolutely no rotation. The key issue is to decouple stage error when there are artifact errors, alignment errors, and measurement noise. Previous approaches are to isolate the alignment error first by using least squares, which is effective when the measurement noise has a normal distribution. Then, polynomial fitting, or Fourier series, or other orthogonal function decompositions are used to decouple the stage error from three measurement views, view 0, view 1(90^{0} rotation) and view 2 (one interval translation). It is difficult to realize these three measurement views in practice.

We present a testbed physical realization of an X-Y stage for nanolithography. This stage is an air-bearing dc motor driven capstan drive system being retrofitted with a NanoGrid XY-grid-based encoder for feedback. The NanoGrid greatly reduces possible effects from air turbulence (as might be seen with laser interferometers) and Abbe offsets from using separate X and Y scales. However, the metrology is based on an artifact: The accuracy of the grid itself. We will present a circle-closure based self-calibration method for grid and stage calibration.

The principle of circle closure (the sum of the angles around any point in a plane equals 360) is for self-calibration of the rotation angle of the rigid artifact (e.g. XY encoder grid). When the mark position of rigid plate is measured in turn with about 90° rotation, difference between each rotation and an unknown reference angle x can be measured using a differential measurement (which can achieve better resolution). Circle closure provides a constraint that the sum of all the rotations must be 360°. This provides sufficient information for the solution for all unknown rotation angles. We only need to decouple the computation of the pivot point position in the self-calibration algorithm. The mathematical model is simpler, and may be easier to implement.

Redundant measurements may be used to improve calibration accuracy. Different combinations from the 4 measurement views (0, 90^{0}, 180^{0}, 270^{0}) can be used to calibrate the stage errors. The final calibration is calculated by using least square methods. The mathematical model and algorithm with and without additive measurement noise are simulated using MATLAB.

Keywords: self-calibration, nanolithography, calibration algorithm, two-dimensional stage

Introduction

Two-dimensional XY stages are used in many precision engineering applications. The specific application that we are interested in is nanolithography, where it is desirable to maintain nm to sub-nm accuracies over a 2-dimensional workspace spanning a wafer (e.g. 300 mm). There is generally a deviation between the measured stage position and its true position; this is called the stage position measurement error.

The stage position measurement error is the sum of random measurement noise and systematic measurement error, which is a function of the stage position. Calibration is the procedure used to

* Lu: luxm@unm.edu Tran: tran@me.unm.edu This work is supported by DOD/ARO DAAD-19-99-1-0196

determine the stage error map, to map the measured stage position to its position in a Cartesian reference frame. In conventional calibration method, two steps are needed: First, measure a standard artifact; i.e, an artifact plate that has marks with known positions. The difference between the measured position and the known position of the marks is the stage error at that mark location. Second, set up a mapping function (a stage calibration function) between the measured coordinates and the actual coordinates; e.g., using a piece-wise linear function or a polynomial. The stage calibration function is an approximation to the stage error map. The accuracy depends on the traceability of the artifact. Precision stages are tested against artifacts which have been certified by a calibration service whose "grand master" are traceable to national measurement or standard institutes which have realized and transferred the base units.

In our nanlithography application, there is no standard artifact available with the desirable nm accuracy. In fact, we are unable to fabricate an artifact with mark positions known better than the metrology system. Here we run into the chicken-or-egg problem: How do you construct an accurate measurement standard without an accurate standard to calibrate the machine that would make it? Therefore we need to do self-calibration.

Previous work on self-calibration of x-y stages

Self-calibration has a long history, and the advent of e-beam mask pattern generators has increased interest in its use. Raugh (1984) developed a mathematical theory for 2-dimensional self-calibration; however, the algorithm presented is not numerically robust, and is also sensitive to random measurement noise. Ye, Pease et al. (1997) developed a discrete Fourier transform-based algorithm, based on Raugh's theory, which is numerically robust. This approach extracts the stage measurement error from the artifact mark position measurements by comparing three different views of a measurement artifact plate. These three views are an initial set of measurements, a set of measurements with the artifact plate rotated 90 degrees with respect to the stage, and a set of measurements with the artifact plate translated by one grid interval with respect to the stage. The problem now is, how do you achieve an exact 90 degree rotation, and how do you translate the artifact plate one grid interval, without adding rotational motion. Finally, the discrete Fourier-transform based algorithm is complicated.

Testbed physical realization of x-y stage for nanolithography

Our interest is to develop a stage for imaging interferometric lithography (IIL, Brueck, 1998). This technique can help optical lithography approach the fundamental linear systems limit of optics of $\lambda/2$. Any lithographic technique will require more accurate and precise stage positioning (and dimensional stability) to enable registration across multiple levels. Hence, the motivation to investigate self-calibration. Current metrology for precision and ultraprecision stages is based on heterodyne laser interferometers; with a HeNe laser at 632.8 nm wavelength, interpolation of 2^{10} to 2^{11} ($\lambda/1024$ to $\lambda/2048$) is available for two-pass interferometers. This corresponds to resolutions of 0.6 to 0.3 nm. Unfortunately, the refractive index of air varies with CO_2 and water-vapor content, and air turbulence. This is the primary limitation on the accuracy with operation in air—while the resolution of the interferometers is nm, the actual accuracy may only be sub-μm. Other metrology systems have been proposed to reduce the errors due to air, such as holographic encoders and two-color interferometers; neither technology has yet been proven accurate. For metrology, we are using the OPTRA NanoGrid XY encoder system, which offers nm-level resolution, but with accuracy at the submicron level. The objective of the self-calibration research is to develop algorithms to improve the accuracy of the encoder to its resolution—from submicron to nanometer accuracy, and to implement these in a lithographic stage. We have obtained, courtesy of Sematech, an air-bearing, capstan drive interferometric lithography stepper, with HP laser interferometers for feedback (shown in Figure 1). This stepper will serve as a testbed for the stage development, which will primarily investigate self-calibration and thermal compensation methods.



Figure 1. Testbed air-bearing stage for self-calibration research

This stage is configured as a stacked-slide X-Y stage, although the moving stage itself is supported by air bearings on the granite surface plate. The wafer chuck is mounted on a piezo-driven flexure stage. The existing laser interferometer will be retained for calibration purposes, although its basic resolution is 5 nm (e.g. coarser than the resolution of the NanoGrid metrology).

The X-Y encoder system resolution is 0.3 nm; it is insensitive to variations in the index of refraction of air—but its basic accuracy is on the order of 1 μm, since its accuracy is based on the absolute accuracy of the encoder artifact plate, whereas the laser interferometer accuracy is fundamentally dependent on the wavelength of the laser. We will mount the NanoGrid system on the stage, with the encoder plate parallel to, and in the same plane as the wafer chuck. The encoder plate will move with the wafer chuck, while the readheads are stationary. This will minimize Abbe offsets. At the same time, since there is a large planar area available around the wafer chuck, multiple encoder plates and readheads may be installed in order to obtain redundant measurements.

Stage error function and artifact plate error function

The stage error function S(x,y) is the systematic measurement error at (x,y), where (x,y) is the true location of the sample site in a Cartesian grid. Assume the field to be calibrated is L×L. The origin of the x-y axis is chosen at the center of this area. So,

$$S(x,y) = S_x(x,y)e_x + S_y(x,y)e_y$$

where ex and ey are the unit vectors of the stage axis. Sx(x,y) and Sy(x,y) are functions on the continous 2-D field L×L. Let the sample sites be in a N×N square array covering the field L×L
(For simplicity, let N be odd). The sample site locations in the Cartesian grid are:

$$x_m = m\Delta$$

$$y_n = n\Delta$$

where m= – (N–1)/2,– (N–2)/2, … ,0 ,…, (N–2)/2 ,(N–1)/2, n= –(N–1)/2,–(N-2)/2, … ,0 ,…, (N–2)/2, (N–1)/2 and $\Delta = (L/N)$, which is called the sample site interval. So the discrete stage error function can be expressed as

$$S(x_m, y_n) = S_x(x_m, y_n)e_x + S_y(x_m, y_n)e_y$$

The artifact plate has a square (N×N) mark array of the same size as the stage sample site array. The origin of the plate x-y axis is located at the center on the mark array. The nominal mark locations in the plate coordinate system are the same as the sample site locations in the stage coordinate system. The plate is rigid; i.e, the relative positions between the marks on the plate do not change when we rotate it on the stage. However it is not perfect, which means that each actual mark at (x_m,y_n) deviates from its nominal location by $A(x_m,y_n)$.

$$A(x_m, y_n) = A_x(x_m, y_n)e_x + A_y(x_m, y_n)e_y$$

It should be noted that every mark on the artifact plate has an identification number (m,n). The mark's identification number does not change during the plate's motions on the stage.

Decoupling the alignment error based on angle calibration using circle closure

Circle closure is used to measure the rotation angle. The XY grid is placed on a rotation table which is mounted on the stage; a schematic is shown in Figure 2. There are two sensor heads above the XY grid which are used to get translation information in each measurement view. We can select two points on the four sides of xy grid plate to determine the direction of each side. The direction change of each side can be measured by an autocollimator or other high resolution sensor. According to circle closure, the sum of 4 inner angles is equal to 360 degrees. The principle of circle closure is a natural conservation law for plane angles in Euclidean geometry. Denoting the deviations from the nominal value by $\Delta\theta_i$ (i=0,1,2,3) and the actual rotation angle θ_i .We have

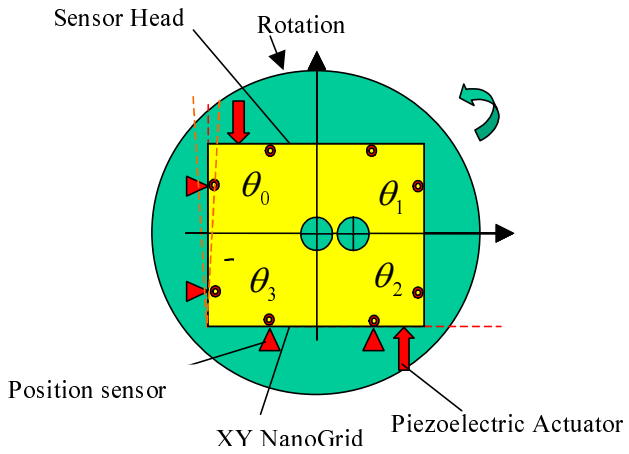


Figure 2. Decoupling alignment errors based on angle calibration via circle closure.

$$\theta_i = x + \alpha_i$$

$$\theta_i = 90^0 + \Delta\theta_i$$

$$\Delta\theta_i = x - 90^0 + \alpha_i$$

$$\sum_{i=0}^{3}\Delta\theta_i = 4(x - 90^0) + \sum_{i=0}^{3}\alpha_i$$

With the constraint of the circle closure $\sum_{i=0}^{3}\Delta\theta_i = 0$, we have $x = 90^0 + -\frac{1}{4}\sum_{i=0}^{3}\alpha_i$

With x, we can calculate the actual rotation angle θ_i with

$$\theta_i = 90^0 - \frac{1}{4}\sum_{i=0}^{3}\alpha_i + \alpha_i$$

We only need to decouple the computation of the pivot point position in the self-calibration algorithm. In this way, only two 90^0-rotation views are needed to exactly calibrate the stage errors at those sites sampled by the mark array. The mathematical model is simpler, and may be easier to implement.

Summary

The mathematical model and algorithm with and without additive measurement noise are simulated using MATLAB; the results of the simulation are shown in the poster. The algorithm gives an exact self-calibration on the discrete sample sites. When there is random measurement noise, different combinations from different rotations and two sensor head (4 rotation measurement views and 4 translation measurement views) are used to calibrate the stage errors. The final calibration is calculated by using least square methods.

References

Brueck, S. R. J., "Imaging Interferometric Lithography," *Microlithography World*, Winter 1998.

Evans, C. J., Hocken, R. J., and Estler, W. T., "Self-Calibration: reversal, redundancy, error separation, and 'absolute testing,' " *CIRP Annals*, Vol45(2), 1996.

Raugh, M.R. "Absolute two-dimensional sub-micron metrology for electron beam lithography" *Prec Eng* vol. 7(1), p. 3-13, 1985.

Ye, J., Takac, M., Berglund, C. N., Owen, G., and Pease, R. F., "An exact algorithm for self-calibration of two-dimensional precision metrology stages," *Prec Eng* vol. 20(1), p. 16-32, 1997.

Carlson, D., Dorval, R. K. , Gargas, J. A., Hercher, M., Mansur, D., and Tran, H. D., "A Versatile XY Stage with a Flexural Six-Degree-of-Freedom Fine Positioner," *Proc. 10th ASPE Annual Meeting*, p. 203-206, October 1995.

MULTIPOINT TEMPERATURE CONTROL USING THERMOELECTRIC MODULES

Jian Ouyang[*], Phuong-Thao Ton-Nu[+], and Hy. D. Tran[* §]
* Dept. of Mechanical Engineering
University of New Mexico
Albuquerque, NM 87131
+ Intel Corporation
Rio Rancho, NM 87124

Abstract

Temperature control is a critical issue in precision engineering, especially for metrology structures and metrology frames. While oil shower and air showers have been shown to achieve millikelvin temperature stability, some applications do not permit oil shower or air shower systems (e.g. clean room environments). We demonstrate that by adding active temperature control to a structure, using thermoelectric devices for cooling or heating, we can maintain a structure at a more uniform temperature despite environmental temperature fluctuations. While it is not possible to maintain completely uniform temperature, it is possible to maintain temperature at selected points in the structure, where the sensors and actuators are located. Our objective is to demonstrate millikelvin temperature stability despite non-uniform external thermal disturbances. Some of the control issues that arise are cross coupling between different sensors and actuators, and effect of bias and drift of sensors (especially at millikelvin resolution). The analysis is based on a state variable finite difference thermal model. While state-space based control design may better handle cross coupling in the dynamics of the thermal system, simpler PID type controllers are easier to implement and test. The major advantages with active temperature control are that by separating the functions of temperature control from structural stability, one can choose materials for metrology structures based on structural criteria, independent of thermal properties. We will show initial experiments using a 4×4 array of thermoelectric coolers to control the temperature of an aluminum plate despite environmental disturbances. The results and control methods can be extended to large structures, such as metrology frames in lithography steppers. The advantages of this approach are not only better temperature control, but also superior dimensional control using inexpensive materials, and also more rapid thermal equilibration during equipment startup and shutdown.

Introduction

Thermally induced errors in precision mechanical structures are a major part of the error budget in precision mechanical systems, especially with lithography applications. With the increasing demand for improved minimum critical dimension (CD) in semi-conductor fabrication, the desire to maintain a dimensional stability of 10% of the CD (currently 0.18 μm) becomes increasingly challenging. In spite of the manufacturers' effort, there are always some residual thermal displacements that are related to coefficient of thermal expansion (CTE) of the materials of the structure, which is subject to the fluctuations of the surroundings. Therefore, it is desirable to maintain structures at a uniform temperature,

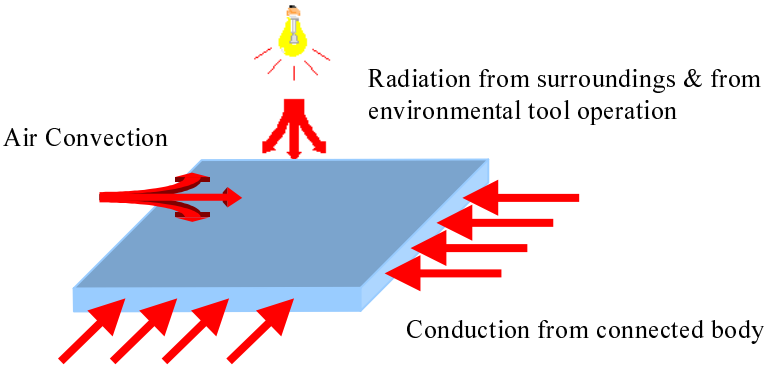


Fig. 1 Environmental Thermal Disturbances

§ Ouyang: ouyanj@unm.edu Tran: tran@me.unm.edu
This work is supported by DOD/ARO DAAD-19-99-1-0196

regardless of environment. As an example, consider a 100mm square structure with a maximum 10nm of dimensional strain allowable to thermal effects. If the structure is made of super-invar (with a CTE of 0.63×10^{-6}/K), it is necessary to maintain its temperature at 160mK; however if the structure is made of aluminum (CTE of 22.5×10^{-6}/K), it is necessary to maintain the temperature should be kept within 5mK. However, aluminum has better specific stiffness, and is therefore more desirable mechanically. Active thermal control using a feedback control system emerges as a logical and practical solution.

Current techniques in the large equipment is to use air or oil shower to maintain the temperature; oil showers are not desirable in a clean room environment. Commercially available thermoelectric coolers (TEC) exploiting the Peltier effect and acting like solid-state heat pump with no moving parts are a very nearly ideal solution for small temperature ranges. With TECs, we can control structure temperature slightly above and below ambient temperature just by controlling current to the TEC; this allows isolation of a structure from possibly turbulent heat exchangers. We will describe the modeling of the system (4×4 TECs), and a proof-of-concept experiment and simulations for multipoint temperature control.

Heat transfer calculation

Consider the calculation shown in Figure 2.

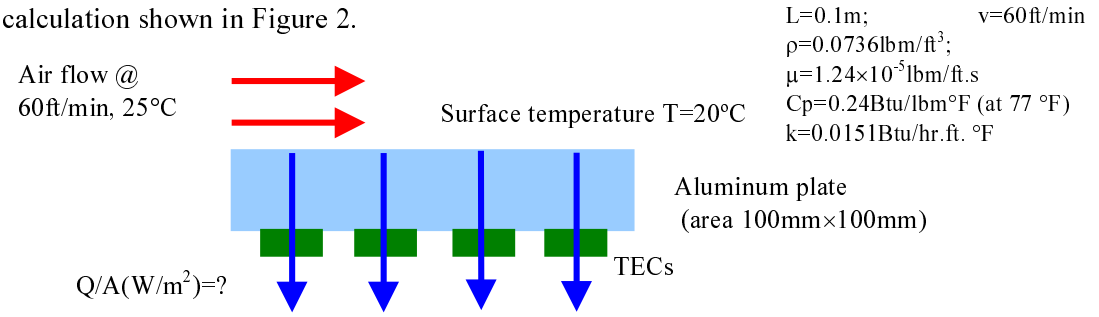


Fig. 2 Heat Transfer Calculation Setup

The heat convection coefficient of the surrounding air $h = \frac{N_u \times k}{L}$ (1)

For the design purpose, take average value for Nusselt number $N_{uL} = 0.664 \times \mathrm{Re}^{1/2} \times \mathrm{Pr}^{1/3}$ (2)

Equation (2) is applied when Reynolds number $\mathrm{Re} < 2\times10^5$, where $\mathrm{Re} = Lv\rho/\mu = 1947$ (3)

We can also calculate Prandtl number Pr as $\mathrm{Pr} = \mu C_p / k = 0.70951$ (4)

From (2), (3) and (4), we can obtain $N_{uL} = 26.132$ and $h = 1.203\,\mathrm{Btu/hr.°F.ft^2} = 7.22\,\mathrm{W/m^2K}$. With a total area of $0.01\mathrm{m}^2$, we have $Q = h \times A \times \Delta T = 0.361\,\mathrm{W}$. The air velocity of 60 ft/min is typical in laminar flow clean rooms (between 60 and 100 ft/min at the face of the HEPA filters). The average cooling power of a Ferrotec-USA module TE6302/065/018A is 5W, so an array of these modules provides more than sufficient cooling or heating power.

Experimental setup

Figure 3 shows a schematic concept for a temperature control experiment.

The temperature range we need to control is not large (±5 K), however the requirement for temperature precision is at millikelvin level. We choose thermistors for their high sensitivity and long term stability.

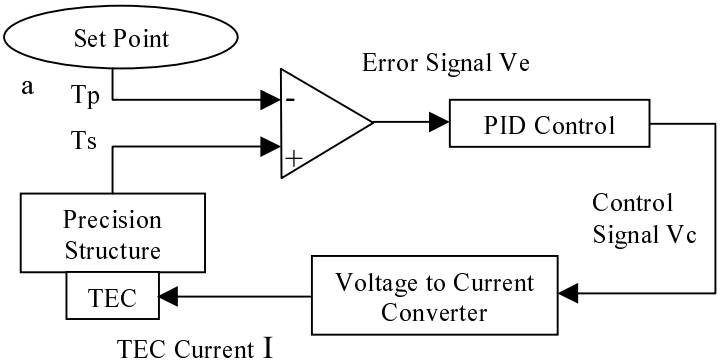


Fig. 3 Experimental block diagram

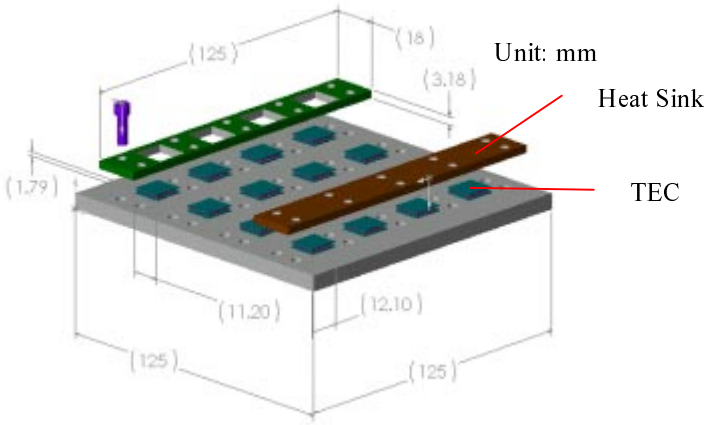


Fig. 4 Test Structure Assembly View



Fig. 5 Photograph of Test Structure

The test structure is an aluminum plate, shown in Figure 4 and Figure 5.

Finite Difference Model

The test structure is virtually cut into 4×4 parts as figure 6 shows and the sixteen elements are controlled by individual TECs and the temperature is measured individually by thermistors. However there exists thermal coupling between these elements. Coupling between diagonal elements (e.g. [1,1] and [2,2]) is smaller than between adjacent elements, and is omitted in the model. The effects of bolting holes are also omitted in the model. Since the system involves both conduction and surface convection effects, it is necessary to determine the Biot number that provides a measure of the temperature drop in the solid relative to the temperature difference between the surface and the fluid.

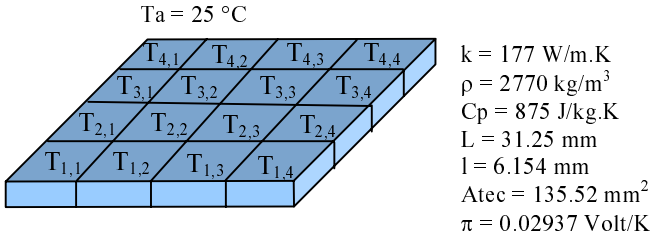


Fig. 6 Finite Difference Model

$$Bi = hL / k = 0.00334 < 0.1 \quad (5)$$

Hence the assumption of a uniform temperature distribution for each element is reasonable.

There are three types of elements for this model: inner ([2,2], [2,3], [3,2], [3,3]); corner ([1,1], [1,4], [4,1], [4,4]); side ([1,2], [1,3], [2,1], [3,1], [4,2], [4,3], [2,4], [3,4]).

- **Inner Nodes**

Take [2,2] as an example. By conservation of energy, we can obtain the following equation

$$\dot{T}_{2,2} = \frac{k}{\rho L^2 C_p}(T_{1,2} + T_{2,1} + T_{2,3} + T_{3,2} - 4T_{2,2}) - \frac{h}{\rho L^2 l C_p}(2L^2 - A_{tec})(T_{2,2} - T_a) - \frac{\pi}{\rho L^2 l C_p} I_{2,2} \quad (6)$$

- **Corner Nodes**

Take [2,2] as an example. By conservation of energy, we can obtain the following equation

$$\dot{T}_{1,1} = \frac{k}{\rho L^2 C_p}(T_{2,1} + T_{1,2} - 2T_{1,1}) - \frac{h}{\rho L^2 l C_p}(2L^2 - A_{tec} + 2Ll)(T_{1,1} - T_a) - \frac{\pi}{\rho L^2 l C_p} I_{1,1} \quad (7)$$

- **Side Nodes**

Take [1,2] as an example. By conservation of energy, we can obtain the following equation

$$\dot{T}_{1,2} = \frac{k}{\rho L^2 C_p}(T_{1,1} + T_{1,3} + T_{2,2} - 3T_{1,2}) - \frac{h}{\rho L^2 l C_p}(2L^2 - A_{tec} + Ll)(T_{1,2} - T_a) - \frac{\pi}{\rho L^2 l C_p} I_{1,2} \quad (8)$$

In equation (6), (7) and (8), the k-term is related with the heat conduction between different elements, h-term heat convection from the surroundings, π-term heat drawn by TEC. Here we omit the resistance and conductance effects of TEC to simplify the problem and assume that heat sinks attached to TECs work perfectly. It is also assumed that the temperature-related coefficients k, h, Cp and π are constant since the temperature range is within ±5 K.

Plugging in the value of parameters, we obtain the state-space equations of the system as follows.

$$\dot{\mathsf{T}} = \mathsf{A}\mathsf{T} + \mathsf{B}\mathsf{I} + \mathsf{D}T_a \tag{9}$$

$$\dot{\mathsf{T}} = \begin{pmatrix} \dot{T}_{1,1} & \dot{T}_{1,2} & \dot{T}_{1,3} & \dot{T}_{1,4} & \dot{T}_{2,1} & \dot{T}_{2,2} & \dot{T}_{2,3} & \dot{T}_{2,4} & \dot{T}_{3,1} & \dot{T}_{3,2} & \dot{T}_{3,3} & \dot{T}_{3,4} & \dot{T}_{4,1} & \dot{T}_{4,2} & \dot{T}_{4,3} & \dot{T}_{4,4} \end{pmatrix}^T$$

$$\mathsf{T} = \begin{pmatrix} T_{1,1} & T_{1,2} & T_{1,3} & T_{1,4} & T_{2,1} & T_{2,2} & T_{2,3} & T_{2,4} & T_{3,1} & T_{3,2} & T_{3,3} & T_{3,4} & T_{4,1} & T_{4,2} & T_{4,3} & T_{4,4} \end{pmatrix}^T$$

$$\mathsf{I} = \begin{pmatrix} I_{1,1} & I_{1,2} & I_{1,3} & I_{1,4} & I_{2,1} & I_{2,2} & I_{2,3} & I_{2,4} & I_{3,1} & I_{3,2} & I_{3,3} & I_{3,4} & I_{4,1} & I_{4,2} & I_{4,3} & I_{4,4} \end{pmatrix}^T$$

$$A = \begin{pmatrix} \Phi & 0.0748E_4 & 0 & 0 \\ 0.0748E_4 & \Gamma & 0.0748E_4 & 0 \\ 0 & 0.0748E_4 & \Gamma & 0.0748E_4 \\ 0 & 0 & 0.0748E_4 & \Phi \end{pmatrix} \qquad \mathsf{B} = -0.00202\mathsf{E}_{16}\,\mathsf{I}$$

$$\mathsf{D} = \begin{pmatrix} \mathsf{D}_1 & \mathsf{D}_2 & \mathsf{D}_2 & \mathsf{D}_1 \end{pmatrix}^T$$

$$\Phi = \begin{pmatrix} -0.1506914 & 0.0748 & 0 & 0 \\ 0.0748 & -0.2253961 & 0.0748 & 0 \\ 0 & 0.0748 & -0.2253961 & 0.0748 \\ 0 & 0 & 0.0748 & -0.1506914 \end{pmatrix} \qquad \mathsf{D}_1 = \begin{pmatrix} 0.0010914 \\ 0.0009961 \\ 0.0009961 \\ 0.0010914 \end{pmatrix}$$

$$\Gamma = \begin{pmatrix} -0.2253961 & 0.0748 & 0 & 0 \\ 0.0748 & -0.3001008 & 0.0748 & 0 \\ 0 & 0.0748 & -0.3001008 & 0.0748 \\ 0 & 0 & 0.0748 & -0.2253961 \end{pmatrix} \qquad \mathsf{D}_2 = \begin{pmatrix} 0.0009961 \\ 0.0009008 \\ 0.0009008 \\ 0.0009961 \end{pmatrix}$$

Where E_4 is a 4×4 identity matrix, $\mathsf{0}$ is a 4×4 zero matrix, E_{16} is a 16×16 identity matrix, T_a is the temperature of the surrounding air. This system is linear and controllable.

Summary

The results of the simulations and experiments will be presented at the conference.

References

1. DeBra, D. B., Victor, R. A., and Bryan, J., "Shower and high pressure oil temperature control," *CIRP Annals* vol 35(1), p. 359-364, 1986.
2. *Technical Reference For Ferrotec Thermoelectric Cooling Modules*, Ferrotec America, Nashua, NH
3. Huang, J. J., and DeBra, D. B., "Liquid Temperature Control for a Hydraulic Turning Machine," *IEEE Control Systems Magazine*, vol 17(4), p. 55-63, 1997.
4. *Internaional Technology Roadmap for Semiconductors*, http://www.itrs.net/ntrs/publntrs.nsf

An experimental investigation of heat pipes at low power inputs

Ajit Borundia and Hy D. Tran[1]
University of New Mexico
Albuquerque, NM 87131

Abstract:
In most precision engineering applications, minimization of thermal distortions is done by the use of either active control using temperature controlled air or liquid showers, artificial heaters or ultra low expansion materials such as Invar, Zerodur. This paper explores the use of heat pipes at low heat loads (0.1 W – 5 W) in precision engineering applications where the requirement of stable and near uniform temperature is stringent. Experiments are performed on a commercial heat pipe and its performance characteristics such as effective thermal conductivity, transient response and temperature gradient at steady state are determined.

Introduction:
The process of miniaturization of existing technologies to the nano and micro level with applications in medical sciences, aerospace, communications, automobiles, electronics have put a tremendous demand for higher accuracy on the manufacturing processes and their concerning metrology. Of all the sources of inaccuracy in a precision machine tool, thermally induced distortions are the most prominent source of error and fortunately they are not random errors. These distortions are worsened by the increase in automation in precision machine tools and instruments since the gauging system is now thermally connected to the heat sources of the machine tool and also its dimensions are large compared to the manual gauging. It is a well-known fact that mitigating the thermal errors in machine tools results in increase in accuracy of the process by an order of magnitude and also improves the repeatability of the process. These effects can be reduced in 3 ways: (a) Compensation (b) Temperature controlled coolant shower (c) Passive cooling using heat pipes, heat sinks.

Compensation requires a thermal deformation model of the machine tool to determine the thermal displacements in real time and to account for these displacements in the process. Unfortunately most of the models that have been developed are either very complex and slow or very simple and inaccurate and do not represent the full range of the operating conditions. Also it is very difficult to compensate for the bending mode of deformation by using the position control system of machine tool.

Temperature controlled coolant shower method is widely used in machine tools since it is very simple to implement and has fast response. It has the added benefit of keeping the work area clean of chips, dust and other extraneous materials, which act as heat source or corrosion agent. Depending on the amount of heat removal and on the requirement of dry work area, the coolant can be air or fluid. Liquid shower has high heat capacity and therefore remove heat under near isothermal conditions thus minimizing thermal distortions. On the other hand, air shower is better in reducing the effect of external heat

[1] Address correspondence to tran@me.unm.edu. This work was supported by ARO DAAD.19.99.0196

sources on the machine tool. It has been found out that in the machine tools, the process of pumping the liquid coolant alone generates a significant amount of heat and acoustic vibration. Bryan et al [1] reports that for the production of mirror finishes a structural motion of less than 2 μ inches is required in the machine tool.

Passive cooling by heat pipes and heat sinks are effective at high operating temperatures and high heat flux. This paper explores the use of heat pipes as a means of passive cooling at low power inputs encountered in the lithography process and other precision engineering applications to maintain stable, very low temperature gradients across the surface. The effective thermal conductivity of heat pipes and the transient response of the heat pipe are determined for different heat inputs ranging from 0.1 W to 5 W at the operating temperatures of 0 C and 17 C respectively.

Experimental Apparatus:

A 4 mm diameter circular copper water heat pipe of length 125 mm from Noron Products with thermistors along its length is insulated as shown as in the Fig 1. A 5W flexible kapton heater (Watlow) is wrapped around the heat pipe at one end. The other end of the heat pipe is inserted into the condenser. A thermocouple is glued 17 mm from the end on the condenser section of the heat pipe. This entire set up is then placed in a vacuum chamber to further reduce the heat loss through convection and radiation to the surrounding environment. The lengths of the condenser and the evaporator sections are 57 mm and 43 mm respectively. The temperature of the condenser was allowed to vary + 0.5 deg C in the case of freezing water (0 deg C) and ± 0.5 deg C in the case of the tap water (17 deg C) to simulate the real life conditions.

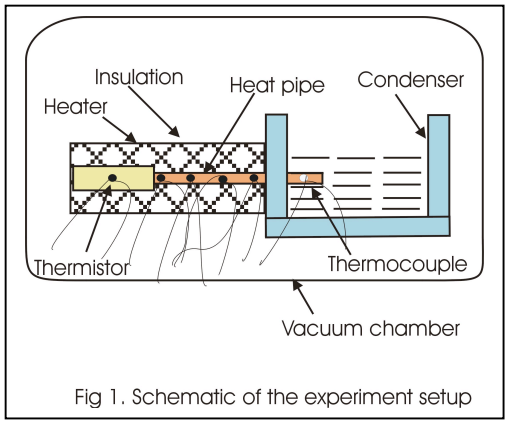


Fig 1. Schematic of the experiment setup

Results & Discussion:

The effective thermal conductivity, transient response, steady state temperature distribution of the heat pipe at 0 deg C and 17 deg C respectively are plotted in figures 2-7. The effective thermal conductivity of the heat pipe was found to be equal to that of copper at 0.4 W and 0.1 W of power input at 0 deg C and 17 deg C respectively. The heat pipe time constant was found to be ranging from 60 to 70 secs for both the cases thus indicating that the transient response was independent of the operating temperature of the heat pipe at low power inputs. The temperature gradient along the adiabatic section of the heat pipe at steady state was found to be varying linearly from –0.068 deg C /mm to –0.28 deg C/mm at 0 deg C and from –0.03 deg C/mm to –0.26 deg C/mm at 17 deg C between power inputs 0.1 W to 5 W respectively. This shows that the temperature gradient is not very sensitive to the operating temperature of the heat pipe.

The conductivity of OFHC copper is approximately 380 W/m-K. With heat flux levels as low as 1 W, the effective thermal conductivity of a heat pipe is several times that of a solid conductor.

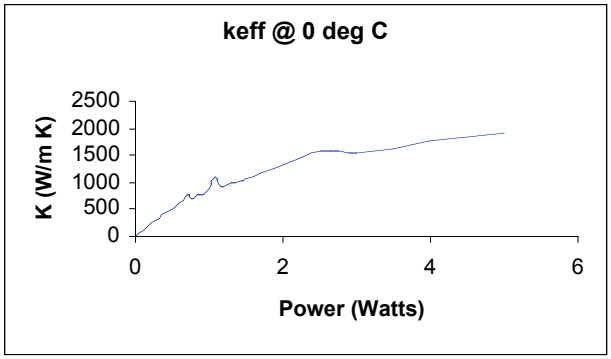


Fig 2. Effective conductivity at 0 deg C.

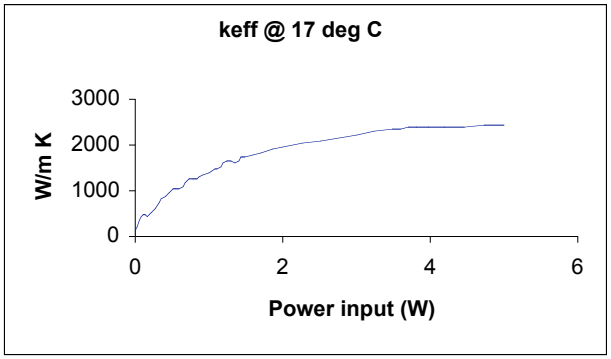


Fig 3 Thermal conductivity at 17 deg C

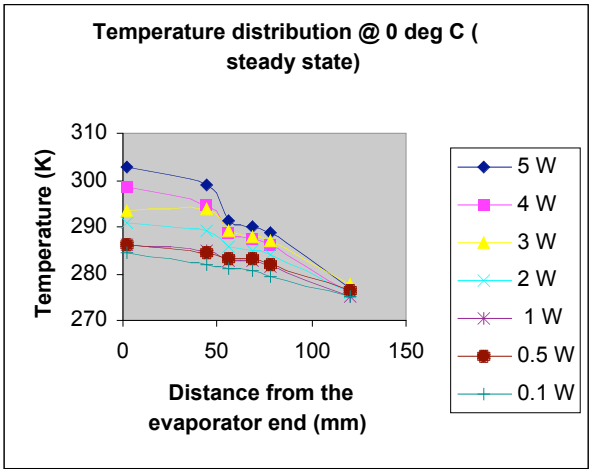


Fig 4 Temperature distribution at 0 deg C

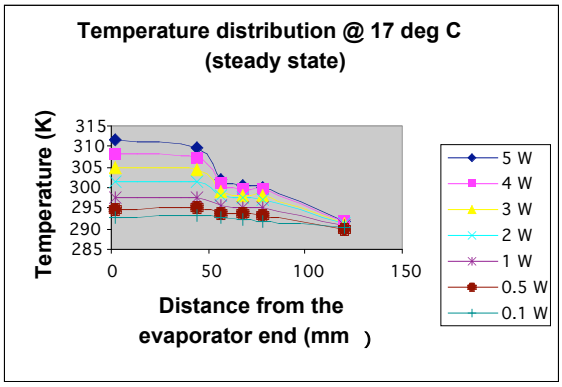


Fig 5 Temperature distribution at 17 deg C

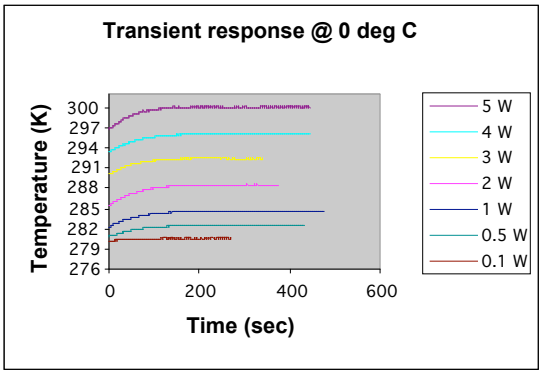


Fig 6 Transient response at 0 deg C

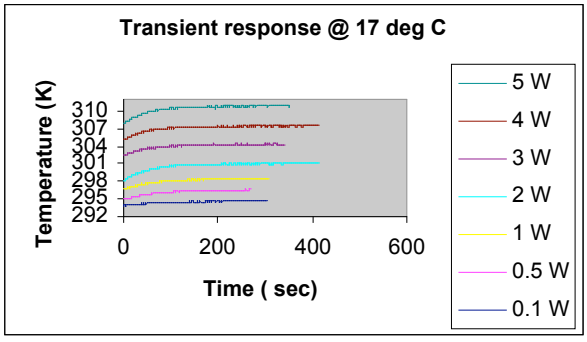


Fig 7 Transient response at 17 deg C

Conclusion:

The effective thermal conductivity, transient response and temperature gradient of the heat pipe at low power inputs were evaluated. From the results we conclude that heat pipes respond quite well in transporting randomly varying low heat loads which occur either due to friction between components in motion or due to concentrated heat input in case of the lithography process at stable temperature gradients. In the future, this would open up the possibility of manufacturing precision equipments by using a structure combining a structure with desirable mechanical properties with heat pipes and active thermal control to minimize thermally induced strain.

References:

1. Bryan J B et.al; A practical solution to the thermal stability problem in machine tools, SME Technical paper, No. MR72-1378 1972.
2. Cao Y, Gao M; Experiments and Analyses of Flat Miniature Heat pipes, Journal of Thermophysics and heat transfer vol 11 no.2 1997, pp 158-164.
3. El-Nasr A A, Haggar S M; Effective thermal conductivity of heat pipes, Heat and Mass transfer vol 32 1996, pp 97-101.
4. Fraser S, Attia M H, Osman M O M; Modeling, Identification and Control of Thermal Deformation of Machine Tool Structures, Part 5: Experimental Verification, Journal of Manufacturing Science and Engineering, vol 121, 1999, pp 517-523.
5. McClure R; Manufacturing Accuracy Through the Control of Thermal Effects, Thesis, LLNL, UCRL-50636 1969.

2-D Self-Calibration for Scale Based Metrology in Nanolithography

Xiaoming Lu & Hy. D Tran[1]
Department of Mechanical Engineering
University of New Mexico
Albuquerque, NM 87131

Abstract

Two-dimensional precision stages for semiconductor lithography require sub-nm-level accuracy. Current ultraprecision stage metrology based on heterodyne laser interferometers with a 632.8-nm HeNe laser and interpolation to λ/2048 provide resolution of 0.3 nm. Unfortunately, the refractive index of air varies with CO_2 and water vapor content, as well as with air turbulence. Without expensive atmospheric compensation, the actual accuracy may only be sub-μm for operation in air.

In our approach two scale based metrology system (OPTRA NanoGrid) are used. The NanoGrid greatly reduces possible effects from air turbulence and Abbe offsets from using separate X and Y scales. However, the metrology is based on an physical artifact, and the accuracy of the metrology is limited to the accuracy of the grid artifact itself: about 1μm. Physical scales may provide greater precision, but not always greater accuracy. Real-time self-calibration is implemented to improve the accuracy of stage to the level limited by repeatability of stage.

Fourier transform is used to decouple the distortion of artifact metrology and stage metrology at sampling points from 3 measurement view. Based on feedback principle high accuracy translation and rotation is implemented to minimize misalignment error. Square reversal is used to calculate rotation error δγ. Iterative algorithm is developed to isolate the rotation error δγ in self-calibration.

Keywords: Self-Calibration ,Nanolithography, two-dimensional stage, scale based metrology

1.Introduction

The key issues in self-calibration is how to solve measurement equations. In other words, isolate the scale and stage distortion from measurement equations when there are misalignment errors during translation and rotation. In a previous approach, Ye, Pease et al. (1997) developed a discrete Fourier transform-based algorithm by approximating A(x,y) and S(x,y) with Fourier series, which is numerically robust to measurement noise. The alignment error is isolated in the statistical meaning by using least square method under several assumption, which is effective when the measurement noise has a normal distribution and required space sample interval is big enough. Second, translation and rotation error will affect the sample position. This will be problematic when space sample interval is small e.g sub-μm. Third, because the x and y translation and rotation values of the stage point set can not be given by the algorithm or known before the self-calibration, it is difficult to find the origin and the axis orientation which meet the assumptions. Thus the stage error map obtained from this method has an unknown coordinate system, which might not be directly related to the machine coordinate system.

[1] luxm@unm.edu; tran@me.unm.edu This work is supported by DOD/ARO DAAD-19-99-0196

2.Testbed physical realization of an X-Y stage for Nanolithography

The testbed is an air-bearing ,capstan drive interferometric lithography stepper, with HP laser interferometers for feedback. This stage is configured as a stacked-slide X-Y stage. The moving stage is supported by air bearings on the granite surface plate. Two scale based metrology systems are mounted on the same plane of moving stage to implement real-time self-calibration. The configuration in the plane is shown in Figure 1.

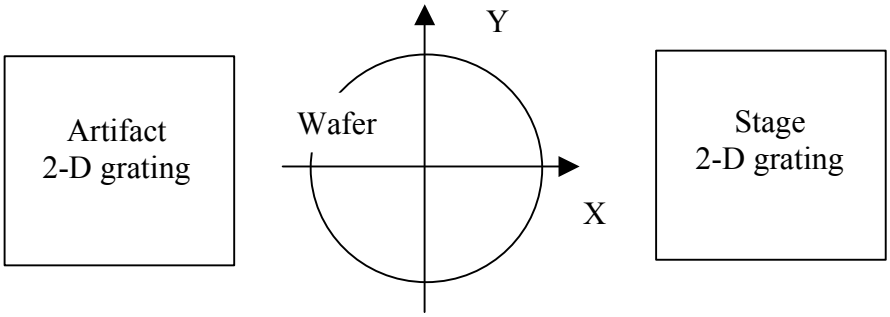


Figure 1 Configuration of 2-D real-time self-calibration

Scale based metrology system is a 2-D ultra-high resolution diffraction interferometer. It can measure X and Y plane motion simultaneously with 0.3 nm resolution and to detect small angle motion around Z axis.

One of the main error source in nanolithography come from the disturbance of guideway. The translation error of guideway in X and Y direction will be canceled by feedback of accurate stage metrology. The disturbance from the rotation error of guideway around Z axis will be minimized based on Abbe principle in X direction. For 1 nm accuracy, compare with traditional configuration, the tolerance of rotation error will increase from 1 μrad to 80 μrad. In Y direction the rotation error of guideway around Z axis will be measured by two scale based metrology system and compensated with software. So the accuracy of nanolithography mainly depends on the accuracy of metrology.

3. Mathematical model and Self-Calibration Algorithm

3.1 Decouple 2-D grating distortion when there is no alignment error

Two scale based metrology system are used to generate measurement view. One is stage metrology .The other is artifact metrology. The distortion s(x,y) of stage grating and distortion a(x,y) of artifact grating at sampling points (x_m,y_n) can be expressed as

$$\begin{aligned} s(x_m, y_n) &= s_x(x_m, y_n)e_x + s_y(x_m, y_n)e_y \\ a(x_m, y_n) &= a_x(x_m, y_n)e_x + a_y(x_m, y_n)e_y \end{aligned} \qquad (1)$$

We need 3 measurement views to solve $s(x_m,y_n)$ and $a(x_m,y_n)$. In View 0 or home view, two metrology system is aligned with the air bearing of the stage by minimizing the angle between the moving direction of air bearing and X-axis of the metrology. $s(x_m,y_n)$ and $a(x_m,y_n)$ is compared at sampling points, so

$$v_x 0(x_m, y_n) = s_x(x_m, y_n) - a_x(x_m, y_n) \qquad (1a)$$

$$v_y 0(x_m, y_n) = s_y(x_m, y_n) - a_y(x_m, y_n) \qquad (1b)$$

In View 1, the sensor head is translated from View 0 by one sample site interval relative to the artifact grating, yield the measurement equations

$$x_x 1(x_m, y_n) = s_x(x_m, y_n) - a_x(x_{m+1}, y_n) \quad (2a)$$

$$v_y 1(x_m, y_n) = s_y(x_m, y_n) - a_y(x_{m+1}, y_n) \quad (2b)$$

Based on translation property of 2-D discrete Fourier transform to solve $A_x(u,v)$ from equation (1) and (2)

$$A_x(u,v) = \frac{V_x(u,v)}{1 - e^{j2\pi\frac{u}{M}}} \quad (3)$$

where

$$A_x(u,v) = \frac{1}{MN} \sum_{x_m=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{y_n=-\frac{M-1}{2}}^{\frac{N-1}{2}} a_x(x_m, y_n) e^{-j(\frac{2\pi}{M}ux_m + \frac{2\pi}{N}vy_n)}$$

$$V_x(u,v) = \frac{1}{MN} \sum_{x_m=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{y_n=-\frac{M-1}{2}}^{\frac{N-1}{2}} (v_x 0(x_m, y_n) - v_x 1(x_m, y_n)) e^{-j(\frac{2\pi}{M}ux_m + \frac{2\pi}{N}vy_n)} \quad (4)$$

Equation (3) only gives solution at $u \neq 0$. Based on the rotation property of 2-D Fourier transform $A_x(0,v)$ will be solved from equation (5) in View2 where artifact grating is counterclockwise rotated 90^0 relative to View 0.

$$V_x 2(x_m, y_n) = s_x(x_m, y_n) - ROT90(a_y(x_m, y_n)) \quad (5a)$$

$$V_y 2(x_m, y_n) = s_y(x_m, y_n) - ROT90(a_x(x_m, y_n)) \quad (5b)$$

where ROT90 denotes "90^0 rotation."

3.2 Minimize the misalignment error based on feedback and iterative algorithm

The sensor head is mounted on 6 degree kinematics stage. The direction and location of sensor head relative to the 2-D grating are determined by the 6 balls on the kinematics stage. A picomotor is used to drive the relative movement of sensor head relative to the 2-D grating with 30 nm resolution . The sensor head is aligned with the 2-D grating stage by minimizing the angle between the moving direction of 6 degree kinematics stage and X-axis of the metrology. The relative translation between sensor head and 2-D grating will be measured by 2-D metrology system itself. The feedback is used to guarantee required one sample site interval translation accuracy determined by sampling requirement.

A 2-D grating is mounted on the bearing and rotation can be driven by a picomotor with 30″ resolution. Square reversal is used to measure the 90^0 rotation error $\Delta\gamma$ which will be reduced by feedback to the level limited by the picomotor resolution.

The accuracy to implement translation and rotation depends on the accuracy of 2-D grating. The translation error can be minimized by implement 1-D self-calibration in X axis first.

The 2-D self-calibration error due to 90^0 rotation error $\Delta\gamma$ can be compensated if $\Delta\gamma$ is accurately measured when $\Delta\gamma$ is tiny. From experiment and computer simulation the measurement accuracy of $\Delta\gamma$ is insensitive to the accuracy of metrology which means error transfer coefficient is less than 1. This will guarantee that an iterative algorithm will be convergent. So an iterative algorithm is developed to minimize disturbance from measurement accuracy of rotation error $\delta\gamma$.

4. Experiment results and summary

Real-time calibration was implemented on the Nanolithography testbed. Dynamic data sampling methodology was developed to minimizes effects of quasi-static disturbance from measurement environment change, such as temperature. The repeatability of scale based metrology is about 8.5 nm (see Fig 2). The initial calibration result is shown in Fig 3. The distortion of 2-D Grating is about 1μm . Real time self-calibration will increase the accuracy of scale based metrology from 1 μm to the level limited by the measurement noise.

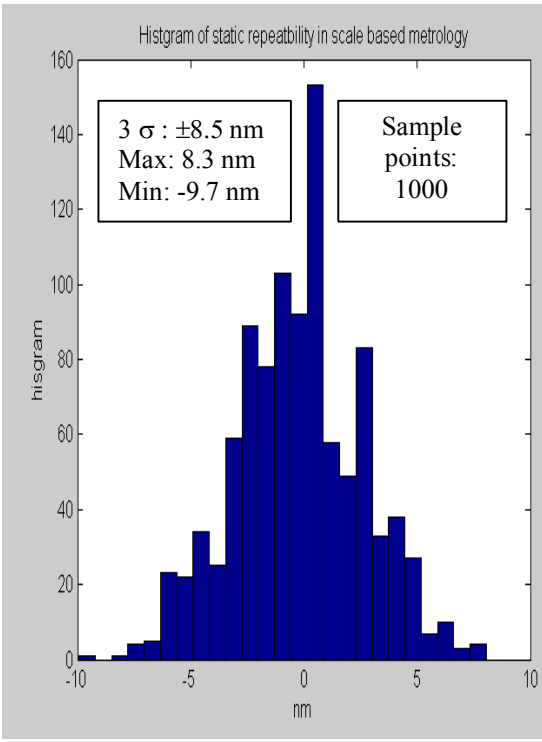


Figure 2 Short term repeatability of scale based metrology

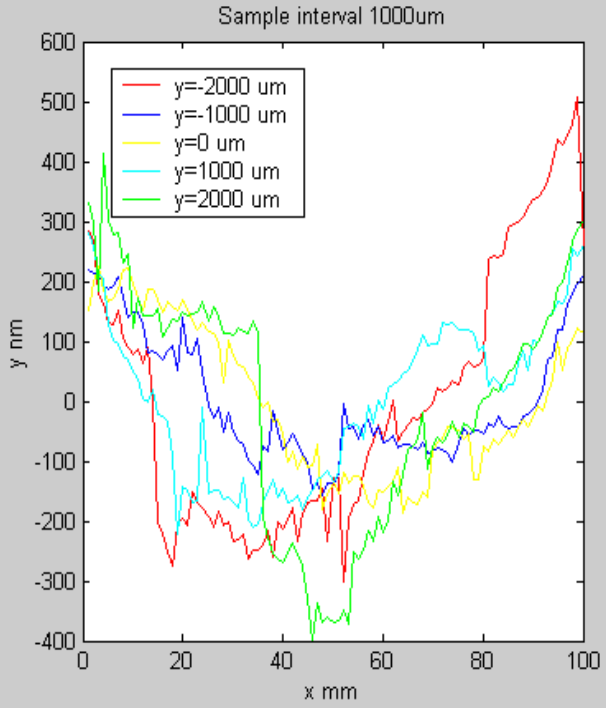


Figure 3 Distortion of 2-D grating

References

Evans,C,J.,Hocken,R,J., and Estler,W.T.,"Self-calibration:reversal, redundancy,error separation, and 'absolute testing,'" CIRP annals,Vol45(2),1996.

Ye,J.,Takac,M.,Berglund,C.N.,Owen,G.,and Pease,R.F.,"An exact algorithm for self-calibration of two-dimensional precision metrology stages," Prec Eng vol.20(1),p.163-32,1997.

A Fuzzy Logic Based Adaptive Feedforward PI Controller for Nanometer Positioning

Xiaoming Lu & Hy D. Tran[1]
Department of Mechanical Engineering
University of New Mexico
Albuquerque, NM 87131

Abstract

A technology driver in high precision manufacturing is a position measurement and control system which allows manufacturers to control the movement of their equipment at extremely high levels of precision. Performances of controllers play important roles in nanometer level positioning. A robust, disturbance resistant controller is necessary for nanometer positioning. However, the time varying and nonlinear properties of plants frequently result in the performance degradation of widely used PID controllers.

The performance of feedback system can be improved considerably by combining PI feedback control with feedforward control together. Feedback controllers are very capable of compensating disturbances because feedback controllers are basically error driven. However, they often suffer from a trade-off between high performance and robust stability. With good dynamic property knowledge available, a feedforward controller may be able to prevent control errors, because feedforward controller output is based on the reference, instead of the error signal. Feedforward controllers improve the control performance by including a feedforward term to the PI output so that the controller can react to the command more quickly in order to bring the plant to the desired setpoint. A fuzzy logic controller will adjust the feedforward and PI feedback gains (K_{ffv} ,K_{ffa} , $K_{ffd,}$ K_P and K_I) to adapt the change of dynamic behavior caused by the time varying and nonlinear properties inherent in the positioning systems.

The developed control algorithm is simulated in MATLAB, and is implemented with a TMS320M67 DSP in an experimental nanolithography stage. A X-Y grating based metrology is used to measure the position of stage, which is more robust to environmental change and more suitable with high performance servomechanisms. Experiments show that stage will follow 10 nm step input with nm level steady state error, which is mainly caused by floor and acoustic vibrations.

1. Introduction

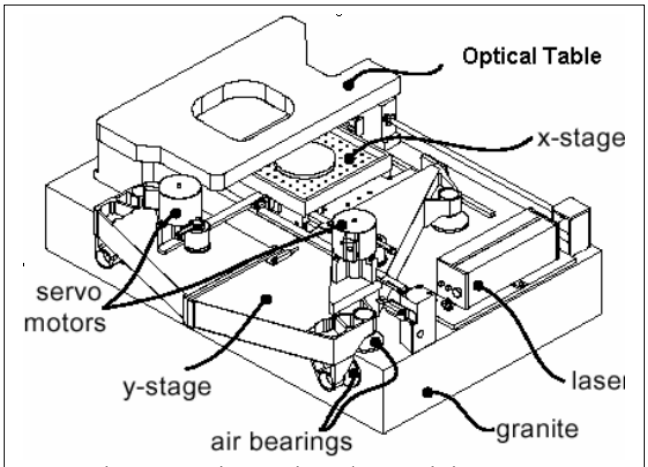


Fig 1 2-Dimensional precision stage

A robust and disturbance resistant controller is necessary for nanometer positioning. However, the time varying and nonlinear properties inherent in the system dynamics frequently result in degradation of PID controller performance. For example, air-bearing supported capstan drives are one of several systems used in precision positioning stages (such as the one presented here). They exhibit very low friction and no backlash. However, capstan friction results in non linearity. Open loop testing indicates that the

[1] Lu: luxm@unm.edu; Tran: tran@me.unm.edu. Supported by DOD/ARO DAAD-19-99-0196

frequency responses of in macro (1 μm maximum amplitude) and micro (50 nm) range motion are different.

Fixed gain PID controllers are not robust enough and further improvement is needed. Rao and Ro stated that state feedback controllers had better robustness and disturbance rejection capability than PID controllers do. Ro and Hubbel adopted Model reference adaptive control (MRAC) theory in controlling a stage. As a result, separate mathematical models and MRACs had to be developed for micro and macro modes according to the scale of the movement. Although this dual-model MRAC did improve the consistency between micro and macro modes as opposed to a single model MRAC , the problem of deciding the proper threshold for switch from micro to macro mode would make a dual model MRAC controller unsuitable for, say, a 500 nm movement.

2 Design of Fuzzy Logic Based Adaptive Feedforward PI Controller

A block diagram of an analog PI controller with fuzzy logic based adaptive feedforward loop is shown in Fig 2. The PI feedback controller is designed for robustness and accuracy. The integrator increases the type number of the system, allowing it to track

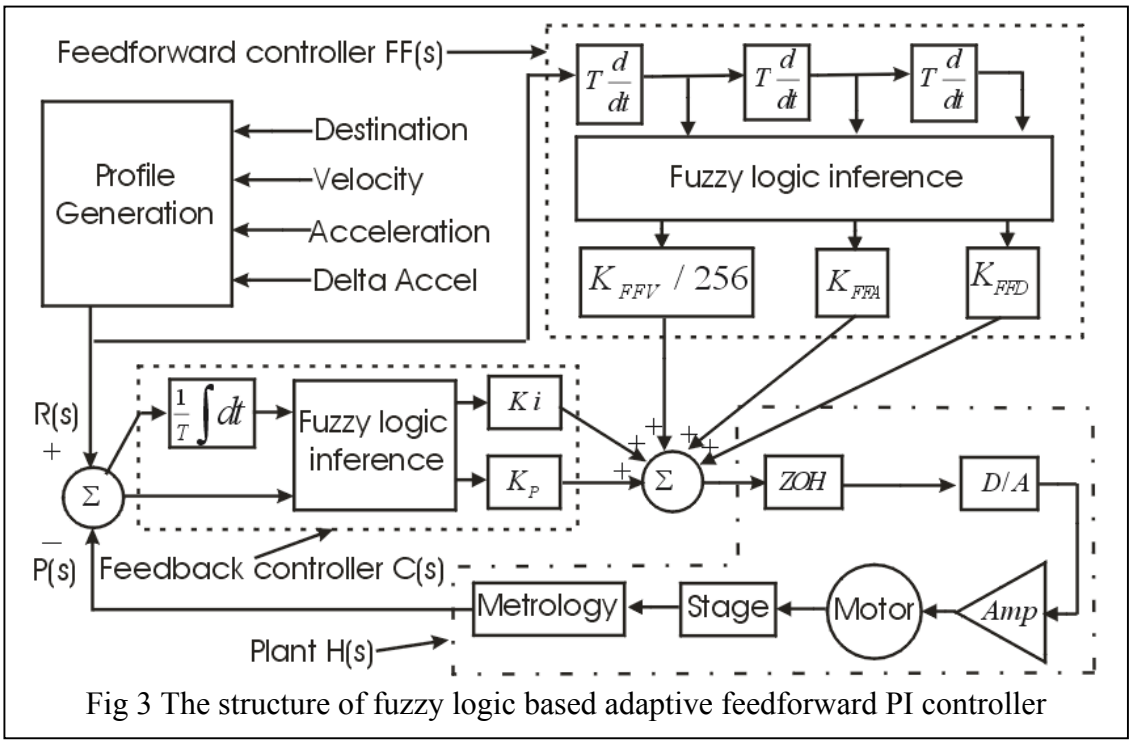


Fig 3 The structure of fuzzy logic based adaptive feedforward PI controller

with no (or limited) error. In other words, the integrator will eliminate steady state error to guarantee nm level control accuracy. However, implementing an I-controller by itself will cause instability, and it must be combined with a proportional action to stabilize the system. On the other hand, the PI controller does not help to speed up the system response. Feedforward controllers improve the control performance by including a feedforward term to the PI output so that the controller can react to the command more quickly, in order to bring the plant to the desired setpoint. Note that this function is outside the feedback control loop, and thus does not affect the system's stability. Feedforward helps to minimize following error during motion, by generating most of the DAC output signal from the motion profile rather than the position error.

The performance of feedforward and feedback controller depends on the dynamic behavior of plant. Also, feedforward controllers may require large control signals which are beyond the permissible maximum output of the drive. A fuzzy logic controller will adjust the

feedforword and feedback gain (deciding the proper threshold for switching from micro to macro mode) according to the reference gain to deal with those issues.

3 Simulation and Experiment Setup

The proposed control algorithm is simulated in MATLAB. Simulation results show that the fuzzy logic adaptive controller will adapt to change of dynamic behavior caused by the time varying and nonlinear properties inherent in the positioning system dynamics. The testbed for optical Nanolithography which is used to investigate the result of adaptive feedforward controller (shown in Figure 1) is an interferometric lithography stepper, featuring an air-bearing stage, with capstan drive, with HP laser interferometers for feedback. This stage is configured as a stacked-slide X-Y stage. The moving stage is supported by air bearings on the granite surface table.

An XY grid-based planar encoder system is used to measure 2-dimensional ultra-precise

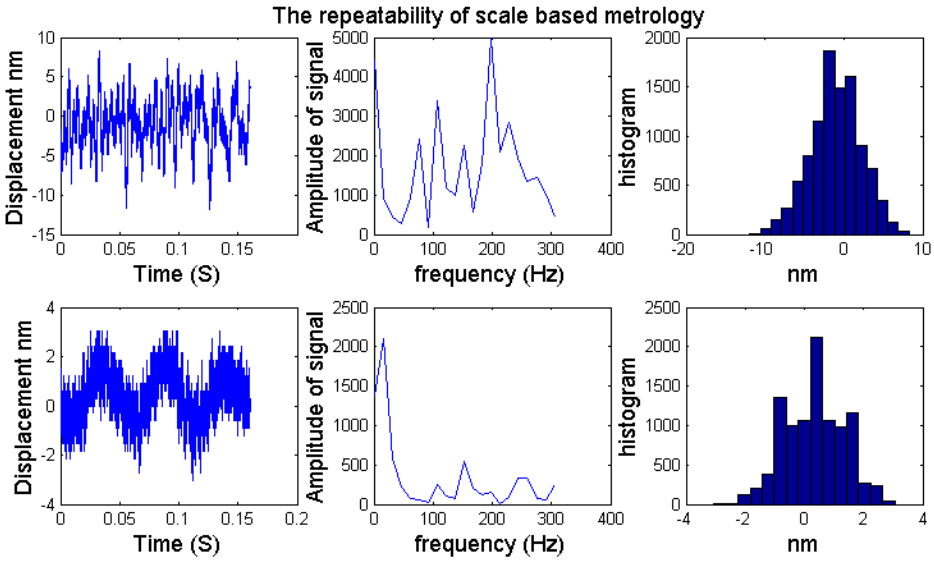


Fig3 Comparing results for repeatability testing of metrology with air bearing turn on (upper) and air bearing turn off (bottom)

planar displacements of stage, which avoids the turbulence effects which are commonly encountered with laser interferometers or the Abbe errors associated with separate linear scales. The 2-D grid measurement resolution is 0.3nm The update rate for this metrology is 375 kHz, which is suitable for high performance servo-mechanisms.

The control algorithm is implemented digitally with TMS320C67 digital signal processor which is mounted on an Innovative Integration M67 card. This is a 32-bit standard full-size PCI card. The digital position outputs of the X-Y grating based metrology are accessed through the DSP card parallel digital I/O connector, which is faster than through the PC's bus. The sampling period is 10μs. The 16 bit D/A output of the DSP is amplified and then drives the DC motor.

4 Results And Future Work

The measurement noise is shown in Fig 3. When the stage is not floated (resting on granite plate), the amplitude of measurement noise is about 2.5 nm. Comparatively, the amplitude of measurement noise is about 8.5 nm when stage is floated. This may be because the air-bearing is more sensitive to the vibration disturbance from floor or the air pressure in the air bearing changes with time. The controller is tested with different size of step input. The stage will follow 10 ,20 ,50 ,500 and 5000 nm step input with the steady state error around nano-meter level after filtering out the vibration disturbance (Shown in Fig 4).

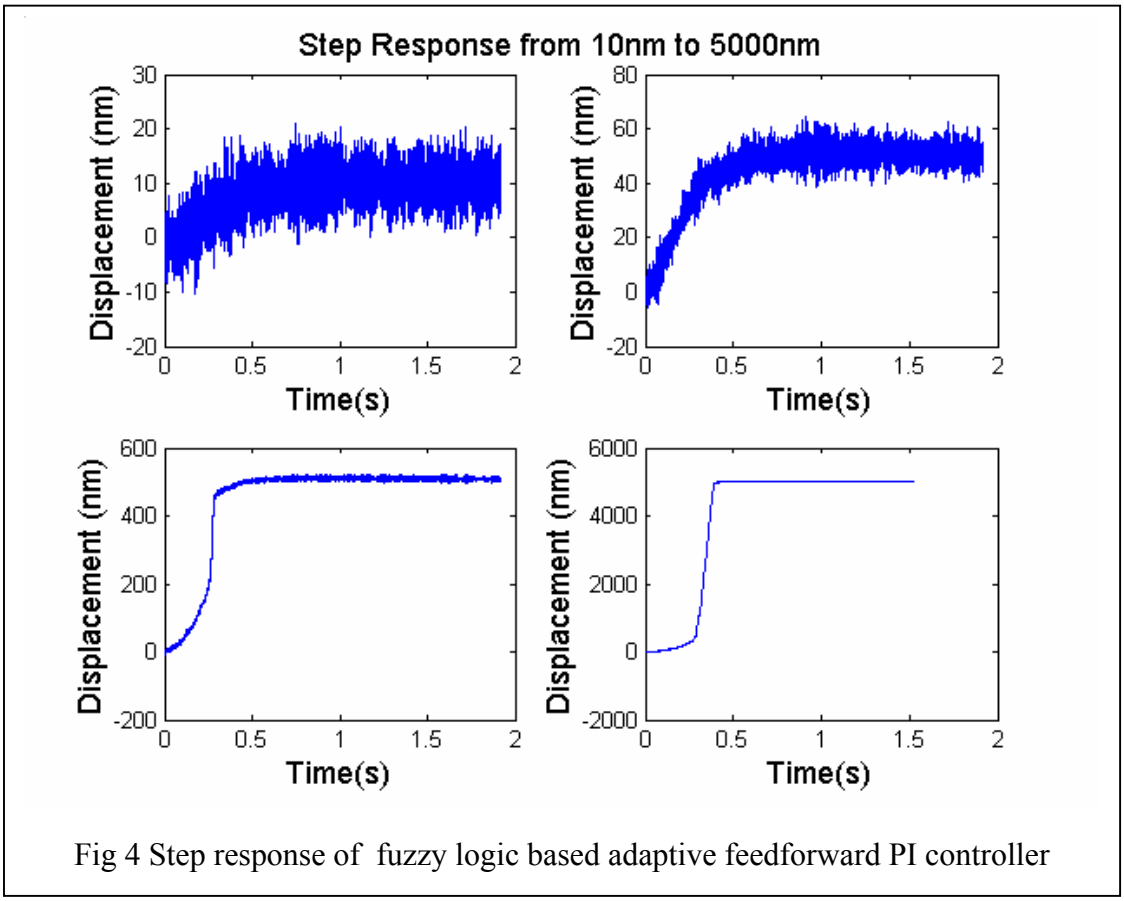


Fig 4 Step response of fuzzy logic based adaptive feedforward PI controller

The PI feedback controller ensures the control accuracy of stage motion. Adaptive feedforward controller improve the transient control performance considerably when there are the changes of dynamic behavior caused by the time varying and nonlinear properties inherent in the positioning systems.

In our current approach, the gain of feedforward controller is determined by trial and error based on experiments. One possible approach in the future is *learning* feed-forward (LFF) controllers. LFF may be able to prevent control errors, because its output is based on the reference, instead of the error signal. They are not necessarily based on a physical process model and are potentially able to learn and reproduce an 'arbitrary' continuous function to any desired degree of accuracy, even if it concerns non-linear and/or time-variant functions.

References

[1] Junhong Mao, Hiroyuki Tachikawa, Akira Shimokohbe," Precision positioning of a DC-motor-driven aerostatic slide system", Precision engineering, Vol 27,No1,January 2003, p32-41

[2] Rao Gs, RoPI. Submicrometer control of a traction drive using state feedback and estimation. Prec Eng 1995; 17: 125-130.

[3] Lin MC. Chen Js. Experiments toward MRAC design with intergral compensation for motor drives. JSMe 1995;38:68-77.

[4] Ro PI, Hubbe, PI. Model reference adaptive control of dual-mode micro/macro dynamics of ball screws for nanometer motion. J Dynamsyst Measure Control 1993;115:103-108.

[5] C.L.Chao,J.Neou. Model reference adaptive control of air-lubricated capstan drive for precision positioning. Prec Eng 2000; 24: 285-290.

Optimizing the fluid dispensing process for immersion lithography

A. Abdo,[a] G. Nellis,[b] A. Wei, M. El-Morsi, and R. Engelstad
Computational Mechanics Center, University of Wisconsin, 1513 University Avenue, Madison, Wisconsin 53706

S. R. J. Brueck and A. Neumann
Center for High Technology Materials, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

(Received 3 June 2004; accepted 4 October 2004; published 14 December 2004)

The concept behind immersion lithography is the insertion of a high refractive index liquid in the space between the final projection lens of an exposure system and the device wafer to improve the overall resolution of the exposure process. Computational fluid dynamics (CFD) simulations were performed in order to investigate the process of initially filling the lens-wafer gap with immersion fluid. The CFD models were used to investigate the effects of dispense velocity, gap height, and fluid dispense angle on the fill process; specifically on the possibility of air entrainment. The simulations revealed that there is an optimal region in the parameter space of gap height and dispense velocity for which the gap fills completely. Outside of this region, either excessive inertial or surface tension forces cause an undesirable, incomplete filling process. The optimal region was found to shift somewhat based on the fluid dispense angle. Finally, experiments were performed to verify the CFD models. The CFD simulations and the experimental results were in good agreement, both qualitatively with regard to the shape and evolution of the free surface and quantitatively with regard to the velocity of the contact line. © *2004 American Vacuum Society.* [DOI: 10.1116/1.1824065]

I. INTRODUCTION

Immersion lithography has been proposed as a method for extending optical lithography resolution to 45 nm and below.[1] The premise behind immersion lithography is to improve resolution by increasing the index of refraction in the space between the final projection lens of an exposure system and the device wafer by inserting a high index liquid in place of the low index air that currently fills the gap. The liquid in the gap is an optical element and therefore must have a uniform index of refraction. To this end, the gap must be completely filled with liquid (water for 193-nm lithography with other fluids being evaluated for 157-nm lithography[2]) and air bubbles cannot be tolerated.

In order for immersion lithography to be practical, the introduction of liquid into the wafer-lens gap must not adversely impact the overall manufacturing process. Since the immersion fluid acts as an optical component during the lithographic process, it must retain a high and uniform optical quality. It is critical that the fluid management system reliably and rapidly fill the entire lens-wafer gap,[3] maintain the fill under the lens throughout the entire exposure process, and ensure that no bubbles are entrained during filling or scanning. Optimization of the fluid dispensing process is essential to the success of immersion lithography; therefore, two- and three-dimensional (2D and 3D) computational fluid dynamics (CFD) models have been developed to simulate the filling process. Several previous studies have used CFD simulations to investigate potential problems that may arise when the immersion lithography process is operating in a quasisteady manner; these problems include fluid heating[4,5] and air entrainment due to free surface flows over wafer topology.[6] The simulations presented in this work parametrically study the filling process that must occur prior to initiation of immersion exposures. The parameters investigated include the fluid dispense velocity and angle as well as the gap height between the lens and the wafer.

An optimal region in the parameter space of fluid velocity and gap height is identified. The region is delineated on one side by a critical Reynolds number, above which the dispense jet impinges on the wafer forming a thin film that does not completely fill the gap. On the other side, the region is delineated by a critical Weber number, below which a meniscus is formed from the dispense port and crawls along the lens. The subsequent rupture of the meniscus upon its eventual contact with the wafer is likely to entrain unacceptable air bubbles. The fluid dispense angle has an effect on the location of this optimal region.

In order to verify some aspects of the CFD simulations, the model predictions were compared with an experiment in which water was used to fill a gap that replicates the lens-to-wafer gap in an immersion lithography tool. There was good agreement between the CFD and experimental results, both qualitatively in terms of the shape of the free surface and its progression across the gap and quantitatively in terms of the velocity of the filling process.

The use of 2D CFD models for the parametric studies and 3D CFD models for experimental verification was required by practical constraints. It was important to verify that the CFD models were predictive and therefore some experimental comparison was required; however, the experimental apparatus involved 3D flow and therefore a 3D CFD model

[a]Currently with IBM Microelectronics Division, East Fishkill, NY.
[b]Electronic mail: gfnellis@engr.wisc

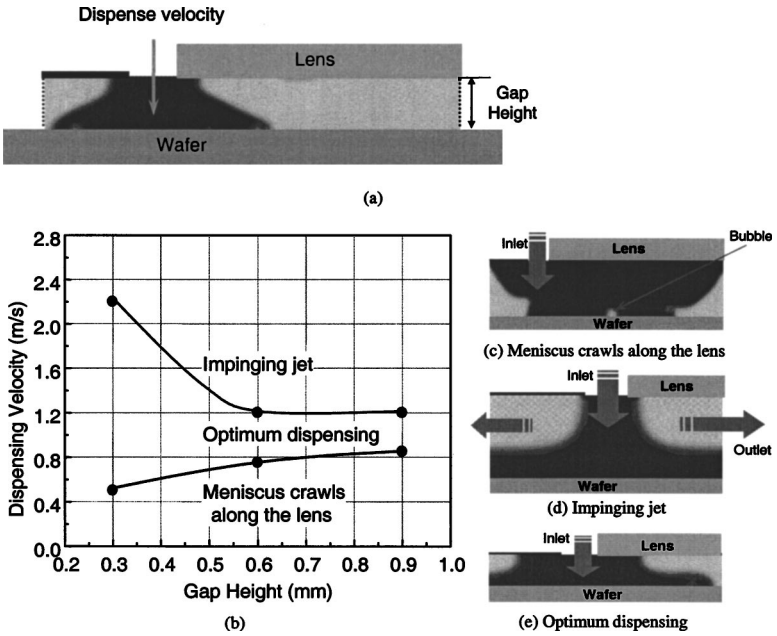


FIG. 1. (a) Schematic of the dispensing process. The behaviors are delineated in (b) which shows the fluid dispense velocity as a function of gap height with three characteristic regions identified. The effects of dispense velocity and gap height on the fill process are shown in (c) for the behavior just after the meniscus breaks for a low dispense velocity (note trapped air bubble), in (d) with the thin, high velocity fluid film associated with a high dispense velocity, and in (e) for the complete and rapid filling process that occurs when an acceptable dispense velocity is used.

was necessary to yield quantitative agreement. Unfortunately, the use of a 3D CFD model to carry out the parametric studies would require a prohibitive amount of computational time. Therefore 2D CFD models were used for the parametric studies of the underlying physics. The correlation between the 2D and 3D results has been observed experimentally; for example, it is possible to experimentally observe the transition between impinging jet and optimal dispense at a constant gap height by increasing the velocity using the experimental test facility described in this paper. In other testing, the transition between meniscus crawl and optimal dispense behavior has been observed when the dispense velocity is reduced. While the experimentally observed transition velocities are not exactly equal to those shown in the paper, they are nominally consistent with the critical Weber number and Reynolds number identified by the 2D CFD model.

II. CFD MODELING RESULTS

Figure 1(a) shows a schematic of the filling process used in the two-dimensional CFD simulations. Immersion fluid (water for these simulations) was dispensed with a uniform velocity from a port adjacent to the lens surface. The lens and dispense port are positioned at the same distance from a wafer which is stationary for the simulations and experiments described in this article. The progress of the filling process is monitored after the initiation of the fluid dispense.

The CFD simulation results showed that the fluid dispense process is sensitive to a number of parameters including the fluid dispense velocity, the wafer-to-lens gap height, and the fluid dispense angle. Three different gap heights (i.e., 0.3 mm, 0.6 mm, and 0.9 mm) were examined in a parametric study; for each height, the fluid dispense velocity was varied from 0.0 to 3.0 m/s. The parametric studies showed that the dispense process behavior falls into different regimes depending on these process variables, as illustrated in Fig. 1(b). The simulation was conducted with the three gap heights only; the connecting line shown in the graph is qualitative, to illustrate the trend of the simulation results.

At low fluid dispense velocity the liquid forms a hanging droplet that does not immediately touch the wafer surface. As more liquid is added to the droplet it tends to crawl along the hydrophilic lens surface; eventually the meniscus touches the wafer and at this point it breaks very rapidly, as shown in Fig. 1(c). This behavior is undesirable because the sudden rupture of the meniscus is unpredictable and may entrain macroscopic air bubbles that can be clearly identified in Fig. 1(c). This air cannot escape from the meniscus to wafer gap during the rupture process and therefore forms small bubbles under the influence of surface tension forces. At each gap height, a series of simulations at progressively lower dispense velocities was used to delineate the transition from a completely filled gap to this meniscus crawl behavior, as shown in Fig. 1(b). The specification of this transition is somewhat qualitative but is based on the extent to which the meniscus has progressed in the direction of the gap at the time when it touches the wafer.

At very low fluid dispense velocities, the inertia of the dispensed fluid is insufficient to overcome surface tension and break the meniscus. Thus it has the chance to crawl along the lens. The transition into this behavior is therefore governed by a balance between fluid inertia and surface tension which constitutes the Weber number (We):

$$\mathrm{We} = \frac{\rho v^2 D}{\sigma}, \tag{1}$$

where ρ and σ are the density and surface tension of the fluid, D is the length scale associated with the dispense port, and v is the dispense velocity. Note that the Weber number as defined in this way is independent of gap height which is consistent with the results shown in Fig. 1(b), in that the gap

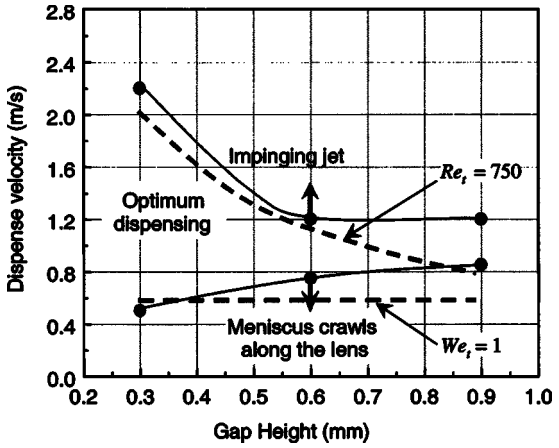


FIG. 2. Delineation of the different physical behaviors of the fluid dispense process based on critical Weber and Reynolds numbers.

height has only a very small effect on the fluid dispense velocity at the onset of the meniscus crawl behavior.

At very high fluid dispense velocities the fluid inertia is sufficient to create a thin, high velocity fluid film on the wafer surface; if the fluid film is sufficiently thin, the gap does not entirely fill with liquid, as shown in Fig. 1(d). This behavior is also undesirable. At each gap height, a series of simulations at progressively higher dispense velocities was used to delineate the transition from a completely filled gap to this impinging jet behavior, also shown in Fig. 1(b). The specification of this transition is more clear than the meniscus crawl transition as it is based entirely on whether an unfilled region persists adjacent to the dispense port.

The thickness and velocity of the fluid layer that is formed on the wafer can be approximately determined by balancing fluid momentum and shear in the thin film while enforcing that the mass carried by the film is consistent with the dispensed mass flow rate. If the fluid momentum is large enough then the thickness of the liquid film is less than the gap height and impinging jet behavior occurs. The transition into this behavior is therefore governed by a balance between fluid inertia and viscous shear which constitutes a Reynolds number (Re):

$$\mathrm{Re} = \frac{\rho v h}{\mu}, \tag{2}$$

where h is the gap height and μ is the fluid viscosity. Notice that this Reynolds number is proportional to the gap height, which is consistent with the results shown in Fig. 1(b) in that the transition velocity decreases as the gap height increases.

These two extreme cases bracket a range of acceptable dispensing velocities. An acceptable dispense velocity is characterized as being sufficiently high so that the meniscus immediately breaks directly under the dispense port but small enough so the entire gap is filled, resulting in the rapid and controlled fill process shown in Fig. 1(e). Figure 1(b) illustrates that the range of acceptable dispensing velocities narrows as the gap height increases.

Figure 2 overlays the CFD results onto lines associated

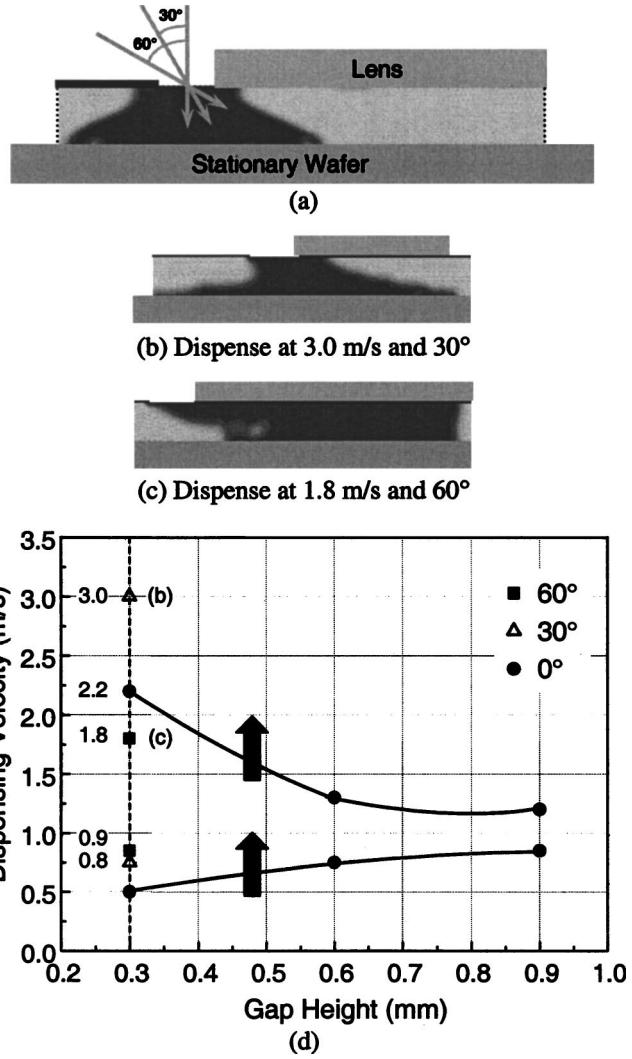


FIG. 3. (a) Schematic illustrating the different dispensing angles investigated, (b) the CFD simulation for a 30° dispense angle with a 3.0 m/s dispense velocity, (c) the CFD simulation for a 60° dispense angle with a 1.8 m/s dispense velocity, and (d) the effect of changing the dispensing angles and velocities on the optimum dispensing zone: the 0.3 mm gap (shown by the dotted line) was used to illustrate the fluid dispensing behavior for three dispensing angles, the velocities indicated on the graph shows the new locations for the transition zones (i.e., 0.8 and 0.9 are the meniscus crawl critical velocities for 30° and 60° angles). The arrows show the new potential location of the transition lines as the dispensing angle increases.

with a constant Weber number and Reynolds number. Notice that the transition from optimal dispense to impinging behavior is approximately consistent with a transition Reynolds number (Re_t) of 750 and the transition from optimal dispense to meniscus crawl behavior is approximately consistent with a transition Weber number (We_t) of 1. There is some discrepancy between the transitions associated with these critical values of Weber and Reynolds number and the transitions predicted by CFD, particularly at larger gap heights. These discrepancies can be partly explained by the fact that the transition between dispense behaviors is actually a gradual change rather than an abrupt one and therefore the determination of the critical dispense velocities that delineate these regions relies both on judgment and the resolution of the

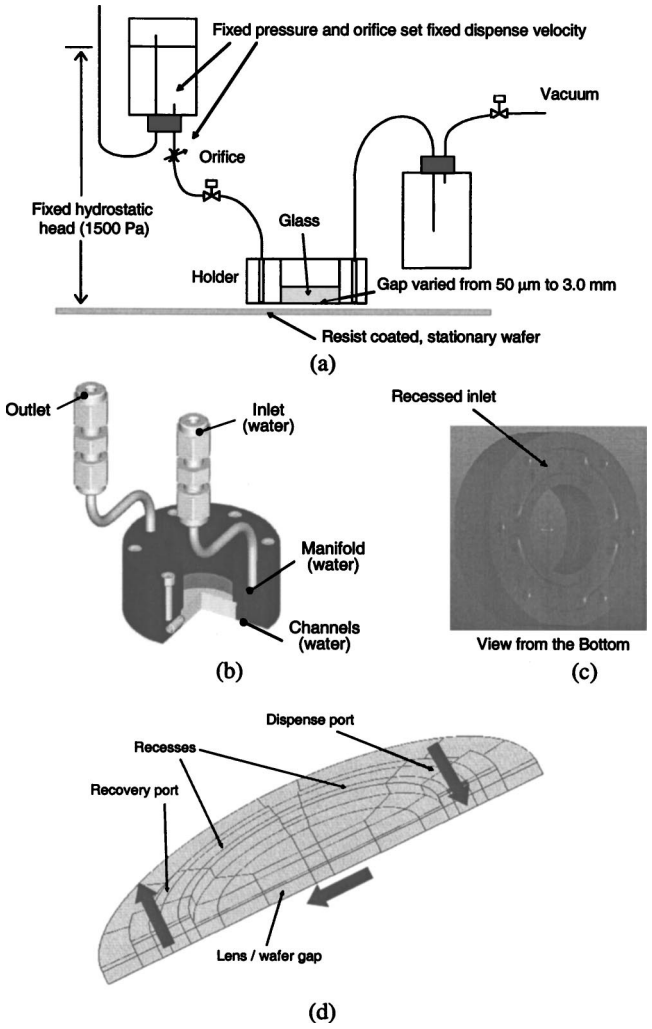


FIG. 4. (a) Schematic illustrating the details of the model verification experiments, (b) a close-up showing the fixture and glass window replicating the lens, (c) a schematic of the lens showing the recesses and the fluid dispense and recover ports, and (d) the CFD half-symmetry model developed to replicate the experiment.

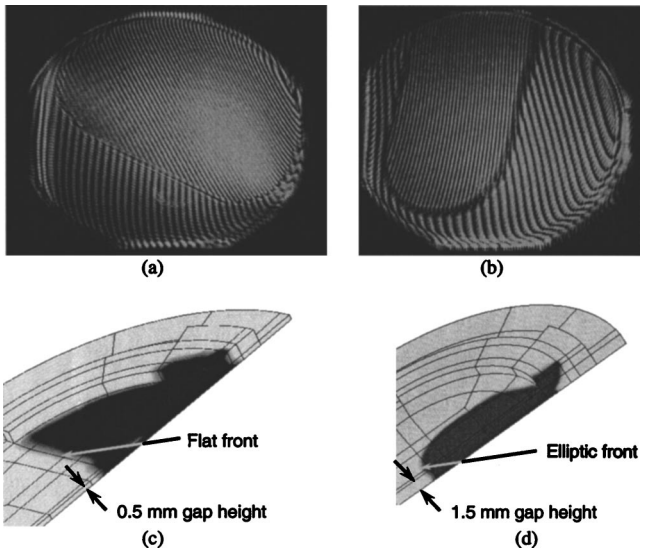


FIG. 5. Photographs showing the progress of the filling water front for (a) 0.5 mm gap, (b) 1.5 mm gap, and the corresponding CFD simulation results showing similar behavior for (c) 0.5 mm gap, (d) 1.5 mm gap.

velocities used in the CFD simulation. Also, the physics of the problem changes as the gap increases; for example, a larger gap allows the meniscus that forms on the lens to expand laterally by a greater amount prior to contacting the wafer and as a result the length scale associated with the Weber number is no longer directly related to the dispense port as given by Eq. (1).

Additional CFD simulations showed that the dispensing behavior is affected by the fluid dispense angle. Figure 3(a) illustrates the three dispensing angles investigated in this work, i.e., 0°, 30°, and 60°, with the positive angles indicating that the fluid is directed towards the lens. For a moderate dispense angle (e.g., 30°), the same regions of behavior exist but the transitions are shifted. The inertia of the fluid in the direction perpendicular to the wafer is reduced when the fluid is dispensed at an angle. Both the meniscus crawl and impinging jet transitions are proportional to this fluid inertia and therefore both lines shift upwards; a higher dispense velocity is required to yield the same inertia. This effect is shown Fig. 3(d) which contains the original points delineating the dispense process regimes for the 0° fluid dispense angle case as well as additional points for a gap height of 0.3 mm and fluid dispense angles of 30° and 60°. Note that for a 30° fluid dispense angle, indicated by the triangles in Fig. 3(d), both the meniscus crawl transition and impinging jet transition, shown in Fig. 3(b), are shifted to higher velocities. For a 60° dispense angle, the meniscus crawl transition again moves to a higher velocity. However, the impinging jet transition no longer occurs. Instead a new behavior is exhibited in which the fluid remains attached to the lens and tends to pull air into the gap, as shown in Fig. 3(c).

III. EXPERIMENTAL VERIFICATION

In order to verify some aspects of the CFD model, the simulated dispensing process was compared with an experiment in which a metered amount of water was used to fill a controlled lens-to-wafer gap. Figure 4(a) shows a schematic of the experiment. The progress of the filling process was monitored using a camera that was mounted over a glass window (which takes the place of the lens element). Figure 4(b) shows a cut-away view of the mounting fixture and indicates passages for the dispense and recovery of water. Figure 4(c) is a solid model of the bottom of the holder which shows the recesses in the structure used to manifold the water to the dispense and recovery ports.

A three-dimensional CFD model was developed with the goal of replicating the actual experimental geometry as closely as possible, as shown in Fig. 4(d). The details of the recessed dispense port, the recessed recovery port, and the gap can be identified in the CFD model. The experiment was carried out at the same flow rate but varying gap heights. Three of these experiments were simulated, including 0.5 mm, 1.0 mm, and 1.5 mm gap heights. Figures 5(a) and 5(b) show pictures of the fluid front that forms as the immersion

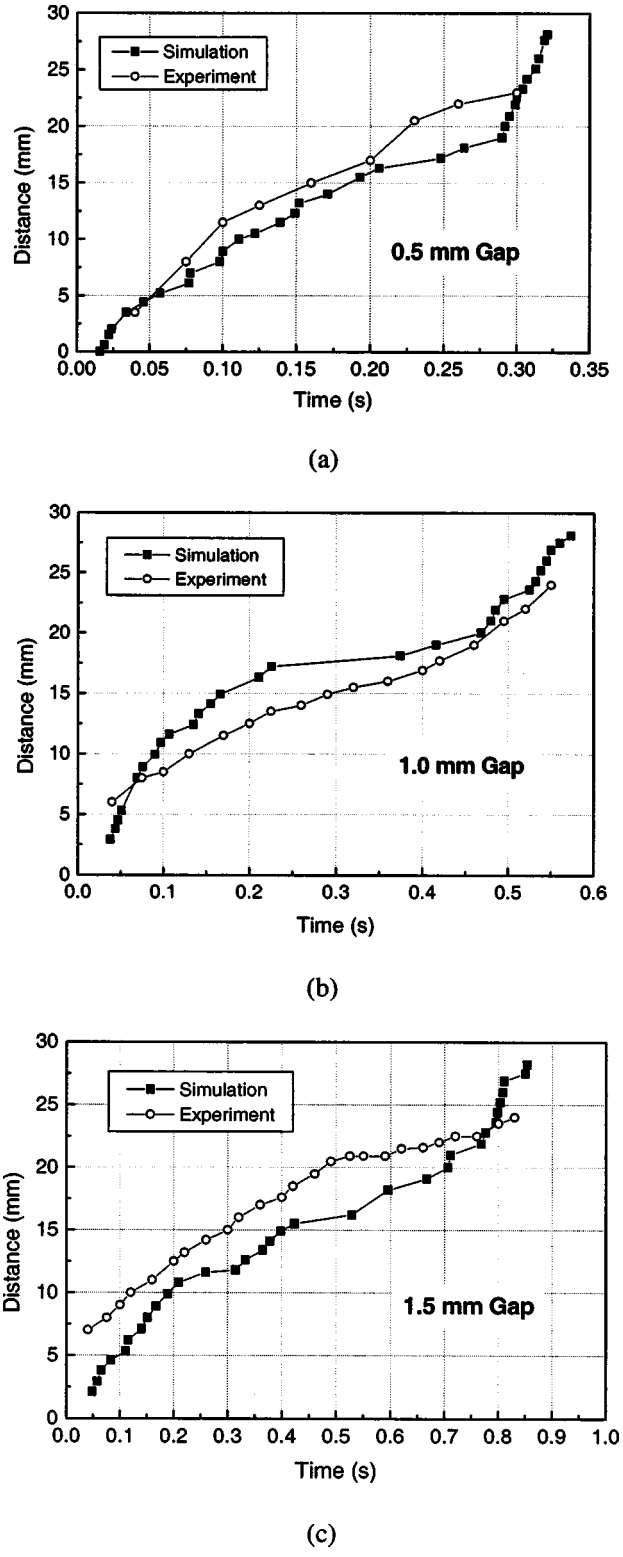


FIG. 6. Distance from the fluid front to the dispense port along the centerline of the glass window measured from experimental results and predicted by the CFD simulations for (a) 0.5 mm gap, (b) 1.0 mm gap, and (c) 1.5 mm gap.

fluid fills a 0.5 mm and 1.5 mm gap, respectively. Notice that the fluid front formed for the smaller gap is much flatter across its advancing edge than for the larger gap. This difference is related to whether the dispense flow rate is sufficient to fill the entire width of the gap at the rate that the leading edge of the meniscus is drawn through the gap by surface tension. The CFD simulations for the filling process for the same gaps, shown in Figs. 5(c) and 5(d), show similar behavior. Quantitative comparisons between the experiments and the CFD simulations are shown in Figs. 6(a), 6(b), and 6(c), where the distance between the water front to the dispense port measured along the center of the circular glass window is shown as a function of time relative to the initiation of the fill process for the 0.5 mm, 1.0 mm, and 1.5 mm gaps, respectively. The results from the CFD simulations and the experiments are in very good agreement.

IV. SUMMARY AND CONCLUSIONS

Two-dimensional CFD models were used to investigate the effects of fluid dispense velocity, gap height, and fluid dispense angle on the process of filling the lens-wafer gap prior to immersion lithography. The simulations revealed that an optimal region exists in the space of fluid dispense velocity and gap. The position of this region shifts as the fluid dispense angle is changed.

A three-dimensional CFD model was developed in order to replicate an experiment in which a controlled fill is monitored through a glass window. The simulation results agreed well with the experimental observations both quantitatively in terms of the progression of the fluid front with time and qualitatively in terms of the shape of the fluid front.

ACKNOWLEDGMENTS

The UW research was funded by DARPA/ARL and the Semiconductor Research Corporation (SRC). The UNM portion of the work was funded by International SEMATECH and by the ARO/MURI program in Deep Subwavelength Optical Nanolithography. Computer support was provided by the Intel Corporation and Microsoft.

[1]M. Switkes and M. Rothschild, J. Microlithog., Microfabr., Microsyst. **1**, 225 (2002).
[2]R. R. Kunz, M. Switkes, R. F. Sinta, J. E. Curtin, R. H. French, R. C. Wheland, C.-P. Chai Kao, M. P. Mawn, L. Lin, P. M. Wetmore, V. J. Krukonis, and K. Williams, J. Microlithog., Microfabr., Microsyst. **3**, 73 (2004).
[3]A. Wei, G. Nellis, A. Abdo, R. Engelstad, C.-F. Chen, M. Switkes, and M. Rothschild, J. Microlithog., Microfabr., Microsyst. **3**, 28 (2004).
[4]A. Wei, A. Abdo, G. Nellis, R. Engelstad, J. Chang, E. Lovell, and W. Beckman, J. Vac. Sci. Technol. B **21**, 2788 (2003).
[5]A. Wei, A. Abdo, G. Nellis, R. Engelstad, J. Chang, E. Lovell, and W. Beckman, Microelectron. Eng. **73-74**, 29 (2004).
[6]A. Wei, M. El-Morsi, G. Nellis, and R. Engelstad, J. Vac. Sci. Technol. B, these proceedings.

Scatterometry: a metrology for subwavelength surface-relief gratings

Petre C. Logof`tu* and John R. McNeil

Center of High Technology Materials, University of New Mexico, Albuquerque NM 87106

ABSTRACT

Phase-modulation scatterometry is a metrology technique for determining the parameters of gratings using as a key device a phase modulator. For measurement purposes the phase modulator requires a complicated calibration procedure that is analyzed here in detail. The main source of error to be dealt with are the fluctuations of the phase modulation amplitude. The measurables are the direct term and the first two harmonics of the output. For the fitting of the experimental data we used the ratio of the harmonics to the direct term because it improves significantly the accuracy. A sensitivity analysis was performed for two samples, one real and one theoretical, to find the measurement configuration that insures optimum determination precision for the grating parameters. For the real sample, comparison of the theoretical predictions for sensitivity with the actual values showed a good agreement. For both samples the sensitivity analysis indicated sub-nanometric precision for the critical dimension (grating linewidth).

Keywords: scatterometry, phase modulation, sensitivity

1. INTRODUCTION

In the ongoing effort to develop scatterometry as a sensitive metrology technique [1-6] we continued trying to improve it by implementing new measurement techniques. In this article the implementation of a phase-modulation technique (PMT) in scatterometry will be discussed. The phase modulation technique (PMT) was first applied for optical measurement by Jasperson et al. [7], and since then was continuously developed [8-10]. The technique was previously applied to scatterometry [10], but an analysis of the sensitivity and simplified synthetic formulae for the measurables, to our knowledge, are not available in the literature and they are thus an original contribution. Also the error analysis of the calibration from paragraph (4) is pushed to a deeper level than that of Acher et al [9]. It turns out that not all of the calibration configurations proposed by them can be applied, except in a modified form. A sensitivity analysis formally identical to the one performed for the ellipsometric-scatterometer in Reference [11], namely the application of the SAF formalism is done for the phase-modulation scatterometer.

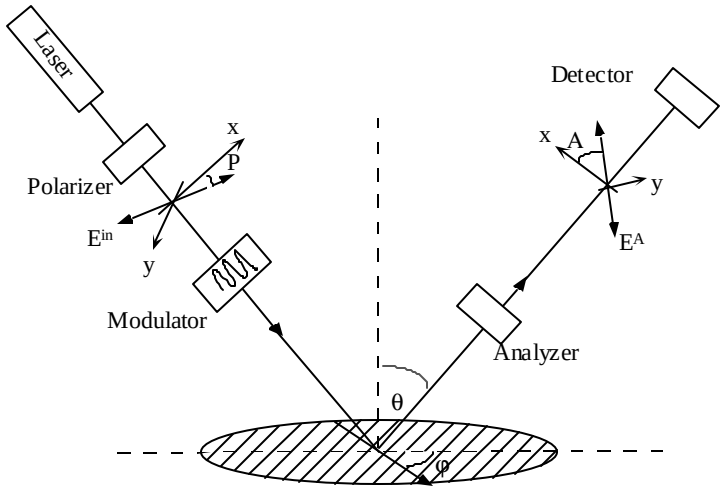


Fig. 1 The experimental arrangement of the phase-modulator scatterometer.

2. EXPERIMENTAL ARRANGEMENT

2.1. General view

The experimental arrangement that we call "phase-modulator scatterometer" is still an ellipsometric-scatterometer. Compared to the scatterometer described elsewhere [11], there is only one new device added in the system, the phase-modulator shown in Fig 1. This one device, however, radically modifies the nature of the experimental arrangement, as we shall see in the following.

The phase-modulator is basically a silica bar glued to a piezoelectric quartz bar transducer. The transducer is subjected to an oscillating voltage

* Correspondence: Email: logofatu@unm.edu;Telephone 505 272 7808; Fax 505 272 7801
Permanent address: National Institute of Laser, Plasma and Radiation Physics (INFLPR), Bucharest-M`gurele, PO-Box MG-36, 76900, Romania

with a frequency of 50 kHz. Because of its piezoelectric properties, the quartz bar then applies a pressure oscillating with the same frequency to the silica bar. The geometry of the homogenous silica bar changes, becomes anisotropic, and the bar behaves like a uniaxial crystal with dychroic properties. The result is that oscillation modes of the refractive index of the silica bar are excited, and oscillating phase retardation is introduced in the light that passes through the bar. Specifically the vector electric field component parallel to the longitudinal axis of the bar is retarded. Because the eccentricity of the polarization ellipse oscillates at the analyzer, the photodetector receives an oscillating signal. A lock-in amplifier measures the 50 kHz harmonics of the signal. These harmonics are the measurables of the experimental arrangement, and they are functions of the geometrical parameters of the sample.

For the phase-modulator scatterometer we need a photo-detector of large bandwidth, because the oscillation frequency of the modulator is 50 kHz. For the ellipsometric-scatterometer [5, 11], where the chopper frequency in use is about 200 Hz, there was no need for a large bandwidth photodetector. It is useful to note that there is such a large gap between the frequency of the chopper and that of the modulator that both can be used together in the experimental arrangement, because they do not practically interfere.

2.2. Alignment

An important experimental issue is to determine the location on the modulator surface where the amplitude of the standing stress wave is maximum, because the spatial variation is minimum there and we have uniform modulation amplitude. We can do this by monitoring the amplitude of the first or second harmonics of the signal when the light spot moves across the modulator surface. The first harmonic S_w varies approximately as $J_1(A)$, where A is the modulation amplitude, while the second harmonic S_{2w} as $J_2(A)$ (Bessel functions). The harmonics of the signal are defined in paragraph 3, where the purport of these considerations will be better understood. For values of A below 90°, J_1 and J_2 are monotonically increasing with A. The location where A is maximum is obviously the place where the harmonics S_w and S_{2w} have a maximum. Also, the beam spot area has to be small to insure the uniformity of the amplitude modulation.

For alignment purposes only, we will use as sample a phase retarder instead of a grating. The orientation angle of the retarder axis is γ and the phase retardation is Δ. The coordinate axes of the system are the axes of the retarder. In this case the I coefficients from Eqs. (9) have the more simple form

$$I_0 = \begin{bmatrix} (1-\cos 2\gamma\cos 2A)+\cos 2(P-M)\cos 2M(\cos 2A-\cos 2\gamma)+ \\ +\sin 2A\cos\Delta\cos 2(P-M)\sin 2\gamma\sin 2M \end{bmatrix}, \tag{1.a}$$

$$I_s = \sin 2(P-M)\sin 2A\sin 2\gamma\sin\Delta, \tag{1.b}$$

$$I_c = \sin 2(P-M)\left[(\cos 2\gamma-\cos 2A)\sin 2M+\sin 2A\cos 2M\sin 2\gamma\cos\Delta\right], \tag{1.c}$$

with P, A and M the orientation angles of the polarizer, analyzer and modulator respectively. We have the polarizer aligned parallel to the modulator when the first and the second I coefficients (1 b, c) are zero no matter what the position of the analyzer is. This happens because I_s and I_c are proportional to $sin2(P–M)$. We also know that the retardation axis is the longitudinal axis of the silica bar. We then set the polarizer so that $P–M=45°$. For a fixed value of $P–M$, we vary M and A until we obtain the extinction of I_s and I_c again. This happens when $M=A=0°$, because I_s is proportional to $sin2A$ and I_c has a term proportional to $sin2A$ and another proportional to $sin2M$. In this way we determine the zero position for the modulator and the analyzer.

3. MEASURABLES

From Fig. 1 we see that the transformation suffered by the beam in the phase-modulator scatterometer is described by

$$\vec{E}^{out} = \hat{A}\hat{r}\hat{M}\hat{P}\vec{E}^{in}, \tag{2}$$

where the Jones matrices for analyzer, sample and modulator are

$$\hat{A} = \begin{pmatrix} \cos^2 A & \cos A\sin A \\ \sin A\cos A & \sin^2 A \end{pmatrix}, \tag{3.a}$$

$$\hat{\boldsymbol{r}} = \begin{pmatrix} r_{\boldsymbol{pp}} & r_{ps} \\ r_{sp} & r_{ss} \end{pmatrix}, \tag{3.b}$$

$$\hat{M} = \begin{pmatrix} \cos M & -\sin M \\ \sin M & \cos M \end{pmatrix} \begin{pmatrix} \exp\left(-j\boldsymbol{d}(t)\right) & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos M & \sin M \\ -\sin M & \cos M \end{pmatrix}, \tag{3.c}$$

respectively, with

$$\boldsymbol{d}(t) = A \sin \boldsymbol{w}t\,, \tag{4}$$

which is the variable phase retardation introduced by the modulator in which A is the amplitude of the modulator. The matrix of the polarizer is simply the analyzer matrix with A substituted by P. $\vec{E}^{out}$ and $\vec{E}^{in}$ are the output and input vector electric fields, respectively. After making the calculations in (2) we obtain

$$r = E^{out} / E^{P} = \cos(P-M)(r_p \cos M + r_s \sin M) + \sin(P-M)(-r_p \sin M + r_s \cos M)\exp\left(-j\boldsymbol{d}(t)\right), \tag{5}$$

where we used the notations

$$r_p = r_{pp} \cos A + r_{sp} \sin A\,, \tag{6.a}$$

$$r_s = r_{ps} \cos A + r_{ss} \sin A\,, \tag{6.b}$$

and

$$\vec{E}^{out} = E^{out} \begin{pmatrix} \cos A \\ \sin A \end{pmatrix}, \tag{7.a}$$

$$\vec{E}^{P} = \left(E_p^{in} \cos P + E_s^{in} \sin P\right) \begin{pmatrix} \cos P \\ \sin P \end{pmatrix} = E^{P} \begin{pmatrix} \cos P \\ \sin P \end{pmatrix}. \tag{7.b}$$

The notations (6) were chosen for convenience reasons, but they have a certain physical meaning as well. Namely, r_p is the contribution of the p component of the beam incident on the sample to the output, and r_s is the similar contribution of the s component. Squaring the absolute value of r in Eq. (5) we obtain in terms of power

$$I(t) = I_0 + I_s \sin \boldsymbol{d}(t) + I_c \cos \boldsymbol{d}(t), \tag{8}$$

where

$$I_0 = |r_p|^2 + |r_s|^2 + \cos 2(P-M)\left[\cos 2M\left(|r_p|^2 - |r_s|^2\right) + 2 \sin 2M\ \mathrm{Re}\left(r_p r_s^*\right)\right], \tag{9.a}$$

$$I_s = 2 \sin 2(P-M)\,\mathrm{Im}\left(r_p r_s^*\right), \tag{9.b}$$

$$I_c = \sin 2(P-M)\left[-\sin 2M\left(|r_p|^2 - |r_s|^2\right) + 2 \cos 2M\ \mathrm{Re}\left(r_p r_s^*\right)\right]. \tag{9.c}$$

The calculations through which we obtained the functions I_0, I_s and I_c are very laborious and even more laborious is to obtain a simple, synthetic expressions of them as in Eqs. (9). However, there is a simple way to verify the accuracy of Eqs. (9). With the help of a programming environment for symbolic calculation as powerful as Mathematica, we can calculate the expanded form of the harmonics, which is in the range of pages of formulae. Then we program the computer to simplify the difference between the expanded forms of the harmonics and their synthetic expression (9). As expected, the result of the simplification is zero. One may notice that *I(t)* from Eq. (8) is not actually the square of *r*, but its double. Since the power is half of the square of the absolute value, it follows that we multiplied (8) with 4. The reason for this choice is that it normalizes the harmonics (9), their maximum absolute value becomes 1. The inspection of Eqs. (9) shows what he have advertised in the introduction of this chapter, the fact that the measurables depend on the signs of the phases. This dependence is insured by the fact that the formulae (9) contain both a term $\mathrm{Re}\left(r_p r_s^*\right)$ and a term $\mathrm{Im}\left(r_p r_s^*\right)$, the cosine and the sine of the phase of the complex term $r_p r_s^*$.

Since in Eq. (8) we have to deal with trigonometric functions with the retardation δ from (4) as argument, it is useful to write the expansion of sine and cosine of δ in Bessel functions

$$\sin \delta(t) = 2J_1(A)\sin \omega t + 2J_3(A)\sin 3\omega t + \ldots \textit{higher harmonics}\,, \tag{10.a}$$

$$\cos \delta(t) = J_0(A) + 2J_2(A)\sin 2\omega t + \ldots \textit{higher harmonics}\,. \tag{10.b}$$

It follows that the harmonics of the output signal are

$$\begin{pmatrix} S_0 \\ S_\omega \\ S_{2\omega} \end{pmatrix} = \begin{pmatrix} 1 & 0 & J_0(A) \\ 0 & 2J_1(A) & 0 \\ 0 & 0 & 2J_2(A) \end{pmatrix} \begin{pmatrix} I_0 \\ I_s \\ I_c \end{pmatrix}. \tag{11}$$

The *S* coefficients of Eq. (11) are what can be measured with the phase-modulator scatterometer in the ideal case. In practice we set the modulation amplitude A so that we have

$$A = A_0 = 137.8^\circ\,, \tag{12.a}$$

which nullifies the Bessel function of zeroth order,

$$J_0(A_0) = 0\,. \tag{12.b}$$

For this value of the modulation amplitude, the measurables *S* are proportional to the *I* quantities. In practice, however, we prefer to work with the quantities

$$R_\omega = \frac{S_\omega}{2J_1(A_0)S_0}\,, \tag{13.a}$$

$$R_{2\omega} = \frac{S_{2\omega}}{2J_2(A_0)S_0}\,. \tag{13.b}$$

There are some obvious advantages of working with the quantities (13) besides the fact that they are normalized. The numerator and the denominator of (13) are measured simultaneously and therefore their ratio is free from systematic errors caused by the spatial displacement of the beam spot on the photodetector area and by fluctuations in time of the laser power and other variables.

Note also that, with the phase-modulation scatterometer, we extract more information than in the case of the ellipsometric-scatterometer because we measure the two quantities (13.a, b) simultaneously. Therefore, the information provided by the phase-modulation scatterometer is richer than that provided by the ellipsometric-scatterometer under two aspects: qualitatively, because the measurements depend on the signs of the phases, and quantitatively, because we take double the number of measurements in a θ–scan.

4. CALIBRATION

Unlike the ellipsometric-scatterometer, the calibration of the phase-modulation scatterometer can be quite a challenge. New problems arise, for instance, because the modulation amplitude A is not constant in time. This is a significantly important new source of errors that we must minimize. For this reason the calibration of the modulation amplitude performed as indicated in Eqs. (12) must be done repeatedly during the experiment. There is a better way to maintain constant the modulation amplitude, by monitoring the ratio of the third and first harmonic [9], but our present experimental arrangement does not yet support this technique.

Another problem is that in practice we must deal with slightly different equations than those corresponding to the ideal case shown in Eq. (11). Namely new non-diagonal elements of the matrix on the right hand side of Eq. (11) become non-zero, and they are too large to be neglected (around few percents). The reason for this non-ideal behavior is still in dispute. According to Jasperson et al. the reason is that the modulator has a residual constant birefringence [7]. Acher et al., on the other hand,

hypothesize that higher vibration modes (i.e. $2w, 3w$ etc. modes) are excited by the transducer in the silica bar resulting in a modulation of the form [9]

$$d(t) = d_0 + A_1 \sin(wt) + A_2 \sin(2wt) + A_3 \sin(3wt) + \dots, \tag{14}$$

where the amplitudes A_1 A_2 A_3, etc. may be complex. A very important point made by Acher et al. is that no matter how many higher modes from (14) perturb the modulation and hence the coefficients of the transmission matrix in (11), the relation between the *S* and *I* coefficients is linear. They show then that Eq. (11) should be rewritten as

$$\begin{pmatrix} S_0 \\ S'_w \\ S'_{2w} \end{pmatrix} = \begin{pmatrix} 1 & C_{s,0} & J_0(A) + C_{c,0} \\ 0 & T_1 & C_{c,w} \\ 0 & C_{s,2w} & T_2 \end{pmatrix} \begin{pmatrix} I_0 \\ I_s \\ I_c \end{pmatrix}, \tag{15}$$

where

$$S'_w = \frac{S_w}{2J_1(A)}, \tag{16.a}$$

$$S'_{2w} = \frac{S_{2w}}{2J_2(A)}. \tag{16.b}$$

The terms T_1 and T_2 are called transmission coefficients and they are approximately 1. The coefficients $C_{c,0}$, $C_{s,0}$, $C_{c,w}$ and $C_{s,2w}$ are called coupling coefficients, and their absolute values are small compared to 1. The ensemble of the coefficients introduced in Eq. (15) will be called adjustment coefficients. Now we have to set the amplitude A to a different value than 137.8°, because of the perturbation introduced by C_{c0}. After setting A so that $J_0(A)+C_{c,0}$ is as close as possible to 0 we have only to determine T_1, T_2, $C_{s,0}$, $C_{c,w}$ and $C_{s,2w}$.

Table 1 Calibration configurations for the phase modulator.

Configuration label	P–M (°)	M (°)	phase retarder	A (°)
a)	±45	0 (90)	no	±45
b)	±45	±45	yes	±45
c)	±45	0 (90)	yes	±45

All these operations can be made with the help of the measurement configurations described in Table 1. Instead of a sample, a quarter-wave phase retarder or nothing are alternately inserted in the experimental arrangement. The *I* coefficients (9) take the form (1). The angles *M* and *A* are measured with respect to the axes of the retarder. For configuration a) the absolute values of the angles do not matter, but only their difference, and we do not need to set reference axes. The void that is in place of the sample in configuration a) can be modeled as a retarder with arbitrary orientation y and with the retardation $D=0°$. The ratios (13) in the non-ideal case become

$$R_w = T_1 I_s + C_{c,w} I_c - C_{s,0} I_s^2 - J_0 I_s I_c, \tag{17.a}$$

$$R_{2w} = T_2 I_c + C_{s,2w} I_s - C_{s,0} I_s I_c - J_0 I_c^2, \tag{17.b}$$

where we noted

$$J_0 = J_0(A) + C_{c,0}. \tag{18}$$

In Eqs. (17) we kept only the first order of the expansion, taking into account that the transmission coefficients are close to 1 and the coupling coefficients are close to 0.

We will particularize Eqs. (9) and (17) for each configuration. Since the coupling coefficients are very small, we have to take into account the first order expansion of the uncertainties of the angles *P, A, M,* γ and Δ, because they are of the same order as the coupling coefficients, and they can infringe in the validity of the calibration.

Configuration a)

Eqs. (9) become

$$I_0 = 1, \tag{19.a}$$

$$I_s = 0, \tag{19.b}$$

$$I_c = \sin 2(P-M)\sin 2(A-M) \cong \pm 1. \tag{19.c}$$

Then the quantities (17) become

$$R_w = C_{c,w} I_c, \tag{20.a}$$

$$R_{2w} = T_2 I_c - J_0 I_c^2. \tag{20.b}$$

Changing *P* or *A* with ±90° changes the sign of I_c. The modulation amplitude is adjusted so that the modulus of R_{2w} remains unchanged for ±90° rotation of the polarizer or the analyzer. Fig. 2 illustrates how R_{2w} varies in two configurations of type a) when the modulation amplitude is varied and how the curves intersect at A_0.

One may notice that the configuration in which the polarizer and the analyzer are parallel is more sensitive to variations of the modulation amplitude. It follows that for routine recalibrations of A, after T_2 has been determined, it is better to use this configuration rather than the one in which the polarizer and the analyzer are perpendicular. All we have to do then for a routine recalibration is to adjust A so that R_{2w} corresponding to the configuration *P=45, A=45* reaches the closest value to R_{2w} corresponding to the configuration *P=45, A=–45,* which we may consider constant because its insensitiveness to changes of A.

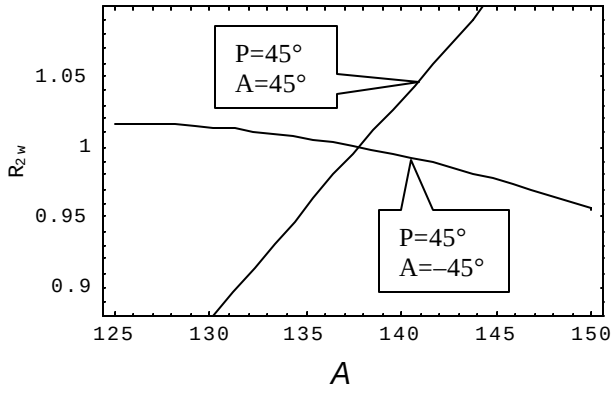


Fig. 2 Illustration of setting the modulation amplitude so that $J_0=0$.

After we nullify J_0 we can easily measure T_2 and $C_{c,w}$ using Eqs. (20). Taking into account the fact that a rotation of *P* or *A* with ±90° changes the sign of I_c, we obtain

$$T_2 = \left| R_{2w}^1 - R_{2w}^2 \right| / 2, \tag{21.a}$$

$$C_{c,w} = \mathrm{sgn}\left(R_{2w}^{1,2}\right) R_w^{1,2}. \tag{21.b}$$

Here the superscripts *1, 2* refer to two cases of configuration a), one with the polarizer or the analyzer rotated with ±90° in rapport with the other.

Configuration c)

We have (9) yielding

$$I_0 = 1, \tag{22.a}$$

$$I_s = \sin 2(P-M)\sin 2A \cong \pm 1, \tag{22.b}$$

$$I_c = \sin 2(P-M)\sin 2A \cos 2M\varepsilon_\Delta \cong 0. \tag{22.c}$$

Here ε_Δ is the uncertainty of the retardation Δ. Thus (17) becomes

$$R_w = T_1 I_s - C_{s,0} I_s^2, \tag{23.a}$$

$$R_{2w} = T_2 I_c + C_{s,2w} I_s \,. \tag{23.b}$$

Changing P or A by $\pm 90°$ changes the signs of both I_s and I_c. We want to change the sign of I_s in order to separate the term with T_1 from the term with $C_{s,0}$ in Eq. (23 a). We also want to separate the term with T_2 from the term with $C_{s,2w}$ from (23.b). Hence we need a transformation that changes the sign of I_s but not that of I_c. This transformation is the rotation of the phase retarder by $\pm 90°$, which means the coordinate axes are rotated with the same amount. This is equivalent to simultaneously rotating P, A and M by $\mp 90°$. Therefore we have

$$T_1 = \left| R_w^1 - R_w^2 \right| / 2 \,, \tag{24.a}$$

$$C_{s,0} = \left(R_w^1 + R_w^2 \right) / 2 \,, \tag{24.b}$$

$$C_{s,2w} = \mathrm{sgn}\left(R_w^1 \right) \left(R_{2w}^1 - R_{2w}^2 \right) / 2 \,, \tag{24.c}$$

where the superscripts correspond to two cases of configuration c), one having the phase retarder rotated with $\pm 90°$ in rapport with the other.

Configuration b)

Although Acher et al. [9] recommend configuration b) for calibration purposes, we rejected it because it cannot provide an accurate determination. Indeed, for configuration b) (9) gives

$$I_0 = 1 \,, \tag{25.a}$$

$$I_s = \sin 2(P - M) \sin 2A \,, \tag{25.b}$$

$$I_c = \sin 2(P - M) \sin 2M \left(\varepsilon_y - \varepsilon_A \right), \tag{25.c}$$

where ε_y and ε_A are the uncertainties in the orientation angle of the phase retarder and the analyzer respectively. Eqs. (17) thus become Eqs. (23), the same as for configuration c). As in the case of configuration c), we need a transformation that changes the sign of I_s but not that of I_c. We cannot move the analyzer or the phase retarder, because the uncertainties ε_y and ε_A would change uncontrollably, prohibiting the use of Eq. (25.c). Rotating the polarizer does not perform the transformation we need. Only rotating the modulator provides the desired transformation. In our experimental system the modulator cannot be precisely rotated. Therefore we did not use configuration b). The transformation used by Acher et al. was the rotation of the analyzer or the polarizer, not of the modulator, which we showed above that does not provide an accurate calibration.

Table 2 The adjustment coefficients for the modulator.

Adjustment coefficients	**T_1**	**T_2**	**$C_{s,0}$ (%)**	**$C_{c,w}$ (%)**	**$C_{s,2w}$ (%)**
Value	0.9962	0.9934	–0.03	–1.79	1.69
Precision	0.0008	0.0007	0.04	0.25	0.32

After performing several calibrations of our modulator we obtained the average values for the adjustment coefficients displayed in Table 2. One comment that can be made about this table is that it is consistent with the hypothesis of Jasperson et al. [7] about the source of the adjustment coefficients, rather than the hypothesis of Acher et al. [9]. Namely, Table 2 shows evidence favoring the hypothesis that the source of errors is a constant residual birefringence of the silica bar. Indeed, if Eq. (14) reduces to

$$\delta(t) = \delta_0 + A \sin(\omega t), \tag{26}$$

then (15) becomes

$$\begin{pmatrix} S_0 \\ S_w' \\ S_{2w}' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -\delta_0 \\ 0 & \delta_0 & 1 \end{pmatrix} \begin{pmatrix} I_0 \\ I_s \\ I_c \end{pmatrix}. \tag{27}$$

This means we have

$$C_{s,0} = 0 \,, \tag{28.a}$$

$$C_{s,2w} = -C_{c,w} = \delta_0 \,, \tag{28.b}$$

which is supported by the data shown in Table 2. We can even calculate δ_0 from Table 2 and obtain that it is equal to 1.00°. However, other characteristics of our system fit neither the analysis of Jasperson nor of Acher. For instance our δ_0 is not a residual birefringence; it disappears when the modulator is not excited. When the modulator is unexcited, it behaves like simple, neutral glass. Moreover, δ_0 is not even constant, it varies in time. This observations lead us to the conclusion that δ_0 is caused by imperfections of the electronic excitation system, in particular of the piezoelectric quartz.

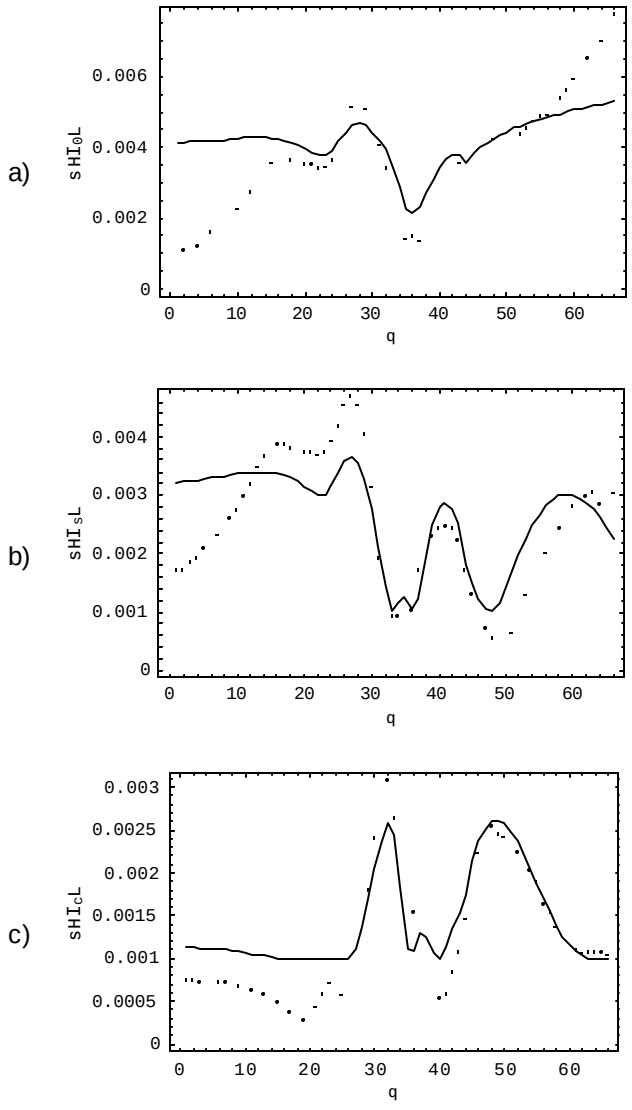


Fig. 3. Empirical error function for the phase-modulator scatterometer: a) I_0: $\sigma^2 = 10^6 + 3.2 \times 10^4\ I_0$, b) I_s: $\sigma^2 = 10^6 + 2.0 \times 10^4\ I_s$ c) I_c: $\sigma^2 = 10^6 + 1.4 \times 10^4\ I_c$. On the right hand of each figure are shown the coefficients for the best fit of the errors. The measurements were performed using the "XL6" sample at ϕ=0° and λ=632.8 nm.

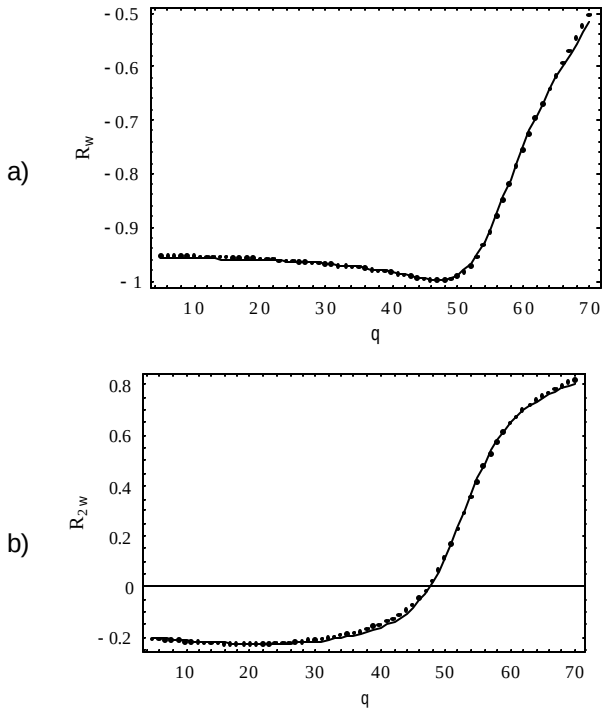


Fig. 4. Example of fitting data from the phase-modulator scatterometer for sample "XL6" at ϕ=90°and λ=632.8 nm. The results of the fitting were lw=1493.1±1.2 Å, h_q=5568.4±3.3 Å, h_1=1604.3±1.9 Å.

It must be noted that the calibration procedures are very sensitive to the quality of the polarizers. Polymeric polarizers transform the incident beam into an elliptically polarized beam with large eccentricity rather than into a perfect linearly polarized beam. For this reason, to perform the calibration one needs better polarizers, such as Glan-Thompson prism polarizers, with a very high extinction ratio.

5. SYMMETRY PROPERTIES AND SENSITIVITY CONSIDERATIONS

The application of the SAF formalism can be a very time-consuming procedure. In order to avoid burdening the computer with unnecessary sensitivity calculations, we will try to reduce the domain of the degrees of freedom of the experimental arrangement [11, 12]. Unlike the ellipsometric-scatterometer, for the phase-modulator scatterometer we can make some *a priori* sensitivity considerations, independent of the sample, which will further reduce the domain of experimental parameters that we have to scan in the search of the optimum sensitivity.

At first sight it seems we have four degrees of freedom for the phase-modulator scatterometer: P, A, M and f. The measurables of the system, the ratios R_w, R_{2w} from (13), are approximately equal to the ratios I_s/I_0 and I_c/I_0. A simple inspection of the I coefficients from Eqs. (9) shows that these ratios are maximum when $P-M=\pm 45°$. For improving the signal to noise ratio factor we need to set the position of the polarizer relative to the modulator to either +45° or –45° (i.e. $P-M=\pm 45°$).

Changing M by ±90° changes the sign of I_s and I_c but not their amplitude, while I_0 remains unchanged. Thus the quantities R_w and R_{2w} changes only their the sign, which is inconsequential for the sensitivity. Therefore we can restrict the sensitivity scan to a domain of $M \in (0°\ldots 90°)$.

6. THE EMPIRICAL ERROR FUNCTION

The SAF formalism requires an extrapolation of the measurement errors obtained is some particular measurement configurations to an empirical error function valid for all the measurement configurations. The empirical error function for the phase-modulator scatterometer was evaluated in the same way it was evaluated for the ellipsometric-scatterometer [6]. In Fig. 3 it is illustrated how we fit the averaged errors for a set of one hundred measurements with the empirical function. The function that best approximates the measurement errors is

$$\sigma^2 = 10^{-6} + 5.0\times 10^{-4} I^2 . \tag{29}$$

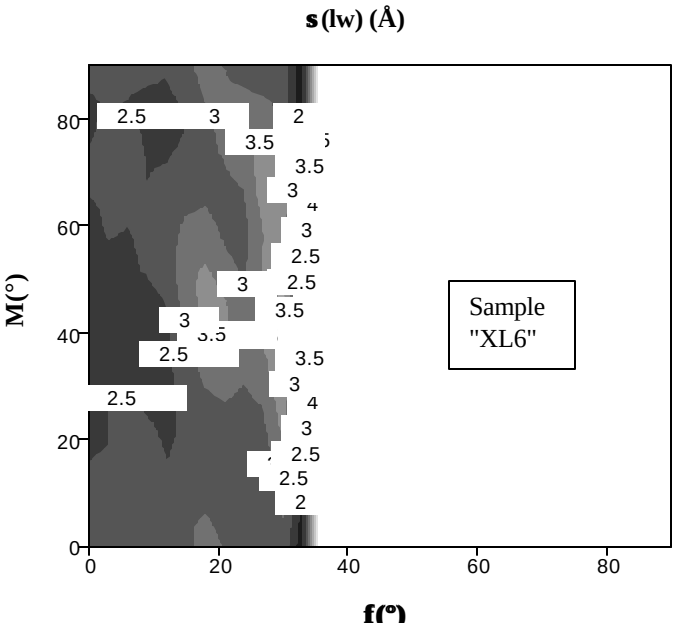


Fig. 5. Sensitivity map from the SAF scan for the phase-modulator scatterometer, sample XL6, at λ=632.8 nm. Here the color code is as follow: deeper the hue, smaller σ.

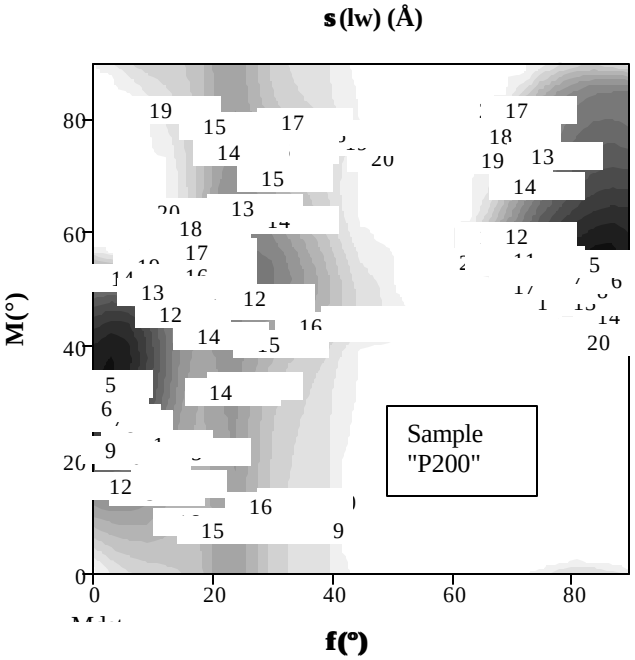


Fig. 6. Sensitivity map from the SAF scan for the phase-modulator scatterometer, sample P200, at λ=632.8 nm. The color code is as in Fig. 5.

7. SAF PREDICTIONS AND COMPARISON WITH EXPERIMENT

The SAF formalism [11] was applied to the samples "XL6" and "P200" presented in Table 3, for the measurables from (13). In Fig. 4 an example is shown of fitting data generated by the phase-modulation scatterometer to theoretical curves calculated with RCWA. In Fig. 5 the sensitivity map is illustrated for sample "XL6", using the configurations with $P=A=45°$, among which there is the optimum configuration. For the time being we are restricted by the fact that we cannot precisely change the orientation angle of the modulator *M*. Additionally, because of the property shown in Fig. 2, we have to maintain the analyzer orientation parallel to that of the polarizer, $P=A=45°$. For this reason we cannot yet provide experimental data for some measurement configurations to compare with the theoretical predictions (for instance configuration 5 in Table 3). In Fig. 6 the sensitivity map for sample "P200" is shown for configurations type $P=-A=45°$. The code color for both Figs. 5 and 6 is the same, namely the deeper the hue of gray, the smaller σ.

Table 3. Geometrical and optical parameters of the samples investigated. The subscripts "g" and "s" refer to the grating and the substrate, respectively. The layers from the grating to the substrate have numerical subscripts starting with 1. Sample "XL6" is real, and sample "P200" is theoretical.

Sample	L (Å)	lw (Å)	h_g (Å)	h_1 (Å)	n_g	n_1	n_s
XL6	3628.8	1470	5700	1600	1.64	1.66	3.882 – 0.019 i
P200	2000	700	4000	–	1.64	–	3.882 – 0.019 i

Table 4. Numerical results of the sensitivity analysis of sample "XL6" for phase-modulator type measurables.

Label	Measurement configuration				lw (experiment)	s(lw) (Å) (theory)	s(lw) (Å) (experiment)
	f(°)	M(°)	P(°)	A(°)			
1	0	90	45	45	1467.2	4.1	2.1
2	45	90	45	45	1419.0	40.2	29
3	90	90	45	45	1507.4	35	2.9
4	21	90	45	45	1464.8	5.0	4.4
5	6	37	45	–45	–	3.0	–

Table 5. Numerical results of the sensitivity analysis of sample "P200" for phase-modulator type measurables.

Label	Measurement configuration				s(lw) (Å) (theory)
	f(°)	M(°)	P(°)	A(°)	
1	3	36	45	–45	4.8
2	87	56	45	–45	4.7
3	90	90	45	45	18.8
4	24	90	45	–45	14.1

In Table 4 the numerical results of the sensitivity analysis performed on the sample "XL6" are summarized. The first four configurations were used for computing the empirical error function discussed in paragraph 6. We tested experimentally these four configurations. Configuration 5, deemed by SAF to be optimum could not be tested. We notice a qualitative agreement between theory and experiment, except for configuration 3. The reason for this discrepancy is because the measurement precision for this configuration was inexplicably better than the measurement precision for the other configurations. The

cause of this (fortunate) anomaly is yet to be determined. In Table 5 the numerical results of the sensitivity analysis for the theoretical sample "P200" are shown.

The inspection of Tables 4 and 5 does not reveal a marked improvement of the sensitivity compared to the ellipsometric-scatterometer. It is not even clear if there is any improvement, although the phase-modulation scatterometer provides more information, both quantitatively and qualitatively, than the ellipsometric-scatterometer. The reason was ascribed to the additional sources of errors introduced by a more complicated experimental arrangement. Specifically the instability of the phase modulation is an important perturbation factor. This assertion is proved by the fact that the empirical error function (29) is considerably higher than for the ellipsometric scatterometer [6].

There is, however, an important aspect in Table 4 that makes our effort of developing a phase-modulator scatterometer worthwhile. Namely, we notice what we believe is an increased accuracy. Actually there is no reason to doubt the absolute accuracy of these results since the dispersion of the experimentally determined values for the linewidth are almost within the theoretical uncertainties. If we ignore the insensitive configurations 2 and 3, we have a both precise and accurate determination of the linewidth, which is clearly bound to lie within the interval 1460-1470 Å. The perceived accuracy is the most important improvement shown by the phase-modulation scatterometry technique, compared to previous scatterometry techniques. This is attributed to the fact that the measurables are real-time ratios. The numerator and the denominator are simultaneously measured, in exactly the same area of the photodetector, unlike the ellipsometric-scatterometer, thus allowing many systematic errors to be eliminated. Contrary to the ellipsometric-scatterometer, the normalization measurement I_0 is taken in the same conditions as I_s and I_c. The increased accuracy that is associated with the phase-modulation scatterometer leads us to the conclusion that in the systematic errors manifest in Table 4, the external causes outweigh the internal ones.

8. CONCLUSION

Scatterometry has been a valuable metrology tool to characterize structures with CD's as small as 100 nm and smaller. To compensate for the loss of sensitivity due to the miniaturization trend in microelectronics, we have designed a methodology for finding the optimum measurement configuration. The technique exploits experimental parameters that have optimum sensitivity to changes of the spatial parameters of the grating. This method, which we introduced previously [6, 11], is termed "sensitivity analysis for fitting" (SAF), and it is a mathematical algorithm that can be applied for optimization purposes regardless of the specific type of measurements that are performed.

In the work discussed here, we applied SAF to a new experimental arrangement that has an element added to the classical experimental arrangement - a phase modulator. This element acts to provide more information regarding grating properties. Additionally, the modulator allows three signals (harmonics of the phase modulation signal) to be measured simultaneously with lock-in amplifiers, which means that the speed of acquiring information is increased. By normalizing the two superior harmonics to the fundamental harmonic (DC term), we eliminate systematic errors, such as the spatial variations of the sensitivity of the photodetector and laser fluctuations. However, phase modulation scatterometry has some disadvantages. It is more sensitive to residual birefringence of the optical elements than classical scatterometry. Because of residual birefringence and higher modulation orders induced by the modulator piezoelectric transducer, correction terms need to be applied to the measurements. These correction terms have to be calculated prior to any data acquisition, and there is not an easy and accurate way to do this initial calibration automatically. Additionally, the precision of the phase modulation impacts the precision of sample parameter measurements.

Carefully weighing the pros and cons, we will illustrate that by joining the phase modulator with a mathematical optimization algorithm (SAF) and classical scatterometry, we obtain an improved method that yields superior results. The technique is able to satisfy tomorrow's metrology needs of the microelectronics industry.

REFERENCES

1. J. R. McNeil, S. S. H. Naqvi, S. M. Gaspar, K. C. Hickman, S. R. Wilson, in *Encyclopedia of Materials Characterization* C. A. E. C. Richard Brundle, Shaun Wilson, Ed. (Manning Publishing Co., 1992).

2. S. S. H. Naqvi, J. R. McNeil, R. H. Krukar, Z. R. Hatab, "Grating parameter estimation using scatterometry," *Miniature and Micro-Optics and Micromechanics, Proc. SPIE 1992*, pp. 170, 1993.

3. M. R. Murnane, C. J. Raymond, S. S. H. Naqvi, J. R. McNeil, "Subwavelength photoresist grating metrology using scatterometry," *Proc. SPIE 2532*, pp. 251-261, 1995.

4. B. K. Minhas, S. A. Coulombe, S. S. H. Naqvi, J. R. McNeil, "Ellipsometric-Scatterometry for the metrology of sub-0.1 μm linewidth structures", *Appl. Opt.* **37**(22), pp. 5112-5115, 1998.

5. S. A. Coulombe, P. C. Logofatu, B. K. Minhas, S. S. H. Naqvi, J. R. McNeil, "Ellipsometric-Scatterometry for sub-0.1 mm CD measurements," *Microlithography, Proc. SPIE 3332*, pp. 282-293, 1998.

6. P. C. Logofatu, J. R. McNeil, "Sensitivity analysis of fitting for scatterometry," *Microlithography XIII, Proc. SPIE 3677*, pp. 177-183, 1999.

7. S. N. Jasperson, S. E. Schnatterly, "An improved method for high reflectivity ellipsometry based on a new polarization modulation technique", *Rev. Sci. Instrum* **40**(6), pp. 761-767, 1969.

8. T. C. Oakberg, "Modulated interference effects: use of photoelastic modulators with lasers", *Optical Engineering* **34**(6), pp. 1545-1550, 1995.

9. O. Acher, E. Bigan, B. Drevillon, "Improvements of phase modulated ellipsometry", *Rev. Sci Instrum.* **60**(1), pp. 65-77, 1989.

10. D. W. Mills, R. L. Allen, "Spectral ellipsometry on patterned wafers," 1995.

11. P. C. Logofatu, dissertation, University of New Mexico (2000).

Measurement precision of optical scatterometry

Petre C. Logofătu* and John R. McNeil

Center of High Technology Materials, University of New Mexico, Albuquerque NM 87106

ABSTRACT

In the work reported here, we discuss the measurement precision of two scatterometry techniques, the variable angle and the variable wavelength techniques. The issue of interest is the measurement precision of the sample parameters. This is determined by both the sensitivity of the diffraction measurable to changes in sample parameters and the precision with which the measurable can be determined. This approach includes taking into account the correlation effect between the contribution to the measurable of the various grating parameters to be determined, such as linewidth and height. The comparison of the theoretical predictions of precision for angle-resolved and wavelength-resolved scatterometer measurements shows no conclusive hierarchy. Practical considerations, however, indicate that angular-resolved scatterometry is a more advantageous technique. For both methods, decreasing the wavelength of the light source improves the determination precision of the sample parameters.

Keywords: scatterometry, ellipsometry, error analysis, spectroscopy, sensitivity, measurement precision

1. INTRODUCTION

Scatterometry is a method of inferring the shape of microstructures from how these structures affect light in reflection[1-3]. The technique represents a metrology tool for determining the geometrical parameters of microelectronics test structures from measurements of the light diffracted by these structures. For clarity, suppose the measurable is a quantity q (generally the reflectance) which is a function of the sample geometrical parameters that we want to determine $\vec{p}^{\,geo}$ (e.g. linewidth, grating height, underlying layer heights, etc.), the sample optical parameters that we know *a priori* $\vec{p}^{\,opt}$ (refractive indices of the materials and grating pitch), and the parameters of the experimental arrangement that we control $\vec{p}^{\,exp}$ (azimuth angle of the sample, incidence angle and wavelength of the beam, and input and output polarization); or

$$q = q\left(\vec{p}^{\,geo}, \vec{p}^{\,opt}, \vec{p}^{\,exp}\right). \tag{1}$$

We can generate theoretical or experimental curves of q versus $\vec{p}_j^{\,exp}$ by varying the experimental parameter $\vec{p}_j^{\,exp}$. The choice of this variable experimental parameter defines a scatterometry technique. If the choice is the incidence angle, we have angle-resolved scatterometry[3-5]. If the choice is the wavelength, we have spectroscopic scatterometry[6].

It is generally recognized that the most important aspect of a metrology technique is the measurement precision of the quantities to be determined. In this paper the two scatterometry techniques mentioned above will be analyzed and compared primarily with this in mind.

2. MEASUREMENT PRECISION OF THE SAMPLE PARAMETERS

A rigorous definition of the measurement precision for scatterometry is expressed by taking into account the determination procedure of the sample parameters. Namely, they are obtained by fitting theoretically generated curves to experimental curves. The most probable values for the sample parameters are those parameters corresponding to the curve which has the minimum weighted mean square difference with the experimental curve, or

* Correspondence: Email: logofatu@unm.edu;Telephone 505 272 7808; Fax 505 272 7801
Permanent address: National Institute of Laser, Plasma and Radiation Physics (INFLPR), Bucharest-Măgurele, PO-Box MG-36, 76900, Romania

$$\chi^2 = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left(q_i^{teo}\left(\vec{p}^{\,geo}\right) - q_i^{exp} \right)^2 , \tag{2}$$

where σ_i is the measurement precision of q_i. This fitting method, also called least squares fit, yields the following determination precision of the parameters $\vec{p}_j^{\,geo}$ [1, 7]

$$\sigma\left(p_j^{geo}\right) = \sqrt{\alpha_{jj}^{-1}} , \tag{3}$$

where α is the curvature matrix[7]

$$\alpha_{k,l} = \frac{1}{2} \frac{\partial^2 \chi^2}{\partial p_k^{geo} \partial p_l^{geo}} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \frac{\partial q_i}{\partial p_k^{geo}} \frac{\partial q_i}{\partial p_l^{geo}} . \tag{4}$$

Here j, k and l range from 1 to M, the number of parameters to be determined. M is the dimension of the vector $\vec{p}_j^{\,geo}$ and of the matrices α and α'. We observe that for the rigorous estimation of (3), one needs both the measurement precision of the measurables and the sensitivity of the measurables q_i with respect to changes of the sample parameters $\vec{p}^{\,geo}$, both of which are expressed in (4). We also note for later references that the curvature matrix is symmetric, i.e. $\alpha_{12}=\alpha_{21}$.

2.1. Measurement precision of the measurables

The measurement precision of the measurables is a purely experimental issue, independent of the sensitivity of the measurables with respect to the sample parameters. There are various factors that influence the measurement precision. For the particular case of the angle-resolved scatterometer in our laboratory, we isolated as possible sources of errors the power fluctuations of the laser, mechanical vibrations due to the movements of parts of the system, and finite resolution of the lock-in amplifier. Typically the measurable is the reflectance of the system, R. The measurement precision of R, assuming influences from all the perturbation factors enunciated above, has the form[1, 8]

$$\sigma^2(R) = \sigma_{res}^2 + c_R R^2 + c_\theta \left(\frac{\partial R}{\partial \theta} \right)^2 . \tag{5}$$

Here σ_{res} is the contribution of the finite resolution, c_R is the coefficient of the laser fluctuations contribution, which is linear in R, and c_θ is a coefficient that describes the contribution of the mechanical vibrations of the system on the incidence plane. The latter term in Eq. (5) integrate the uncertainties of the incidence angle θ in the uncertainty of the measurable R. The terms in the sum (5) are squared because we have no reason to believe that the various contributions are correlated[7]. Using (5) as a fitting function for the measurement error of R, one obtains the coefficients from the right hand side of (5). We will refer to Eq. (5) as "the empirical error function".

2.2. Sensitivity

The sensitivity of the measurable with respect to the parameters of the sample is an intrinsic characteristic of the sample, irrespective of the experimental conditions. As expressed above in (4), the situation in which $k=l$ accounts for the sensitivity of the measurable to changes in a single sample parameter. The situation in which $k \neq l$ accounts for the correlation effects of changes in multiple sample parameters. All situations should be considered in the determination of (3).

Inspecting Eqs. (3) and (4) we notice that it is difficult to analytically define the sensitivity unless all σ_i are equal and can be factored. For an arbitrary empirical error function, sensitivity and measurement error cannot be uncoupled. This situation has been accommodated in the results presented below.

For illustration purposes, let us assume that the empirical error function (5) is a constant, i.e.

$$\sigma_i = \sigma^q = const \tag{6}$$

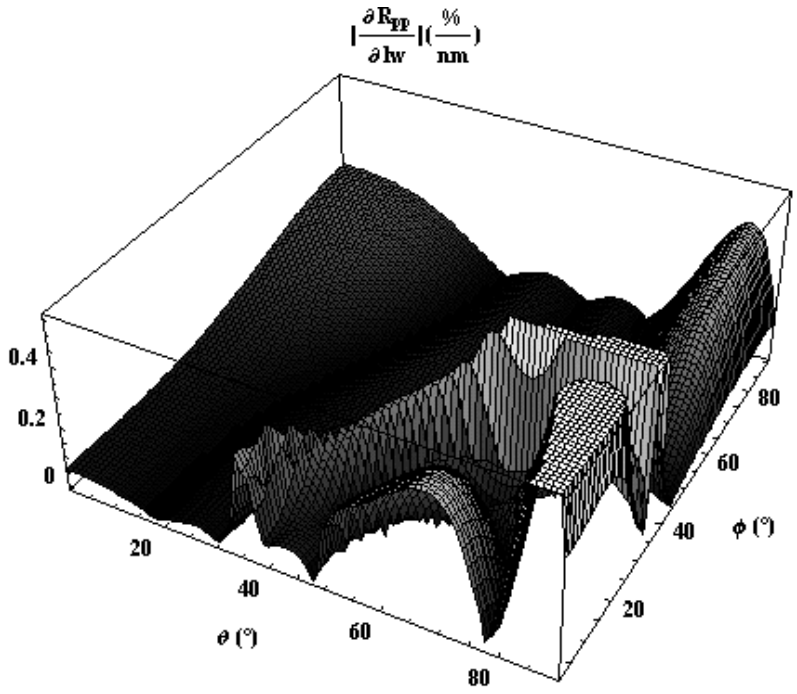


Fig. 1 Partial derivative of the measurable R_{pp} with respect to the linewidth *lw* for the sample "XL6" versus the incidence angle θ and azimuth angle ϕ.

Then we can rewrite Eqs. (3, 4) as

$$\sigma\left(p_j^{geo}\right) = \sigma^q \sqrt{\gamma_{jj}^{-1}} = \sigma^q / sens \,, \quad (7)$$

where γ is

$$\gamma_{k,l} = \sum_{i=1}^{N} \frac{\partial q_i}{\partial p_k^{geo}} \frac{\partial q_i}{\partial p_l^{geo}} = \left(\sigma^q\right)^2 \alpha_{k,l} \,. \quad (8)$$

From Eq. (7) it follows quite intuitively that *sens*, the inverse of $\sqrt{\alpha_{jj}^{'-1}}$, can be chosen as the quantitative definition of the intrinsic sensitivity of the measurables with respect to the sample parameter p_j. Eq. (7) clearly illustrates our statement that a good determination precision of the sample parameters is achieved when we have good measurement precision of the measurables (small σ^q) and high sensitivity (large *sens)*.

More generally we have two aspects that need to be taken into account for a correct evaluation of the sensitivity. First, the absolute magnitude of the partial derivatives $\partial q_i / \partial p_k^{geo}$ must be large, and second, their correlation that must be low. Fig. 1 shows the partial derivative of the ellipsometric measurable R_{pp} (the reflectance for *p* input polarization and *p* output polarization) for the sample labeled "XL6", with parameters defined in Table 1. Intuitively one is drawn to the conclusion that the measurements should be made at pairs of angles θ and ϕ where the partial derivative is large (i.e. the sensitivity is high). In many cases this is true, but not necessarily. The reason is that generally more than one sample parameter is fit. The variation of the measurement curve caused by the variation of a certain parameter p_1 might be similar to that for the variation of another parameter p_2. If their correlation is strong, then two curves corresponding to two different sets of parameters (p_1, p_2) may be indistinguishable. This can happen even if we have large partial derivatives, as in Fig. 1, for large θ and small ϕ. In this latter case, if we change the sign of the variation of one of the parameters, the curves might be well separated (see Fig. 2. for illustration). However, this does not change the situation because the worst case defines the real sensitivity.

Table 1. Geometrical and optical parameters of the samples investigated. The subscripts "g" and "s" refer to the grating and the substrate, respectively. The layers from the grating to the substrate have numerical subscripts starting with 1. Sample "XL6" is real, and sample "P200" is theoretical.

Sample	**Λ (Å)**	**lw (Å)**	**h_g (Å)**	**h_1 (Å)**	**n_g**	**n_1**	**n_s**
XL6	3628.8	1470	5700	1600	1.64	1.66	3.882 – 0.019 i
P200	2000	700	4000	–	1.64	–	3.882 – 0.019 i

The above considerations can be better understood if we look at the two-dimensional case, in which we need to fit two parameters. In this situation Eq. (3) becomes

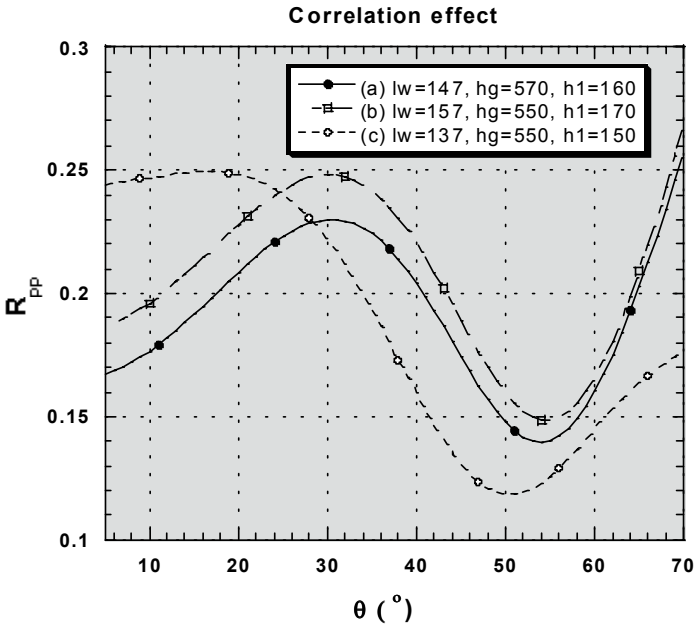


Fig. 2. Example of the correlation effect. Curve (b) is much stronger correlated with (a) than (c), although the corresponding sample parameters values are equally separated. The values for the parameters displayed on the graph are in nanometers. Case (a) "XL6".

$$\sigma\left(p_1^{geo}\right)=\sqrt{\frac{\alpha_{22}}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}}}\,, \qquad (9.a)$$

$$\sigma\left(p_2^{geo}\right)=\sqrt{\frac{\alpha_{11}}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}}}\,. \qquad (9.b)$$

We have small determination errors if the diagonal terms of the curvature matrix α are large, and the non-diagonal one is small. The diagonal terms are the sum of the squares of the partial derivative amplitudes, while the non-diagonal term is their correlation, which can be seen by particularizing Eq. (4) as

$$\alpha_{jj}=\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\left(\frac{\partial q_i}{\partial p_j^{geo}}\right)^2\,, \qquad (10.a)$$

$$\alpha_{12}=\alpha_{21}=\sum_{i=1}^{N}\frac{1}{\sigma_i^2}\frac{\partial q_i}{\partial p_1^{geo}}\frac{\partial q_i}{\partial p_2^{geo}}\,. \qquad (10.b)$$

The above discussion is important because, when studying the sensitivity, one may be tempted to overlook the correlation term (10.b) and concentrate only on the magnitude terms (10.a)[6].

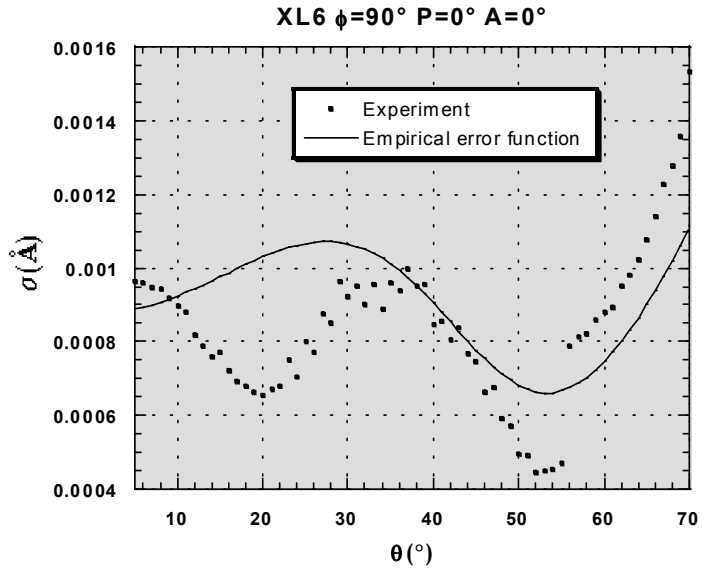


Fig. 3 Determination of the empirical error function; no intensity stabilizer utilized in the experimental arrangement.

3. COMPARISON OF SPECTROSCOPIC AND ANGULAR SCATTEROMETRY

Now that we have a method for determining the measurement precision of a metrology technique, we can apply it to compare the performances of spectroscopic and angular scatterometry. A difficulty is posed by the fact that we do not have an empirical error function for the spectroscopic scatterometer; we will approximate this subsequently. We determined the empirical error function for the angle-resolved scatterometer by repeatedly measuring the reflectance, R, over a wide range of values. The statistical interpretation of the batch of data in the light of Eq. (5) lead us to the empirical error function

$$\sigma^2\left(\mathring{A}\right)=1.0\times10^{-8}+1.3\times10^{-5}R^2\,. \qquad (9)$$

We did not notice significant correlation of the experimental errors with $\partial R/\partial\theta$. Even for the other two terms, the correlation was very weak (see Fig. 3). Eq. (9) is just the

Table 2. Experimental results illustrating the improvement in the linewidth measurement precision provided by the use of the laser intensity stabilizer in the case of the sample "XL6". The wavelength was 633 nm.

Configuration			σ (Å)	
P (°)	A (°)	ϕ (°)	Without stabilizer	With stabilizer
0	0	0	4.9	1.9
0	0	45	–	3.2
0	0	90	6.3	4.0
90	90	0	19.5	0.6
90	90	45	–	1.9
90	90	90	7.6	1.4

Table 2 illustrates results from applying the sensitivity analysis techniques discussed above to sample "XL6", with 147 nm linewidth. We performed repeated (50) measurements of the sample in each of 6 measurement configurations to illustrate the influence of the two empirical error functions on measurement precision. The measurement precision improves significantly (see Table 2) with the improved intensity stabilization and the smaller empirical error function (10), up to 2-3 times improvement depending upon the measurement configuration. The wavelength used was 633 nm. It can be seen that in all but one situation, sub-nm precision was observed for both empirical error functions (9) and (10).

closest approximation (best fit) obtained. For this functional form, the scatterometer apparatus did not contain a means to stabilize the laser intensity or to normalize the measurements to input intensity. Using a laser intensity stabilizer (THORLABS "Noise Eater"), the empirical error function becomes

$$\sigma^2\left(\text{Å}^2\right)=1.0\times10^{-8}+5.0\times10^{-7}R^2+1.0\times10^{-4}\left(\partial R/\partial\theta\right)^2 . \quad (10)$$

Again, Eq. (10) is just the closest approximation (best fit); see Fig. 4. The additional term in (10) is the result of much lower intensity variations that cause additional effects to be influential.

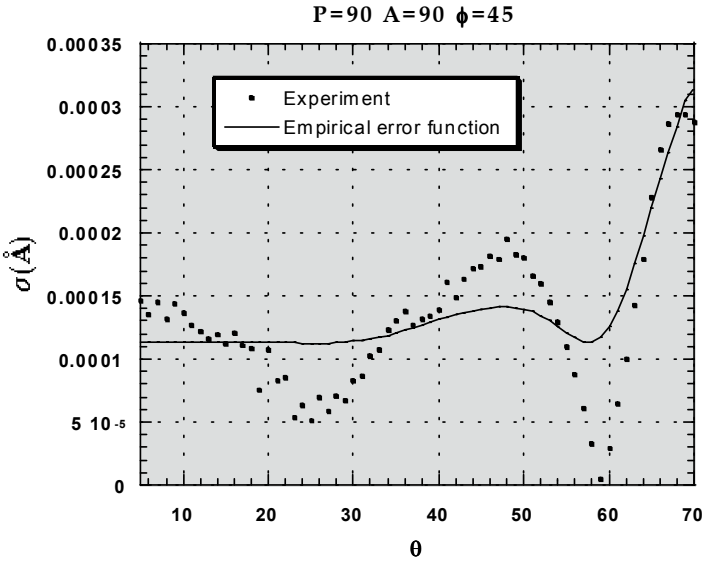


Fig. 4 Determination of the empirical error function, with a stabilizer is inserted in the experimental arrangement.

Obtaining empirical error functions for a spectroscopic scatterometer is less straightforward. The material optical properties must be determined over the wavelength range of interest. This may not be an issue if the technique is used in concert with a spectroscopic ellipsometer. However, one must be confident that process changes that might occur do not significantly influence material optical properties, especially at UV and blue wavelengths. The wavelength range reported in recent work involving this technique[6] is 230 nm to 450 nm. Additionally, if a detection arrangement is used that consists of an array, all elements of the array would optimally have the same empirical error function. Another distinction of the spectroscopic scatterometer is that it utilized a low intensity, incoherent light sources.

To compare measurement precision of the two scatterometer techniques discussed here, we need an empirical error function for the spectroscopic scatterometer. Not having one, we approximated it to be the same as that for the angular scatterometer. In light of the remarks from the previous paragraph about the measurement precision of the spectroscopic scatterometer technique, this approximation is likely quite optimistic. In particular, for the simulated results shown in Table 3, we used the empirical error function (9). Table 3 indicates results for each of the four measurement configurations indicated. However, as opposed to Table 2, which contains experimental results, Table 3 has only calculated simulations. The improvement provided by shorter wavelength is apparent.

The conclusion of interest that inspection of Table 3 provides is that there is no conclusive improvement of measurement precision of one technique over the other. In their recent paper, Allgair et al[6] state that spectroscopic scatterometry is more sensitive than variable-angle scatterometry. However, this conclusion was drawn from comparing the sensitivity of variable-angle measurements utilizing 633 nm light to variable-wavelength measurements utilizing light over the range of 230 nm to 450 nm. Spectroscopic scatterometry would be indeed more sensitive than variable-angle scatterometry if the variable-angle technique utilized wavelengths significantly longer than those utilized in variable-wavelength, such as these. Variable-angle scatterometry shows similar sensitivity and precision as spectroscopic scatterometry if light sources of roughly comparable wavelengths are utilized. Additionally, for the correct assessment of sensitivity we need to take into account of correlation effects as well, which the authors neglected. Finally, as we shown, superior sensitivity is not sufficient for superior results.

Table 3. Simulation results for variable wavelength scatterometry (θ=45° and wavelength range λ=230…450 nm) and variable-angle scatterometry (angular range θ =5°…70°). In both cases the empirical error function Eq. (9) was assumed.

Configuration		**σ (Å)**							
		Variable-wavelength scatterometer λ=230…450 nm		**Variable-angle scatterometer**					
				λ=325 nm		**λ=442 nm**		**λ=633 nm**	
Measurable	**Azimuth ϕ(°)**	**XL6**	**P200**	**XL6**	**P200**	**XL6**	**P200**	**XL6**	**P200**
R_{pp}	0	0.85	0.43	6.60	1.51	2.65	8.30	1.8	5.0
R_{ss}	0	0.35	0.38	0.50	0.28	0.95	5.93	0.6	1.4
R_{pp}	90	0.42	3.35	0.42	3.38	4.30	6.55	15.0	1.5
R_{ss}	90	0.06	1.37	0.42	1.56	0.72	1.67	8.4	8.4

4. CONCLUSIONS

The results presented above illustrate the importance of understanding the two contributions to measurement precision of sample parameters: sensitivity of the measurable to changes in sample parameters, and the precision that the measurable quantity can be determined (i.e. the empirical error function). A correct approach of the sensitivity analysis must also include the correlation effects that one may be tempted to overlook. The empirical error function can be easily determined for the angle-resolved scatterometer configuration. The use of an intensity stabilizer can significantly improve the empirical error function and, consequently, the measurement precision of the sample parameters. Experimental results illustrate how angle-resolved scatterometry can provide sub-nm measurement precision of a nominal 147 nm linewidth sample. Using this same empirical error function together with calculations of sensitivity of *R* to changes in a 70 nm linewidth sample, simulations indicate sub-nm measurement precision is provided by the variable-angle scatterometer using visible light. Parallel simulations for the spectroscopic scatterometry and the angle-resolved scatterometry show comparable precision results. Practical considerations indicate the angle-resolved technique is advantageous to the spectroscopic technique. As a general trend for both scatterometry techniques, the use of a shorter wavelength improves the sensitivity and the measurement precision of the sample parameters.

REFERENCES

1. P. C. Logofatu, dissertation, University of New Mexico (2000).
2. J. R. McNeil, S. A. Coulombe, P. C. Logofatu, C. J. Raymond, S. S. H. Naqvi, "Application of optical scatterometry to microelectronics," *Conference on Lasers and Electro-Optics*, 1998.
3. S. A. Coulombe, P. C. Logofatu, B. K. Minhas, S. S. H. Naqvi, J. R. McNeil, "Ellipsometric-Scatterometry for sub-0.1 mm CD measurements," *Microlithography*, *Proc. SPIE 3332*, pp. 282-293, 1998.

4. J. R. McNeil, S. S. H. Naqvi, S. M. Gaspar, K. C. Hickman, S. R. Wilson, in *Encyclopedia of Materials Characterization* C. A. E. C. Richard Brundle, Shaun Wilson, Ed. (Manning Publishing Co., 1992).

5. S. S. H. Naqvi, J. R. McNeil, R. H. Krukar, Z. R. Hatab, "Grating parameter estimation using scatterometry," *Miniature and Micro-Optics and Micromechanics*, *Proc. SPIE 1992*, pp. 170, 1993.

6. J. Allgair, et al., "Manufacturing considerations for implementation of scatterometry for process monitoring," *Metrology, inspection and process control for microlithography XIV*, *3998*, pp. 126-134, 2000.

7. P. R. Bevington, D. K. Robinson, *Data reduction and error analysis for the physical sciences*. (McGraw-Hill, Boston, 1992).

8. P. C. Logofatu, J. R. McNeil, "Sensitivity analysis of fitting for scatterometry," *Microlithography XIII*, *Proc. SPIE 3677*, pp. 177-183, 1999.

Journal of optoelectronics and advanced materials **3**(1), 45-49 (2001)

Analysis of phase measurement error for null generalized ellipsometry using the phase compensator

P.C. LOGOF~TU[a]
Center of High Technology Materials/ University of New Mexico
1313 Goddard SE, Albuquerque NM 87106, USA

Abstract: An error analysis of null generalized ellipsometry for phase measurement using a Babinet compensator is presented. It turns out that there are errors only in a second order approximation, which means this technique provides high sensitivity for phase measurements. The sensitivity can be improved by equalizing the amplitudes of the x and y components of the field at the Babinet compensator but this complicates considerably the experimental procedure.

1. Introduction

Phase measurement techniques, although known for more than a century, are increasingly used today for their sensitivity and accuracy [1-4]. On the other hand null measurements are reputed for being very sensitive [5]. Therefore null measurements of the phase using a Babinet compensator (null ellipsometry) are likely to provide high sensitivity. In our paper ellipsometry is used for the determination of the phase change in reflection for an anisotropic surface. This type of surfaces cannot be described just by two complex reflection coefficients r_p and r_s. Polarization conversion of light requires four complex reflection coefficients for a complete description of the optical properties of the surface, r_{pp}, r_{sp}, r_{ps} and r_{ss}, a formalism called generalized ellipsometry [6-8]. We derive an expression for the error of phase change measurement in these conditions and it turns out that there is an error only on a second order approximation, which is an eloquent argument in favor of the sensitivity of this measurement technique.

The experimental arrangement for phase measurement is illustrated in Fig. 1. The beam coming from left is linearly polarized by the polarizer P and becomes elliptically polarized after reflection on the grating. The compensator C compensates for the phase shift between x and y component and the beam is again linearly polarized. Another polarizer A (analyzer) is set perpendicular on the incident beam, which vanishes on transmission. The phase measurement is a null measurement and we nullify the light intensity by an appropriate combination of phase shift introduced by the compensator and polarization angle of the analyzer.

The ultimate goal is the determination of the optical properties of the material, i.e. the reflection coefficients. The measurable must be a function of these coefficients, $\Delta = \Delta(r_{pp}, r_{ps}, r_{sp}, r_{ss})$.

2. Error analysis

For calculations we will use the Jones matrix approach [9]. The input components of the beam are

$$\begin{aligned} E_x^{in} &= E^{in}\cos\psi_P, \\ E_y^{in} &= E^{in}\sin\psi_P. \end{aligned} \tag{1}$$

Upon reflection the field becomes

$$\begin{pmatrix} E_x^{ref} \\ E_y^{ref} \end{pmatrix} = \begin{pmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{pmatrix} \begin{pmatrix} \cos\psi_P \\ \sin\psi_P \end{pmatrix} E^{in}, \tag{2}$$

or

$$\begin{aligned} E_x^{ref} &= \left(r_{pp}\cos\psi_P + r_{ps}\sin\psi_P\right)E^{in}, \\ E_y^{ref} &= \left(r_{sp}\cos\psi_P + r_{ss}\sin\psi_P\right)E^{in}. \end{aligned} \tag{3}$$

The phase shift due to reflection on the sample is

$$\Delta = \arg\left(\frac{E_y^{ref}}{E_x^{ref}}\right) = \arg\left(\frac{r_{sp}\cos\psi_P + r_{ss}\sin\psi_P}{r_{pp}\cos\psi_P + r_{ps}\sin\psi_P}\right). \tag{4}$$

The compensator will introduce a phase shift Δ_C which in the end is supposed to be equal to Δ but for now we assume is arbitrary. Then we have

$$\begin{aligned} E_x^C &= E_x^{ref}, \\ E_y^C &= E_y^{ref}\exp\left(-i\Delta_C\right). \end{aligned} \tag{5}$$

After the analyzer the beam is

$$\begin{pmatrix} E_x^{out} \\ E_y^{out} \end{pmatrix} = \begin{pmatrix} \cos^2\psi_A & \cos\psi_A\sin\psi_A \\ \sin\psi_A\cos\psi_A & \sin^2\psi_A \end{pmatrix} \begin{pmatrix} E_x^C \\ E_y^C \end{pmatrix} = \left(\cos\psi_A E_x^C + \sin\psi_A E_y^C\right)\begin{pmatrix} \cos\psi_A \\ \sin\psi_A \end{pmatrix}. \tag{6}$$

The output intensity measured by the detector D is

$$\begin{aligned} I^{out} &= \frac{1}{2}\left(\left|E_x^{out}\right|^2 + \left|E_y^{out}\right|^2\right) = \frac{1}{2}\left|\cos\psi_A E_x^C + \sin\psi_A E_y^C\right|^2 = \\ &\frac{1}{2}\left|\cos\psi_A E_x^{ref} + \sin\psi_A E_y^{ref}\exp\left(-i\Delta_C\right)\right|^2 = \\ &\cos^2\psi_A I_x^{ref} + \sin^2\psi_A I_y^{ref} + 2\sin\psi_A\cos\psi_A\sqrt{I_x^{ref}I_y^{ref}}\cos\left(\Delta - \Delta_C\right), \end{aligned} \tag{7}$$

where

$$\begin{aligned} I_x^{ref} &= I^{ref}\cos^2\alpha, \\ I_y^{ref} &= I^{ref}\sin^2\alpha, \end{aligned} \tag{8}$$

and

$$\tan\left(\alpha\right) = \left|\frac{r_{sp}\cos\psi_P + r_{ss}\sin\psi_P}{r_{pp}\cos\psi_P + r_{ps}\sin\psi_P}\right|. \tag{9}$$

I^{ref} is the intensity that can be measured by a detector right after the reflection on the sample and also happens to be the maximum output at the detector D in Fig. 1 (when the analyzer A is

parallel to the linearly polarized output beam and if there are no pure transmission losses in the compensator and the analyzer). Obviously, from Eq. (3), I^{ref} is of the form

$$I^{ref} = \left(\left| r_{sp} \cos\psi_P + r_{ss} \sin\psi_P \right|^2 + \left| r_{pp} \cos\psi_P + r_{ps} \sin\psi_P \right|^2 \right) I^{in}, \tag{10}$$

with

$$I^{in} = \frac{1}{2} \left| E^{in} \right|^2. \tag{11}$$

α is the polarization angle of the beam when the phase shift is compensated and also the angle at which there is one of the semiaxes of the polarization ellipse. Using Eqs. (8) and (9) we can rewrite (7) as

$$I^{out} = I^{ref} \left(\cos^2\psi_A \cos^2\alpha + \sin^2\psi_A \sin^2\alpha + 2\sin\psi_A \cos\psi_A \sin\alpha \cos\alpha \cos(\Delta_C - \Delta) \right). \tag{12}$$

To obtain an expression for the phase measurement error we derive the output I^{out} on the phase Δ:

$$\frac{\partial I^{out}}{\partial \Delta} = -\frac{1}{2} \sin(2\psi_A) \sin(2\alpha) \sin(\Delta - \Delta_C) I^{ref}. \tag{13}$$

At null conditions one has $\psi_A \approx \alpha \pm \pi/2$ and $\Delta \approx \Delta_C$. Then it follows

$$\delta I^{out} = \frac{1}{2} \left| \sin^2(2\alpha) I^{ref} \right| \delta\Delta^2, \tag{14}$$

or

$$\delta\Delta = \sqrt{\frac{2\delta I(0)}{\sin^2(2\alpha) I^{ref}}}, \tag{15}$$

where $\delta I(0)$ is the measurement error for output intensity when this intensity is zero. Generally this is the minimum value of the error.

3. Comments

It is important to note that the error appears only in the second order approximation. There is no measurement error for the phase in a first order approximation. This is a direct consequence of the fact that a null measurement type technique is used. It must be also mentioned that these considerations apply to precision not to accuracy. They are valid only if random, unbiased errors are assumed.

This relation makes sense, and we can derive it just from intuitive considerations. Since the phase measurement is a null measurement, the larger is the input to nullify, the more precise is the measurement of the phase. And we can see that for large values of I^{ref}, $\delta\Delta$ is indeed smaller. Also, the precision of the phase compensation is better when the components E_x^{ref} and E_y^{ref} have comparable amplitudes and is maximum when they are equal. The precision of the phase compensation is obviously related to how large is the difference between the maximum and the minimum output at the detector, because a large difference allows you a better discrimination of the minimum. Or this difference is a maximum when

$$\left| E_x^{ref} \right| = \left| E_y^{ref} \right| \tag{16}$$

and is minimum when one of the components is zero. Indeed the smaller error is for $\alpha=45°$, when Eq. (16) is fulfilled and becomes infinite for $\alpha=0°$ or $\alpha=90°$, i.e. E_x^{ref} or E_y^{ref} is zero. From Eq. (12) if we allow the compensating phase shift Δ_C to vary we obtain

$$I^{\max} - I^{\min} = \left|\sin(2\psi_A)\sin(2\alpha)\right| I^{ref} . \tag{17}$$

This quantity is obviously maximum when $\alpha=45°$ and is minimum when $\alpha=0°$ or $\alpha=90°$.

Therefore, an important enhancement of the precision of the phase measurement can be done if we rotate the axes of the compensator with an amount θ necessary to equilibrate the x and y components. In this case Eq. (15) becomes

$$\delta\Delta = \sqrt{\frac{2\delta I(0)}{I^{ref}}} . \tag{18}$$

However, Δ becomes, instead Eq. (4),

$$\Delta = \arg\left(\frac{E_x^{ref}{}'}{E_y^{ref}{}'}\right) = \arg\left(\frac{E_x^{ref}\cos\theta + E_y^{ref}\sin\theta}{-E_x^{ref}\sin\theta + E_y^{ref}\cos\theta}\right), \tag{19}$$

where $E_x^{ref}{}'$ and $E_y^{ref}{}'$ are the components of the reflected beam in the rotated frame. Although the rotation of the compensator axes can dramatically improve the precision, in practice is difficult to implement and it may not be worth the trouble. To find the angle θ, you need to scan the output at the detector for various angles ψ_A with the compensator inactive until you find two positions say ψ_1 and ψ_2 separated by 90° where the output is equal. Then you have to rotate the compensator so that the x and y axes matches the two positions. Then θ must be noted for further use in calculations. This procedure must be repeated for each measurement and for each measurement we have a different value for θ.

One must be aware that Eq. (15) is valid only in the approximation that there are no pure transmission losses in the compensator and the analyzer. Generally we do have this kind of losses and Eq. (15) must be modified accordingly

$$\delta\Delta = \sqrt{\frac{2\delta I(0)}{\sin^2(2\alpha) T_C T_A I^{ref}}} , \tag{20}$$

where T_C is the transmission of the compensator and T_A is the maximum transmission of the analyzer, when the beam is linearly polarized and the analyzer is parallel to the beam polarization.

4. Conclusion

We have shown using quantitative arguments that null ellipsometry using a Babinet compensator is a highly sensitive phase measurement technique. The first order approximation of the error is zero and he have to go to the second order approximation. By rotating the compensator to a position where the x and y components are of equal amplitude the sensitivity can be considerably increased but this complicates the experimental procedure and the computations and introduces supplementary sources of errors.

References

1. R. M. A. Azzam, *Perpendicular-incidence null ellipsometry of surfaces with arbitrary anisotropy*, Optical Engineering **20**(1), 58 (1981).

2. W. M. Duncan and S. A. Henck, *Insitu spectral ellipsometry for real-time measurement and control*, Applied Surface Science **63**, 9 (1993).

3. S. A. Coulombe, P. C. Logofatu, B. K. Minhas, S. S. H. Naqvi and J. R. McNeil, "Ellipsometric-Scatterometry for sub-0.1 mm CD measurements," in Microlithography, Proc. SPIE 3332, 282 (1998).

4. O. Acher, E. Bigan and B. Drevillon, *Improvements of phase-modulated ellipsometry*, Rev. Sci. Instrum. **60**(1), 65 (1989).

5. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and polarized light*. North-Holland, Amsterdam, 1977.

6. R. M. A. Azzam, *Application of generalized ellipsometry to anisotropic crystals*, J. Opt. Soc. Am. **64**(2), 128 (1974).

7. R. M. A. Azzam and N. M. Bashara, *Generalized ellipsometry for surfaces with directional preference: Application to diffraction gratings*, J. Opt. Soc. Am **62**(12), 1521 (1972).

8. D. J. DeSmet, *Generalized Ellipsometry and the 4x4 Matrix Formalism*, Surface Science **56**, 293 (1976).

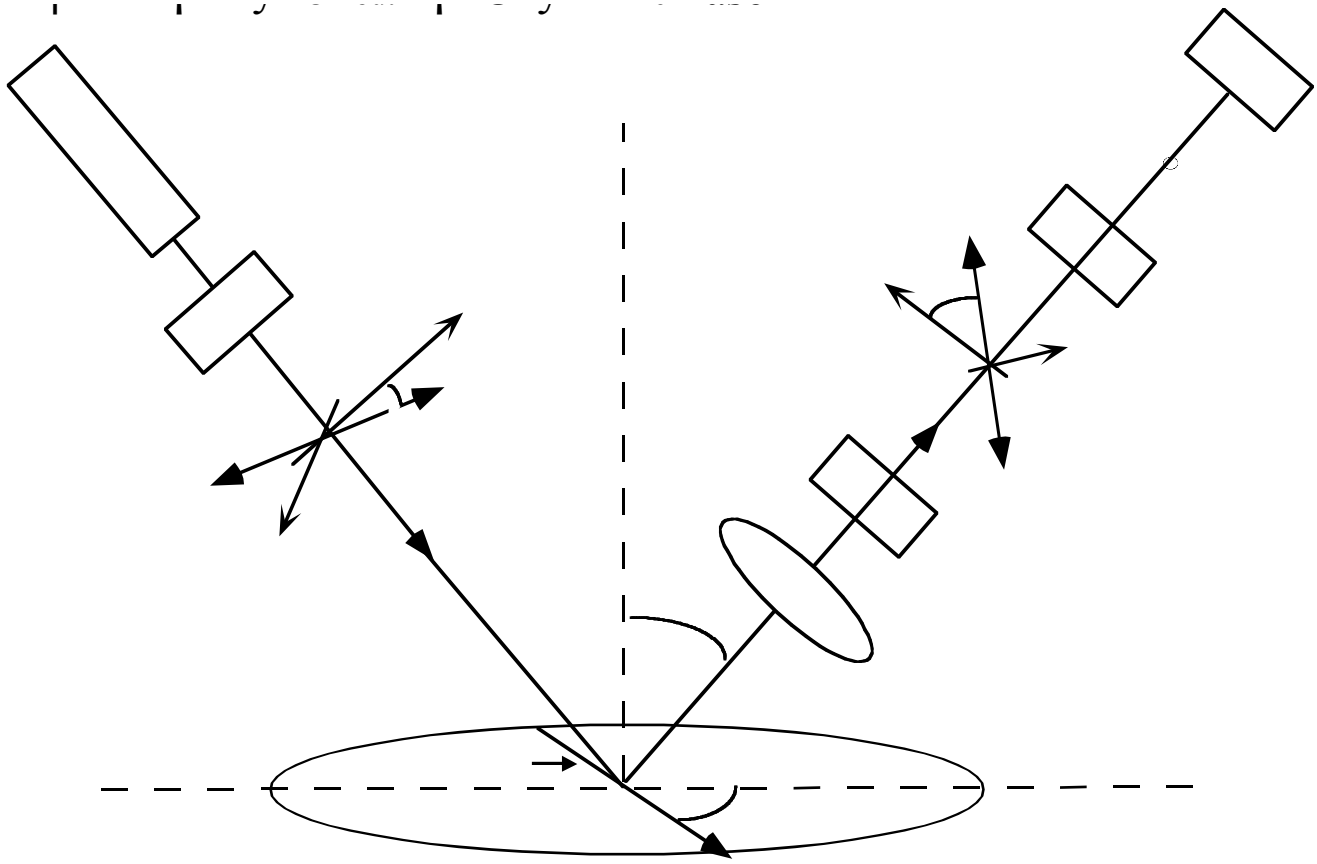
9. E. Collett, *Polarized light - Fundamentals and applications*. Marcel Dekker, Inc., New York, 1993.

Figure captions

Fig.1 Experimental arrangement for phase measurement using a Babinet compensator. P is the polarizer, A the analyzer, C the phase compensator, D the detector, **K** the grating vector, φ the azimuth angle, θ the incidence angle, ψ_P the polarizer angle, ψ_A the analyzer angle and E^{in}, E^{ref}, E^{C} and E^{out} are the light field at different positions in the system.

Footnote

a) Permanent address: National Institute of Laser, Plasma and Radiation Physics (INFLPR), Bucharest-M`gurele, PO-Box MG-36, Romania



Simple method for determining the fast axis of a wave plate

Petre-Catalin Logofatu

Publication: Optical Engineering 41(12), p. 3316-3318, Donald C. O'Shea; Ed.

Publication Date: Dec 2002

Abstract
Based on the metals property of modifying the phase of an incident beam, a simple method for identifying the fast axis of a wave plate independent of a standard retarder is demonstrated.

©2005 SPIE--The International Society for Optical Engineering.

Phase-modulation scatterometry

Petre C. Logofătu

Phase-modulation scatterometry is a metrology technique for determining, by means of a phase modulator as a key device, the parameters of gratings. The main source of error to be dealt with are the fluctuations of the phase-modulation amplitude. The grating zeroth-order reflectance modulated by the phase modulator is converted into a signal by the photodetector. The measurables are the direct term and the first two harmonics of the signal. For experimental data fitting, we used the ratio of the harmonics over the direct term because it significantly improves the accuracy. A sensitivity analysis was performed for two samples, one real and one theoretical, to find the measurement configuration that insures optimum determination precision for the grating parameters. For the real sample, comparison of the theoretical predictions for sensitivity with the actual values showed a good agreement. For both samples the sensitivity analysis indicated subnanometric precision for the critical dimension (grating linewidth). © 2002 Optical Society of America

OCIS codes: 050.1950, 120.0120, 120.5060, 260.5430, 120.5700, 260.2130.

1. Introduction

In the ongoing effort to develop scatterometry as a sensitive metrology technique, researchers continue to improve it by implementing new measurement techniques.[1–6] In this paper the implementation of the phase-modulation technique (PMT) in scatterometry is discussed. PMT was first applied for optical measurement by Jasperson and Schnatterly[7] and has been continuously developed since then.[8,9] The technique was previously applied to scatterometry,[9] but an analysis of the sensitivity and simplified synthetic formulas for the measurables, to our knowledge, are not available in the literature and are thus an original contribution. Generally, we stressed the new and original elements adduced in this paper, but, when required by the intelligibility of the text, we briefly reviewed other authors' research. The main contribution presented here is the sensitivity analysis, formally identical to the one performed for the ellipsometric scatterometer discussed in Ref. 6, namely, the application of the sensitivity analysis for fitting (SAF) formalism for the phase-modulation scatterometer. The same samples that were analyzed in Ref. 6, labeled XL6 and P200, are analyzed here as well, and comparison between the sensitivity of the ellipsometric scatterometer and the phase-modulation scatterometer is made. As one can see from Ref. 6, sample parameters have two categories. First, one has the optical parameters that are known *a priori* from other types of measurement than scatterometry, the pitch Λ and the refraction indices of the materials of which the sample is made. Second, one has the geometric parameters that must be estimated from scatterometry measurements: the linewidth lw (or the critical parameter), the grating height h_g, and heights of the underlying layers h_i. Although all the parameters in the second category are calculated from scatterometry measurements, the sensitivity analysis is performed only for the parameter of interest in the microelectronics industry, the critical parameter lw. The diffraction calculations in this paper were made with the rigorous coupled-wave analysis (RCWA).[10,11]

2. Experimental Arrangement

The experimental arrangement that we call phase-modulator scatterometer is still an ellipsometric scatterometer. Compared with the scatterometer described in Ref. 6, there is only one new device added to the system, the phase modulator (see Fig. 1). This one device, however, radically modifies the nature of the experimental arrangement, as we shall see in the following.

The phase modulator we used was a PEM-90 phase modulator, manufactured by Hinds Instruments

P. C. Logofătu (logofatu@unm.edu) is with the Center of High Technology Materials, University of New Mexico, 1313 Goddard Street, Southeast, Albuquerque, New Mexico 87106 and also with the National Institute of Laser, Plasma and Radiation Physics (INFLPR), Laser Department, P.O. Box MG36, Bucharest 76900, Romania.

Received 11 March 2002; revised manuscript received 23 May 2002.

0003-6935/02/347187-06$15.00/0

© 2002 Optical Society of America

(Hillsboro, Oregon),[12] with an oscillation frequency of 50 kHz. The modulator introduces a variable phase retardation along one of the polarization axes, causing the eccentricity of the polarization ellipse to oscillate. The analyzer converts the light with oscillating polarization into light with oscillating amplitude, which is collected by the photodetector. A lock-in amplifier measures the harmonics of the signal. These harmonics are the measurables of the experimental arrangement, and they are functions of the geometric parameters of the sample.

The experimental parameters of the phase-modulation scatterometer are, as seen in Fig. 1, the polarizer polarization angle P, the orientation of the modulator axis of variable retardation M, the incidence angle of the light onto the sample θ, the azimuth angle of the sample ϕ, and the analyzer polarization angle A. Parameters P, M, ϕ, and A are constant during the measurement process, and they define a measurement configuration. The domain of these parameters is scanned to find the configuration with optimum sensitivity. The angle θ is variable during the experiment and is typically varied from 5° to 70° in steps of 1°. A He–Ne laser (the wavelength $\lambda = 633$ nm) was the light source used in this experiment.

3. Measurables

From Fig. 1 one sees that the transformation incurred by the beam in the phase-modulator scatterometer is described by

$$\mathbf{E}^{\text{out}} = \hat{\mathbf{A}}\hat{\mathbf{r}}\hat{\mathbf{M}}\hat{\mathbf{P}}\mathbf{E}^{\text{in}}, \tag{1}$$

where the Jones matrices for analyzer, sample, and modulator

$$\hat{\mathbf{A}} = \begin{bmatrix} \cos^2 A & \cos A \sin A \\ \sin A \cos A & \sin^2 A \end{bmatrix}, \tag{2a}$$

$$\hat{\mathbf{r}} = \begin{pmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{pmatrix}, \tag{2b}$$

$$\hat{\mathbf{M}} = \begin{bmatrix} \cos M & -\sin M \\ \sin M & \cos M \end{bmatrix} \begin{bmatrix} \exp[-j\delta(t)] & 0 \\ 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos M & \sin M \\ -\sin M & \cos M \end{bmatrix}, \tag{2c}$$

respectively, were used, with

$$\delta(t) = \mathcal{A} \sin(\omega t), \tag{3}$$

which is the variable phase retardation introduced by the modulator and in which $\mathcal{A}$ is the phase-modulation amplitude. The matrix of the polarizer is simply the analyzer matrix with A substituted with P. Values $\mathbf{E}^{\text{out}}$ and $\mathbf{E}^{\text{in}}$ are the output and the input vector electric fields, respectively. The coefficients r_{uv} of the reflection matrix $\hat{\mathbf{r}}$ describe the matrix according to generalized ellipsometry.[13] After making the calculations in Eq. (1) we obtain

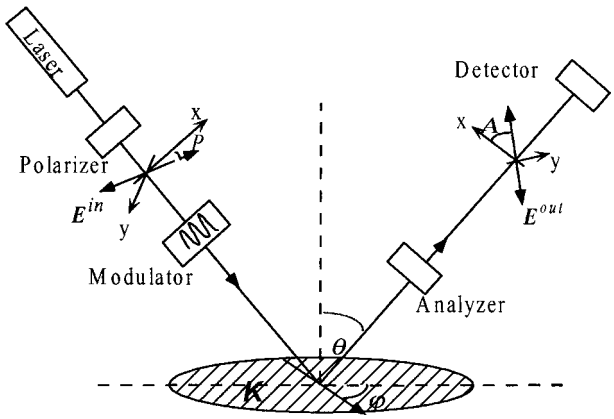


Fig. 1. Experimental arrangement of the phase-modulator scatterometer.

$$\begin{aligned} r = E^{\text{out}}/E^P = & \\ &= \cos(P - M)(r_P \cos M + r_s \sin M) \\ &\quad + \sin(P - M)(-r_P \sin M \\ &\quad + r_s \cos M)\exp[-j\delta(t)], \end{aligned} \tag{4}$$

where we used the notations

$$r_p = r_{pp} \cos A + r_{sp} \sin A, \tag{5a}$$

$$r_s = r_{ps} \cos A + r_{ss} \sin A, \tag{5b}$$

$$\mathbf{E}^{\text{out}} = E^{\text{out}} \begin{bmatrix} \cos & A \\ \sin & A \end{bmatrix}, \tag{6a}$$

$$\begin{aligned} \mathbf{E}^P &= (E_p^{\text{in}} \cos P + E_s^{\text{in}} \sin P) \begin{bmatrix} \cos & P \\ \sin & P \end{bmatrix} \\ &= E^P \begin{bmatrix} \cos & P \\ \sin & P \end{bmatrix}. \end{aligned} \tag{6b}$$

The notations in Eqs. (5a) and (5b) were chosen for convenience reasons, but they have a certain physical meaning as well: namely, r_p is the contribution of the p component of the beam incident on the sample to the output, and r_s is the similar contribution of the s component.

Squaring the absolute value of r in Eq. (4) we obtain in terms of power

$$I(t) = I_0 + I_s \sin \delta(t) + I_c \cos \delta(t), \tag{7}$$

where

$$\begin{aligned} I_0 = |r_p|^2 + |r_s|^2 + \cos 2(P - M)[\cos 2M(|r_p|^2 - |r_s|^2) \\ + 2 \sin 2M \operatorname{Re}(r_p r_s^*)], \end{aligned} \tag{8a}$$

$$I_s = 2 \sin 2(P - M) \operatorname{Im}(r_p r_s^*), \tag{8b}$$

$$\begin{aligned} I_c = \sin 2(P - M)[-\sin 2M(|r_p|^2 - |r_s|^2) \\ + 2 \cos 2M \operatorname{Re}(r_p r_s^*)]. \end{aligned} \tag{8c}$$

The calculations through which we obtained the functions I_0, I_s, and I_c are very laborious, and it is even more laborious to obtain a simple, synthetic expression of them as in Eqs. (8a)–(8c). However, there is a simple way to verify their accuracy. With the help of a programming environment for symbolic calculation as powerful as Mathematica,[14] we can calculate the expanded form of the harmonics, which is in the range of pages of formulas. Then we program the computer to simplify the difference between the expanded forms of the harmonics and their synthetic expression. As expected, the result of the simplification is zero. One may notice that $I(t)$ from Eq. (7) is not actually the square of r but is its double. The reason for this choice is that it normalizes the harmonics in Eqs. (8a)–(8c); their maximum absolute value becomes 1.

After expanding the trigonometric functions in Eq. (7) in Bessel functions we obtain the harmonics' expression

$$\begin{bmatrix} S_0 \\ S_\omega \\ S_{2\omega} \end{bmatrix} = \begin{bmatrix} 1 & 0 & J_0(\mathcal{A}) \\ 0 & 2J_1(\mathcal{A}) & 0 \\ 0 & 0 & 2J_2(\mathcal{A}) \end{bmatrix} \begin{bmatrix} I_0 \\ I_s \\ I_c \end{bmatrix}. \tag{9}$$

The S coefficients of Eq. (9) are what can be measured with the phase-modulator scatterometer in the ideal case. For convenience the value of $\mathcal{A}$ is chosen to be $\mathcal{A}_0 = 137.8°$, because then we have $J_0(\mathcal{A}_0) = 0$, thus simplifying the equations. In practice we prefer to work with the quantities

$$R_\omega = \frac{S_\omega}{2J_1(\mathcal{A}_0)S_0}, \tag{10a}$$

$$R_{2\omega} = \frac{S_{2\omega}}{2J_2(\mathcal{A}_0)S_0}. \tag{10b}$$

There are some obvious advantages of working with the quantities in Eqs. (10a) and (10b), besides the fact that they are normalized. Their numerator and denominator are measured simultaneously; therefore their ratio is free from systematic errors caused by the spatial displacement of the beam spot onto the photodetector area and by the fluctuations in time of the laser power and other variables.

Note also that, with the phase-modulator scatterometer, we extract more information than in the case of ellipsometric scatterometer,[6] because we measure the two quantities in Eqs. (10a) and (10b) simultaneously. Therefore, the information provided by the phase-modulator scatterometer is richer than that provided by the ellipsometric scatterometer because we take double the number of measurements in a θ scan.

4. Calibration

In practice, it turns out that Eq. (9) describes only an ideal behavior of the phase modulator. Corrections for the nonideal behavior were introduced in the calculations according to Acher *et al.*[8]

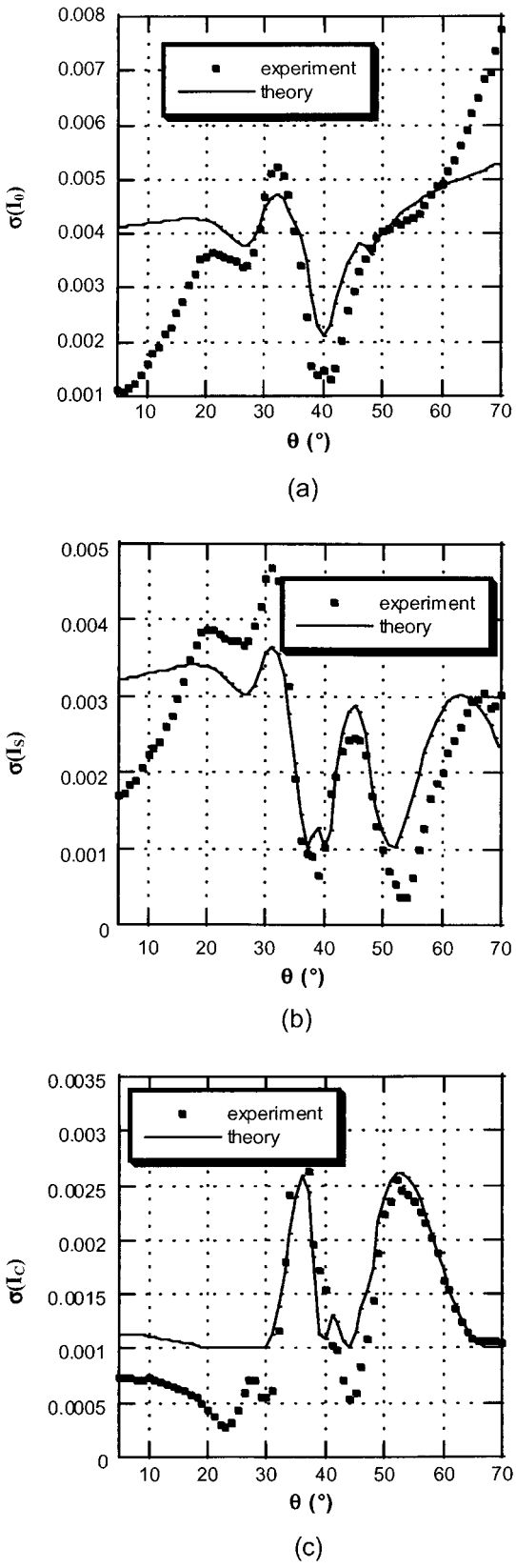


Fig. 2. Empirical error function for the phase-modulator scatterometer. (a) I_0: $\sigma^2 = 10^{-6} + 3.2 \times 10^{-4} I_0$, (b) I_s: $\sigma^2 = 10^{-6} + 2.0 \times 10^{-4} I_s$, (c) I_c: $\sigma^2 = 10^{-6} + 1.4 \times 10^{-4} I_c$. The measurements were performed with the XL6 sample at $\phi = 0°$ and $\lambda = 632.8$ nm.

Table 1. Numerical Results of the Sensitivity Analysis of Sample XL6[a]

Label	Measurement Configuration				lw (Å) (Experiment)	$\sigma(lw)$ (Å) (Theory)	$\sigma(lw)$ (Å) (Experiment)
	ϕ(°)	M(°)	P(°)	A(°)			
1	0	90	45	45	1467.2	4.1	2.1
2	45	90	45	45	1419.0	40.2	29
3	90	90	45	45	1507.4	35	2.9
4	21	90	45	45	1464.8	5.0	4.4

[a]For phase-modulator type measurables

5. Symmetry Properties and Sensitivity Considerations

The application of the SAF formalism can be a very time-consuming procedure. To avoid burdening the computer with unnecessary calculations, we reduced the domain of the degrees of freedom of the experimental arrangement.[5,6] Unlike the ellipsometric scatterometer, for the phase-modulator scatterometer we can make some *a priori* sensitivity considerations, independent of the sample, which further reduce the domain of experimental parameters that we have to scan in search of the optimum sensitivity.

At first sight it seems we have four degrees of freedom for the phase-modulator scatterometer: P, A, M, and ϕ. The measurables of the system, the ratios R_ω and $R_{2\omega}$ from Eqs. (10a) and (10b), are approximately equal to the ratios I_s/I_0 and I_c/I_0. A simple inspection of the I coefficients from Eqs. (8a)–(8c) shows that these ratios are maximum when $P - M = \pm 45°$. To improve the signal-to-noise ratio, we need to set the position of the polarizer relative to the modulator to either $+45°$ or $-45°$ (i.e., $P - M = \pm 45°$).

Although changing M by $\pm 90°$ changes the sign of I_s and I_c but not their amplitude, I_0 remains unchanged. Thus the quantities R_ω and $R_{2\omega}$ change only their the sign, which is inconsequential for the sensitivity. Therefore we can restrict the sensitivity scan to a domain of $M \in (0° \ldots 90°)$.

6. Empirical Error Function

Since the number of measurement configurations to be scanned in the search for the optimum sensitivity is huge, it is necessary to extrapolate measurement errors obtained from some particular measurement configurations to an empirical error function valid for all the measurement configurations. The empirical error function for the phase-modulator scatterometer was evaluated in the same way as it was for the ellipsometric scatterometer.[6] Figure 2 illustrates how we fit the averaged errors for a set of 100 measurements with the empirical function. The function that best approximates the measurement errors is

$$\sigma^2 = 10^{-6} + 5.0 \times 10^{-4} I^2. \qquad (11)$$

7. Sensitivity Analysis for Fitting Predictions and Comparison with Experiment

The SAF formalism was applied to the samples XL6 and P200, described in Table 1 from Ref. 6, for the measurables from Eqs. (10a) and (10b). Figure 3 shows an example of fitting data generated by the phase-modulator scatterometer to theoretical curves

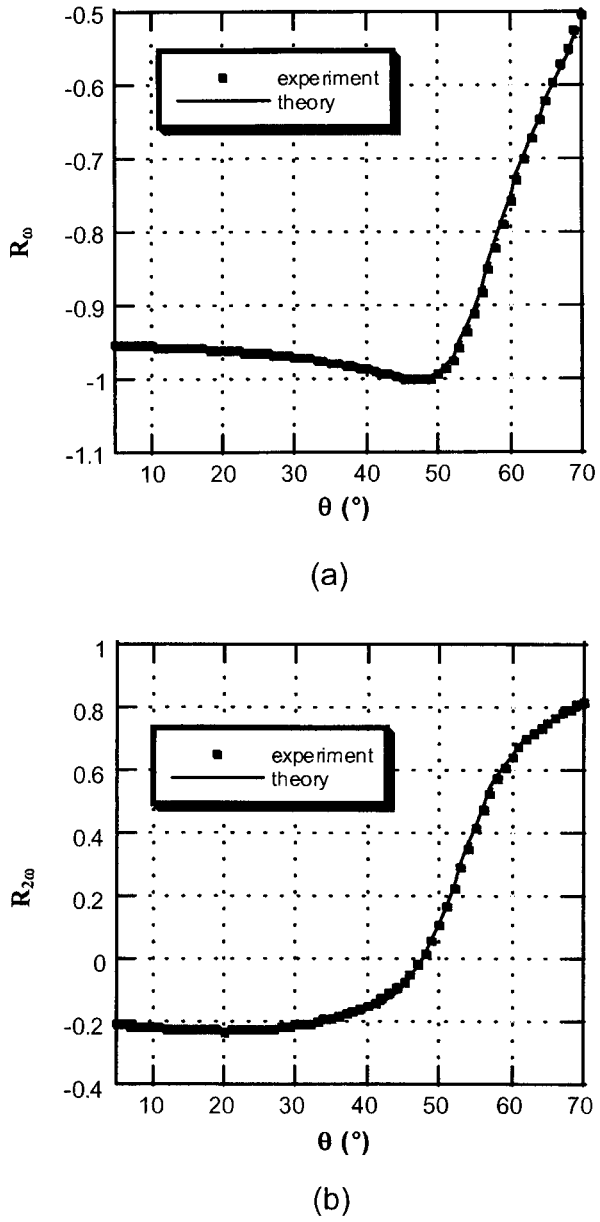


Fig. 3. Example of fitting data from the phase-modulator scatterometer for sample XL6 at $\phi = 90°$ and $\lambda = 632.8$ nm. The results of the fitting were $lw = 1493.1 \pm 1.2$ Å, $h_g = 5568.4 \pm 3.3$ Å, and $h_1 = 1604.3 \pm 1.9$ Å.

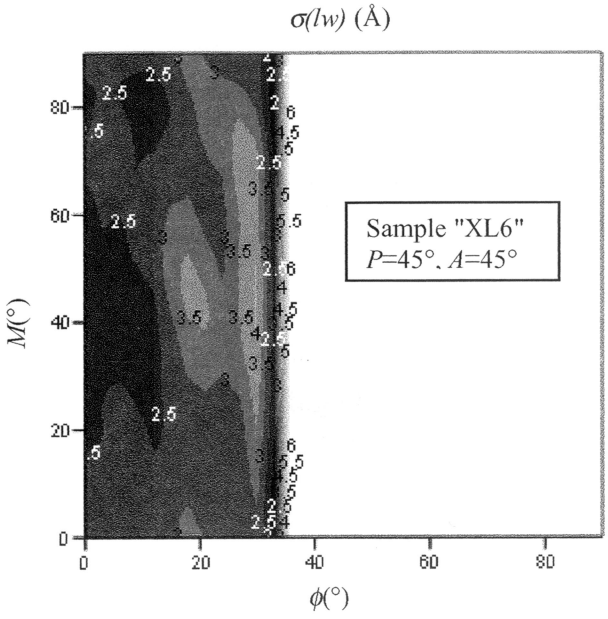


Fig. 4. Sensitivity map from the SAF scan for the phase-modulator scatterometer, sample XL6, at λ = 632.8 nm. Here the color code is as follows: white means large σ (bad sensitivity), black means small σ (good sensitivity).

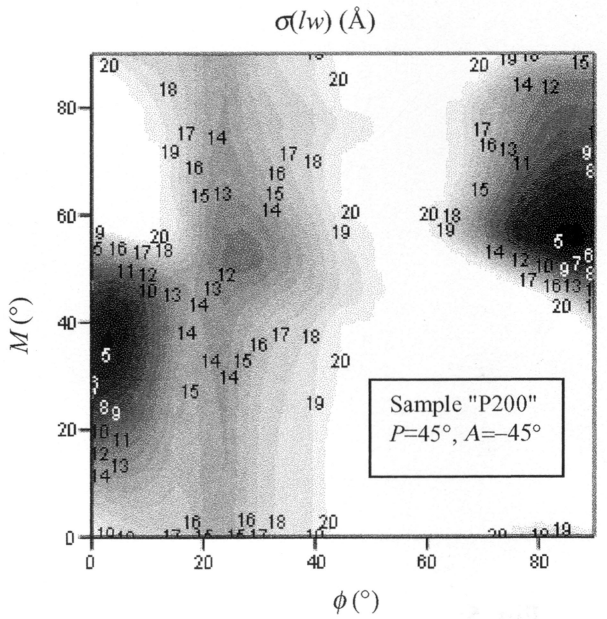


Fig. 5. Sensitivity map from the SAF scan for the phase-modulator scatterometer, sample P200, at λ = 632.8 nm. The color code is the same as in Fig. 4.

calculated with RCWA. Figure 4 illustrates the sensitivity map for sample XL6, with the measurement configurations $P = A = 45°$, among which there is the optimum configuration, and Fig. 5 shows the sensitivity map for sample P200 for configurations type $P = -A = 45°$. The code color for both Figs. 4 and 5 is the same: namely, the darker the hue of gray, the smaller σ.

In Table 1 the numerical results of the sensitivity analysis performed on the sample XL6 for the linewidth *lw* are summarized. We tested experimentally the sensitivity predictions for the measurement configurations and noticed a qualitative agreement between theory and experiment, except for configuration 3. The reason for this discrepancy is because the measurement precision for this configuration was inexplicably better than the measurement precision for the other configurations. The cause of this (fortunate) anomaly is yet to be determined. In Table 2 the numerical results of the sensitivity analysis for the theoretical sample P200 are shown.

The inspection of Tables 1 and 2 does not reveal a significant improvement of the sensitivity compared with that of the ellipsometric scatterometer. This lack of significant improvement was ascribed to the additional sources of errors introduced by a more complicated experimental arrangement. Specifically, the instability of the phase modulation is an important perturbation factor. This assertion is proved by the empirical error function (11) being considerably higher than that for the ellipsometric scatterometer.[6]

Table 2. Numerical Results of the Sensitivity Analysis of Sample P200[a]

	Measurement Configuration				σ(*lw*) (Å)
Label	φ(°)	*M*(°)	*P*(°)	*A*(°)	(Theory)
1	3	36	45	−45	4.8
2	87	56	45	−45	4.7
3	90	90	45	45	18.8
4	24	90	45	−45	14.1

[a]For phase-modulator type measurables

There is, however, an important aspect in Table 1 that makes our effort of developing the phase-modulator scatterometer worthwhile. Namely, we notice what we believe is an increased accuracy. Actually, there is no reason to doubt the accuracy of these results since the dispersion of the experimentally determined values for the linewidth are almost within the theoretical uncertainties. If we ignore the insensitive configurations 2 and 3, we have both a precise and an accurate determination of the linewidth, which is clearly bound to lie within the interval of 1460–1470 Å. The perceived accuracy is the most important improvement shown by the phase-modulation scatterometry technique, compared with previous scatterometry techniques. This accuracy is attributed to the measurables being real-time ratios. The numerator and the denominator are simultaneously measured, in exactly the same area of the photodetector, unlike in the ellipsometric scatterometer, thus eliminating many systematic measurement errors. Contrary to the ellipsometric scatterometer, the normalization measurement S_0 is made in the same conditions as those for S_ω and $S_{2\omega}$.

8. Conclusion

Scatterometry has been a valuable metrology tool for characterizing structures with CDs as small as 100

nm and smaller. To compensate for the loss of sensitivity due to the miniaturization trend in microelectronics, we have designed a methodology for finding the optimum measurement configuration. The method determines the experimental parameters that provide optimum sensitivity of the measurables to changes of the geometric parameters of the sample. This method, which we introduced previously,[4–6] is termed sensitivity analysis for fitting (SAF), and it is a mathematical algorithm that can be applied for optimization purposes regardless of the specific type of measurements that are performed.

Here SAF is applied to a new experimental arrangement that has a new element added to the experimental arrangement of the ellipsometric scatterometer—a phase modulator. This element acts to provide more information regarding grating properties. Additionally, the modulator allows three signals (harmonics of the phase-modulation signal) to be measured simultaneously with lock-in amplifiers, which means that the speed of acquiring information is increased. By normalizing the two superior harmonics to the zeroth harmonic (dc term), we eliminate systematic errors, such as the spatial variations of the sensitivity of the photodetector and laser fluctuations. However, phase-modulation scatterometry has disadvantages. It is more sensitive to residual birefringence of the optical elements of the experiment than the ellipsometric scatterometer, and therefore requires better optics. Because of the residual birefringence and the higher modulation orders induced by the modulator piezoelectric transducer, correction terms must be applied to the measurements. These correction terms have to be calculated prior to any data acquisition, and an easy and accurate method of performing this initial calibration automatically does not exist.

Carefully weighing the advantages and disadvantages, we illustrated that by joining the phase modulator with a mathematical optimization algorithm (SAF) and ellipsometric scatterometry we obtained an improved method that yields superior results. The technique can satisfy tomorrow's metrology needs of the microelectronics industry.

References

1. J. R. McNeil, S. S. H. Naqvi, S. M. Gaspar, K. C. Hickman, and S. R. Wilson, "Optical scatterometry," in *Encyclopedia of Materials Characterization*, C. R. Brundle, C. A. Evans, and S. Wilson, eds. (Manning, Boston, Mass., 1992), pp. 711–722.
2. B. K. Minhas, S. A. Coulombe, S. S. H. Naqvi, and J. R. McNeil, "Ellipsometric scatterometry for the metrology of sub-0.1-μm-linewidth structures," Appl. Opt. **37,** 5112–5115 (1998).
3. S. A. Coulombe, P. C. Logofatu, B. K. Minhas, S. S. H. Naqvi, and J. R. McNeil, "Ellipsometric-scatterometry for sub-0.1 mm CD measurements," in *Metrology, Inspection, and Process Control for Microlithography XII*, B. Singh, ed., Proc. SPIE **3332,** 282–293 (1998).
4. P. C. Logofatu and J. R. McNeil, "Sensitivity analysis of fitting for scatterometry," in *Metrology, Inspection, and Process Control for Microlithography XIII*, B. Singh, ed., Proc. SPIE **3677,** 177–183 (1999).
5. P. C. Logofatu, *Sensitivity Optimized Scatterometry*, Ph.D. dissertation (University of New Mexico, Albuquerque, N. Mex., 2000).
6. P. C. Logofatu, "Sensitivity analysis of grating parameter estimation," Appl. Opt. **41,** 7179–7186 (2002).
7. S. N. Jasperson and S. E. Schnatterly, "An improved method for high reflectivity ellipsometry based on a new polarization modulation technique," Rev. Sci. Instrum. **40,** 761–767 (1969).
8. O. Acher, E. Bigan, and B. Drevillon, "Improvements of phase modulated ellipsometry," Rev. Sci. Instrum. **60,** 65–77 (1989).
9. D. W. Mills, R. L. Allen, and W. M. Duncan, "Spectral ellipsometry on patterned wafers," in *Process, Equipment, and Materials Control in Integrated Circuit Manufacturing*, A. G. Sabnis and I. J. Raaijmakers, eds., Proc. SPIE **2637,** 194–203 (1995).
10. M. G. Moharram, E. B. Grann, and D. A. Pommet, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," J. Opt. Soc. Am. A **12,** 1068–1076 (1995).
11. M. G. Moharram, D. A. Pommet, and E. B. Grann, "Stable implementation of the rigorous coupled-wave theory for surface-relief gratings: enhanced transmittance matrix approach," J. Opt. Soc. Am. A **12,** 1077–1086 (1995).
12. T. C. Oakberg, "Modulated interference effects: use of photoelastic modulators with lasers," Opt. Eng. **34,** 1545–1550 (1995).
13. R. M. A. Azzam and N. M. Bashara, "Generalized ellipsometry for surfaces with directional preference: application to diffraction gratings," J. Opt. Soc. Am. **62,** 1521–1523 (1972).
14. S. Wolfram, *Mathematica, a System for Doing Mathematics by Computer* (Addison-Wesley, Reading, Mass., 1992).

Sensitivity analysis of grating parameter estimation

Petre C. Logofătu

An optimization method for the sensitivity of diffraction efficiency measurements is presented. I define the sensitivity as the estimation precision of the grating parameters. The optimization method called sensitivity analysis for fitting scans all the possible measurement configurations and selects the configuration that yields the best sensitivity. The scan is made over the domain of the experimental parameters of the arrangement, such as the azimuth angle of the grating and the orientation angles of the analyzer and the polarizer. These parameters can be freely varied, and among the multitude of possible combinations there is one configuration that provides optimum sensitivity. Comparison with experimental results reveals a qualitative agreement between theory and practice. © 2002 Optical Society of America

OCIS codes: 050.1950, 120.0120, 260.5430, 120.5700, 260.2130.

1. Introduction

The development of the microelectronics industry has created the need for reliable metrology techniques for the characterization of chip microstructures. Critical elements of the microstructures, such as the critical dimension, the reproducibility, and the uniformity of the microstructures, are areas of concern. Contact metrologies (atomic force microscopy[1]) and noncontact metrologies (scanning electron microscopy[2,3] and scatterometry[4–7]) were developed to fulfill this need. The optical metrology technique, improperly called scatterometry, consists in the estimation of grating parameters from diffraction efficiency measurements. It is a diffraction-based technique and has nothing to do with diffuse scattering, as one may be mislead to think. The experimental arrangement illustrated in Fig. 1 is called a scatterometer. Scatterometers for both commercial and scientific research purposes were developed and are in use. As one can see in Fig. 2, gratings generally diffract multiple orders. However, scatterometry generally limits itself to measuring the 0th reflected order (specular reflection) because most of the reflected power is contained in that order. An additional reason to limit the measurements to the 0th order is that, because of the shrinking geometry of microelectronics samples, oftentimes the 0th order is the only order available for measurement. In this paper I deal with the measurement of the 0th-order reflection only. Scatterometry has the advantages of being noncontact, nondestructive, and it does not require special environmental conditions such as a vacuum. Conventional scatterometry[6,8,9] is limited to only specific values of the experimental parameters illustrated in Fig. 1: the azimuth of the sample ϕ, the polarizer angle P, and the analyzer angle A. For this reason, important information that can be extracted from the sample remains ignored. Another consequence is that there is little room for optimization of the sensitivity of the experimental arrangement. Ellipsometric scatterometry[10,11] is a generalization of conventional scatterometry in which full use of the experimental degrees of freedom of the experimental arrangement is made. By full use I mean that an optimization algorithm called sensitivity analysis for fitting (SAF)[12] was developed to ensure that the scatterometer is in the measurement configuration that provides the best sensitivity. In this paper I undertake a rigorous and formal investigation of the ellipsometric scatterometer and illustrate the application of SAF for real and theoretical samples. The SAF algorithm was presented in its initial form in Ref. 12. In Section 5 I present a modified approach in which the search for sensitivity and stability are done together rather than separately. This modified approach is more time-consuming but its results are much more stable.

P. C. Logofătu (logofatu@unm.edu) is with the Center for High Technology Materials, University of New Mexico, 1313 Goddard Street SE, Albuquerque, New Mexico 87106 and also with the Department of Lasers, National Institute for Laser, Plasma and Radiation Physics, Bucharest 76900, Romania.

Received 20 February 2002; revised manuscript received 16 May 2002.

0003-6935/02/347179-08$15.00/0

© 2002 Optical Society of America

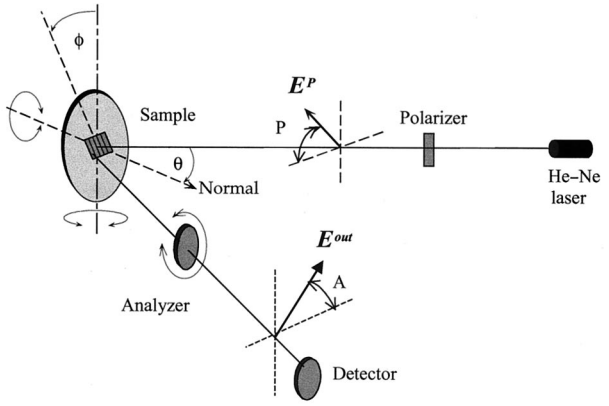


Fig. 1. Experimental arrangement of the ellipsometric scatterometer.

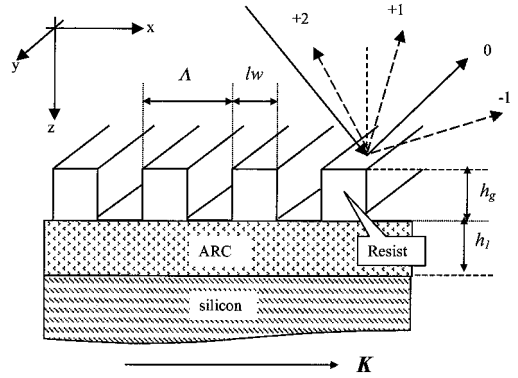


Fig. 2. Example of a test structure for scatterometry studies. A unidimensional grating is placed on top of a stack of thin films ending with a silicon substrate. The materials in the figure are common for microelectronics samples. ARC, ●●●.

Sections 2 and 3 are purely descriptive but contain absolutely necessary definitions of the building elements of the optimization method. Section 4 is the application of the Jones formalism[13] and the generalized ellipsometry[14,15] to the specifics of the scatterometer. The main contribution of this paper, which is the sensitivity analysis formalism called SAF, is outlined in an abstract manner in Section 5 and in a concrete manner in Section 8 where it is applied to the specifics of scatterometry. In Section 5 I show that the sensitivity analysis formalism stands by itself and can be applied to various optimization problems. Section 6 contains theoretical considerations that considerably reduce the calculation time required by the optimization procedure in the specific case of scatterometry. In Section 7 I present an empirical function for the measurement errors, a necessary ingredient for scanning the domain of measurement configurations.

2. Sample Geometry

Scatterometry can be defined as a method of inferring the shape of microstructures from how these structures affect light in reflection. More completely and specifically, scatterometry is a metrology technique used to determine the geometric parameters of microelectronics test structures from efficiency measurements of the light diffracted by these structures. In the sample, the critical features of the chip are reproduced periodically. A schematic of a test sample can be seen in Fig. 2. On a silicon substrate a stack of thin films is deposited and on top of them is placed a rectangular etched grating. The geometric parameters of the test sample in Fig. 2 are as follows. The pitch is the period of the grating and is denoted by Λ. The pitch of the gratings studied was around a few hundred nanometers. The grating vector $\mathbf{K}$ in Fig. 2 points in the periodicity direction, and its modulus is $2\pi/\Lambda$, analogous to a wave vector corresponding to a wavelength Λ. The width of the grating ridge is called the linewidth and is denoted as lw. The height of the grating is noted as h_g, and the rest of the film heights have similar notations. These geometric parameters of the sample, together with the refractive indices of the materials out of which the layers are made, completely characterize the sample. The rectangular profile is not the only profile that a grating can have. However, in this paper I use the rectangular profiles for reasons of simplicity and because they are the most common profiles. Some of the parameters of the sample can be found precisely by means other than scatterometry. For example, the refractive index is usually determined from ellipsometric measurements of unetched samples with smooth surfaces made from the same materials. The pitch is the only geometric parameter that is usually measured independently with good precision by methods other than scatterometry.

3. Experimental Arrangement

In Fig. 1 the experimental arrangement for the ellipsometric scatterometer is shown.[11,12] The diffraction efficiency of the 0th reflected order (the specular reflectance) is measured. The polarizer and the analyzer, with polarization angles P and A that can take any values, allow us to control the polarization of the input and the output. We also have the ability to vary continuously the azimuth ϕ. The measurements are done against the incidence angle θ. The degrees of freedom of the system, or the experimental parameters, are therefore the angles P, A, and ϕ.

Typically we take a set of measurements corresponding to incidence angles θ ranging from 5° to 70° with 1° steps. While the incidence angle is varied, the other experimental parameters are kept constant. The constant experimental parameters define a measurement configuration. To express this definition in a formal manner, we denote the measurable quantities of the scatterometer by

$$q_i = q_i(\mathbf{p}_{\text{geo}}, \mathbf{p}_{\text{opt}}, p_{\text{scan}}, \mathbf{p}_{\text{exp}}),\ i = 1 \ldots N, \quad (1)$$

where $\mathbf{p}_{\text{geo}}$ are the geometric parameters of the sample that we wish to determine (for example, lw, h_g), $\mathbf{p}_{\text{opt}}$ are the optical parameters and other parameters

of the sample that are estimated through alternative means (refractive indices, Λ), p_{scan} is the experimental parameter of the arrangement that we vary in the measurement (usually the incidence angle θ), $\mathbf{p}_{\text{exp}}$ are the experimental parameters of the system that we control and keep constant during the measurements (P, A, and ϕ), and N is the number of measurements. The Bold type indicates an array of quantities. We have, for example, $\mathbf{p}_{\text{geo}} = (p_1, p_2, \ldots, p_M)$. The measurement configuration is defined by $\mathbf{p}_{\text{exp}}$. It is $\mathbf{p}_{\text{exp}}$ that are the degrees of freedom of the scatterometer. It is $\mathbf{p}_{\text{exp}}$ that we explore to find the optimum measurement configuration. Each $\mathbf{p}_{\text{exp}}$ corresponds to a different measurement configuration.

4. Measurables

The Jones matrix calculus[13] was used to obtain the expression of the measurables. The vector electric field output is

$$\mathbf{E}^{\text{out}} = \hat{A}\hat{r}\hat{P}\mathbf{E}^{\text{in}}, \tag{2}$$

or

$$\begin{pmatrix} E_p^{\text{out}} \\ E_s^{\text{out}} \end{pmatrix} = \begin{bmatrix} \cos^2 A & \cos A \sin A \\ \sin A \cos A & \sin^2 A \end{bmatrix} \begin{bmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{bmatrix} \times \begin{bmatrix} \cos^2 P & \cos P \sin P \\ \sin P \cos P & \sin^2 P \end{bmatrix} \begin{pmatrix} E_p^{\text{in}} \\ E_s^{\text{in}} \end{pmatrix}. \tag{3}$$

Here $\hat{A}$ is the Jones matrix of the analyzer, $\hat{P}$ is the Jones matrix of the polarizer, and $\hat{r}$ is the reflection matrix of the sample as defined in generalized ellipsometry.[15,16] Making the calculations in Eq. (3), we obtain

$$\begin{pmatrix} E_p^{\text{out}} \\ E_s^{\text{out}} \end{pmatrix} = rE^P \begin{pmatrix} \cos A \\ \sin A \end{pmatrix}, \tag{4}$$

where

$$r = r_{pp} \cos A \cos P + r_{ps} \cos A \sin P + r_{sp} \sin A \cos P + r_{ss} \sin A \sin P, \tag{5}$$

$$E^P = E_p^{\text{in}} \cos P + E_s^{\text{in}} \sin P. \tag{6}$$

The expression in Eq. (6) is the amplitude of the beam field after passing through the polarizer. In the following I refer to E^P as the input field, because then the ratio output over input takes a more convenient form. The relation in Eq. (5) is fundamental for the study of the ellipsometric scatterometer, containing in a simple, synthetic formula all the information that we need. In practice we measure powers, so we use the square absolute value of Eq. (5). The measurements taken by the ellipsometric scatterometer are normalized to the output power for when the sample is removed from the system. In this case we have $r = 1$. We can conclude that, for the ellipsometric scatterometer, the ratio of output over input is r from Eq. (5) and the measurable is the reflectance

$$R = |r|^2. \tag{7}$$

In Eq. (5) the dependence of the measurable with respect to the polarization angles P and A is explicit. The dependence with respect to the third degree of freedom, the azimuth, is implicit because the reflection coefficients are functions of the azimuth.

5. Sensitivity Analysis for Fitting

For sensitivity analysis purposes we define sensitivity as the precision of the fitting of the standard error for the parameter estimation error,[17]

$$\sigma(p_{\text{geo}}{}^m) = \sqrt{A_{mm}{}^{-1}}\, \Delta\chi_N, \tag{8}$$

where $p_{\text{geo}}{}^m$ ($m = 1, M$) are the geometric parameters to be determined, A is the curvature matrix,

$$A_{k,l} = \sum_{i=1}^{N} \frac{1}{\sigma_i{}^2} \frac{\partial q_i}{\partial p_k} \frac{\partial q_i}{\partial p_l}, \tag{9}$$

σ_i are the measurement errors of the measurables q_i, $i = 1 \ldots N$ (which, in our case, are the reflectance at various incidence angles θ_i), and $\Delta\chi_N$ is a coefficient that determines the probability $P(\Delta\chi_N)$ that the real value of the parameter $p_{\text{geo}}{}^m$ falls within the estimated precision $\sigma(p_{\text{geo}}{}^m)$,

$$P(\Delta\chi_N) = \frac{\Gamma(M/2,\, 0,\, \Delta\chi_N{}^2/2)}{\Gamma(M/2)}, \tag{10}$$

with M the number of parameters to be estimated and Γ the Euler gamma function.[18] With the help of Eq. (8) we can proceed to the sensitivity scanning. Considering Eq. (1), we can define the SAF scan as follows. The optimum measurement configuration $\mathbf{p}_{\text{exp}}{}^{\text{optimum}}$ is that configuration for which we have

$$\sigma_m{}^{\text{optimum}}(\mathbf{p}_{\text{exp}}{}^{\text{optimum}}) = \underset{\mathbf{p}_{\text{exp}}}{\text{Min}} \left\{ \underset{\mathbf{p}_{\text{geo}}}{\text{Max}}[\sigma_m(\mathbf{p}_{\text{geo}}, \mathbf{p}_{\text{exp}})] \right\}, \tag{11}$$

where σ_m is the estimation precision [Eq. (8)] for the geometric parameter $p_{\text{geo}}{}^m$. (The dependence of σ_m on the optical parameters was omitted because they are known and have fixed values.) This formula requires some explanation. In the optimization procedure we first scanned the values of the fitting precision σ_m in the domain of the geometric parameters $\mathbf{p}_{\text{geo}}$ where we suspected the real geometric parameters are to be found, looking for the maximum. Because we did not know the exact values of these parameters, and we wanted good fitting precision for all possible values, we assumed the worst situation in this domain for any measurement configuration $\mathbf{p}_{\text{exp}}$. This first part ensures that the measurement configuration is stable for changes of the geometric parameters. Second, we scanned the values of the fitting precision in the domain of $\mathbf{p}_{\text{exp}}$, or the domain of the experimental degrees of freedom searching for the minimum. This part ensures the optimization of the sensitivity. The experimental parameters $\mathbf{p}_{\text{exp}}{}^{\text{optimum}}$ corresponding to this minimum define the optimum measurement configuration that provides

us with good sensitivity even for the most unfavorable case of $\mathbf{p}_{geo}$ values.

We wanted to verify that the predictions of SAF are accurate. For this purpose we analyzed the distribution of the results for the fitted parameters to see if it agrees with the distribution assumed in Eq. (10). Then we determined if the precision given by theory agrees with the precision that the experimental error distribution indicates.

6. Symmetry Properties that Reduce the Domain of Experimental Parameters

We have determined that the ellipsometric scatterometer has three degrees of freedom: the polarizer orientation angle, the analyzer orientation angle, and the azimuth. The optimization procedure is time-consuming, and it is a priority to minimize the calculation time. We sought to reduce the three-dimensional domain of the degrees of freedom that has to be scanned by eliminating the redundant measurement configurations. A rectangular grating has rotation symmetry C_{2z} (Ref. 19) because it remains unchanged after 180° rotation in azimuth. The grating also has a reflection symmetry relative to the plane xz, or σ_y.[19] In other words, switching ϕ by $-\phi$ changes nothing. (Actually, because of some sign conventions, there are some sign changes in the reflection matrix; however, careful management of these signs should avoid any confusion[20].) We can then restrict the azimuth to the first quadrant because every quadrant simply repeats the information contained in the others. This is true regardless of the actual measurables of the experimental arrangement because it is an intrinsic property of the grating and independent of the rest of the experimental system. Turning to the degrees of freedom of the ellipsometric scatterometer (P, A) that are intrinsic to the arrangement, one can further reduce the domain of nonredundant measurement configurations. Taking into account the identity of the cross-reflection coefficients,[19,21] we can rewrite Eq. (5) as

$$r = r_{pp} \cos A \cos P + r_{ps} \sin(P - A) + r_{ss} \sin A \sin P. \quad (12)$$

Because the measurable is the square modulus of Eq. (12), a sign change of it will have no effect on the reflectance. Note that a change of $\pm 180°$ of P or A transforms r into $\pm r$, hence leaving reflectance unchanged. Actually, this is the simple consequence of the fact that the Jones matrix of a polarizer is left unchanged by a 180° rotation.

Another symmetry is provided by the transformation

$$P \rightarrow -A, A \rightarrow -P. \quad (13)$$

The transformation in relation (13) leaves r unchanged, and it is equivalent to a mirror reflection in the (P, A) plane relative to the line $P + A = 0$ in this plane. The redundancy of the measurement configurations is illustrated in Fig. 3. A map of the scan of the entire domain of experimental parameters shows multiple periodicity and symmetries. The first two transformations show that the reflectance has a periodicity of 180° with respect to P and A. The transformation in relation (13) shows that R is symmetric for a mirror reflection in rapport with the line $P + A = 0$. Each of these three transformations reduce the initial domain of (P, A) to half. Overall, the three transformations reduce the domain of meaningful degrees of freedom to one eighth of the initial area. By convention we chose among multiple available choices this eighth portion of the initial domain to be the area

$$P + A \in (0°, 180°), \quad P - A \in (0°, 180°), \quad (14)$$

corresponding to the diamond-shaped area in Fig. 3. For convenience, we work with the normal coordinates $P_n = P + A$ and $A_n = P - A$.

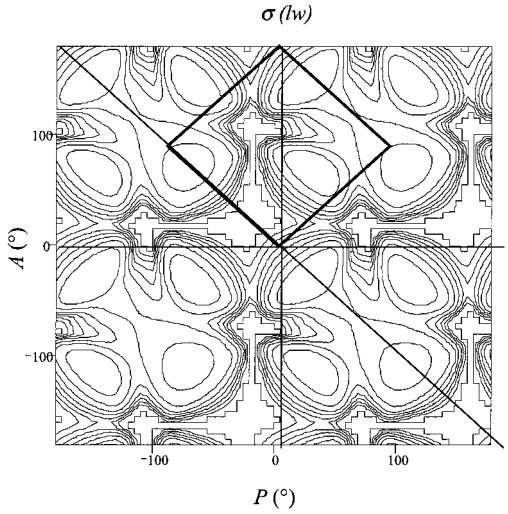


Fig. 3. Map of a sensitivity (σ) scan of the whole range of polarization angles illustrating the redundancy of the measurement configurations. The redundancy occurs no matter what particular sample and azimuth correspond to the map.

7. Empirical Error Function

The sensitivity analysis has to be performed for all the possible measurement configurations $\mathbf{p}_{exp}$. To do this we have to know *a priori* the measurement errors σ_i from Eq. (9), corresponding to all the measurement configurations. Because this is not possible, we have to extrapolate, to calculate the measurement errors in certain configurations, to infer an empirical formula from them that we then extrapolate to all the other measurement configurations.

To determine the empirical measurement error function of the scatterometer,[12,20] I took statistical batches of measurements from various samples and in various measurement configurations. In the particular case shown in Fig. 4, the batch consists in 50 curves; the measurement configuration is $\phi = 90°$, $P = 0°$, and $A = 0°$. I varied the initial conditions, by realigning the system and allowed significant time intervals between every ten sets of measurements.

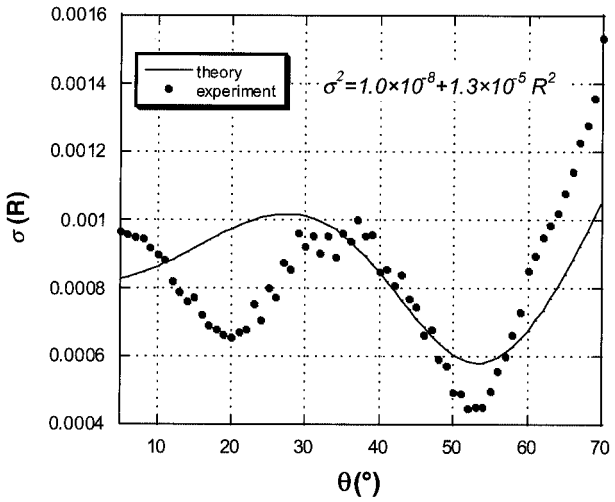


Fig. 4. Measurement precision (dots) fitted by the empirical error function for the ellipsometric scatterometer (solid curve). In the figure are shown the coefficients for the empirical error function, which here are the coefficients of the best fit. The dots are the average values for a batch of 50 measurement sets in the configuration $\phi = 90°$, $P = 0°$, and $A = 0°$ at $\lambda = 632.8$ nm.

I assumed various perturbation factors, such as power fluctuations of the laser, mechanical vibrations that cause fluctuations of the incidence angle, finite reading resolution, and quantified these factors as

$$\sigma^2 = \sigma_{\text{res}}{}^2 + \sigma_{\text{pow}}{}^2 R^2 + \sigma_{\text{vib}}{}^2 (\partial R/\partial\theta)^2. \qquad (15)$$

I calculated the coefficients σ_{res}, σ_{pow}, and σ_{vib} by fitting the actual measurement errors with the formula of Eq. (15). I did not find conclusive correlation of the errors with the partial derivative of the reflectance with respect to the incidence angle, so I eliminated the hypothesis of the mechanical vibrations as a meaningful perturbation factor. In addition, I was not able to find a good correlation for even the remaining perturbation factors. This might have been caused by insufficient size of the statistical batch, incorrect randomization of the experimental parameters of the system, or simply by other perturbation factors that were overlooked. However, an overall correlation of the errors with the finite-resolution reading and the power fluctuation factors can be detected, as one can see in Fig. 4. I generally assumed the worse. I took exceptions when the fitted factor was obviously erroneous. Finally I arrived at the empirical error function that I used in the SAF calculations for samples measured with the ellipsometric scatterometer expressed as

$$\sigma^2 = 1.0 \times 10^{-8} + 1.3 \times 10^{-5} R^2. \qquad (16)$$

8. Sensitivity Analysis for Fitting Predictions and Comparison with Experimental Results

The SAF formalism was applied to the two samples described in Table 1. I worked with He–Ne lasers only at the wavelength $\lambda = 632.8$ nm. For sample XL6, a real sample, I was able to compare the experimental results for the actual error with those for the predicted error, as explained at the end of Section 5. For the sample P200, a theoretical sample, I give only the theoretical predictions. The results for XL6 are summarized in Table 2. The values of the estimation precision σ of the linewidth, theoretical and experimental, were calculated for a confidence level of 95%, which corresponds to a $\Delta\chi$ of 2.79548. By convention, the estimation precision was always calculated for a probability of 95%.

Figure 5 illustrates the fitting of a particular set of data taken from sample XL6. Figure 6 shows an example of a SAF scan for the optimum sensitivity for the sample XL6, and Fig. 7 illustrates another example for the sample P200. The maps in Figs. 6 and 7 are sections with

$$\phi = \text{const of } \underset{\mathbf{p}_{\text{geo}}}{\text{Max}}[\sigma_m(\mathbf{p}_{\text{geo}}, \mathbf{p}_{\text{exp}})]$$

Table 1. Geometric and Optical Parameters of the Samples Referred to in Section 5

Sample	Λ (Å)	lw (Å)	h_g (Å)	h_1 (Å)	n_g	n_1	n_s
XL6	3628.8	1470	5700	1600	1.64	1.66	$3.882 - 0.019i$
P200	2000	700	4000	—	1.64	—	$3.882 - 0.019i$

Table 2. Numerical Results for the Sensitivity Analysis of the Sample XL6 for Ellipsometric Scatterometer Measurements at λ = 632.8 nm

	Measurement Configuration					
Label	φ(°)	P(°)	A(°)	lw (Å)	σ(lw) (Å) (Theory)	σ(lw) (Å) (Experiment)
1 (conventional)	0	0	0	1519.8	1.8	4.9
2 (conventional)	0	90	90	1451.6	0.6	6.3
3 (conventional)	90	0	0	1536.5	15.0	19.5
4 (conventional)	90	90	90	1390.0	8.4	7.6
5 (optimum)	24	−63	63	1494.4	0.5	2.6

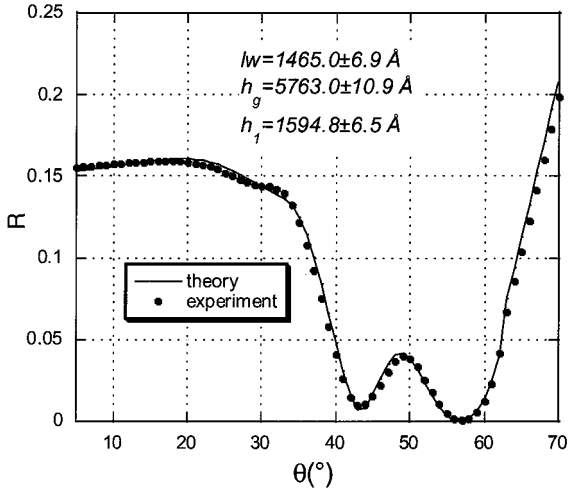


Fig. 5. Example of the fitting of a curve measured by the ellipsometric scatterometer in the configuration $\phi = 24°$, $P = -63°$, and $A = 63°$ at $\lambda = 632.8$ nm. In the plot the geometric parameters of the sample (parameters of the best fit) are shown.

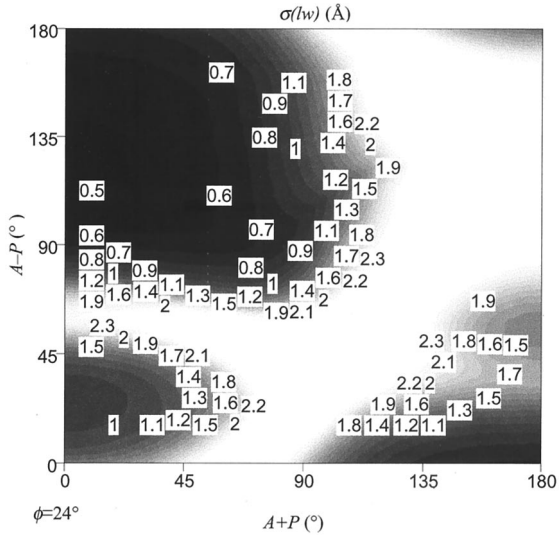


Fig. 6. Illustration of the sensitivity (σ) scan for the sample XL6 at λ = 632.8 nm. In the figure is a section of the investigated domain for ϕ = 24°. The optimum measurement configuration is ϕ = 24°, P = 63°, and A = 117°.

from Eq. (11) in the domain of degrees of freedom of the ellipsometric scatterometer. These sections were chosen because they contain the optimum configuration $p_{\text{exp}}^{\text{optimum}}$ defined also in Eq. (11). In Figs. 6 and 7, the dark hues mean low values of σ, whereas light hues are large values. We are interested therefore in the darker regions of the map.

We can see that Fig. 7 is symmetric relative to the diagonal line. This happens because the azimuth is 0° and we have additional symmetry properties because r_{ps} is zero. For an azimuth different from 0° or 90°, the degeneracy is removed, and it was not considered necessary to make a special case out of these two situations.

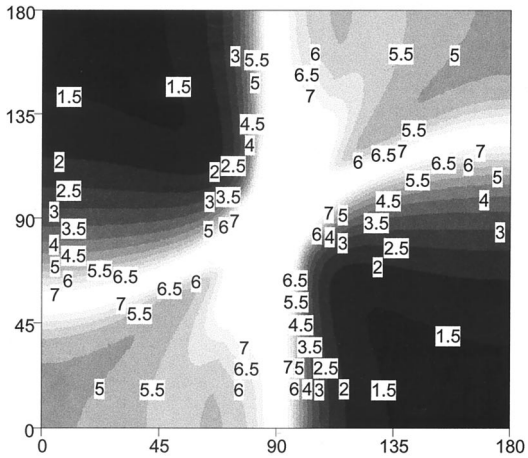


Fig. 7. Illustration of the sensitivity (σ) scan for the sample P200 at λ = 632.8 nm. In the figure is a section of the investigated domain for ϕ = 0°. The optimum measurement configuration is ϕ = 0°, P = −78°, and A = 102°. The lower corners of the plot correspond to the conventional measurement configurations 1 and 2 from Table 3.

Inspection of the last two columns of Table 2 shows that there is not a rigorous agreement between experiment and theory. However, we do have a qualitative agreement. Importantly, the measurement configuration that was predicted by the SAF formalism to be optimum is indeed the optimum configuration.

Configurations 1–4 are called conventional because they are in use in conventional scatterometry discussed in Section 1. For these configurations the conical diffraction simplifies to diffraction in a plane, and the calculations of the diffracted fields are simpler. Because the calculation time is an issue for scatterometry, these configurations are used more often than the rest.

The fact that we are able to have subnanometer precision can be considered a success. However, some concerns are raised by the large discrepancy among the results for the linewidth obtained through different measurement configurations. The discrepancy is larger than the precision, theoretical or experimental, displayed in Table 2, which leads us to the conclusion that we are strongly affected by systematic errors. Systematic errors could be due to either internal or external causes. Internal causes might be from the simplified modeling of the grating (a rectangular profile was assumed, but the real profile may be more complicated) and nonuniformity of the grating. External causes might be deficient alignment, inaccurate optical parameters, imperfection of the optical elements used in the system, and the fact that the ellipsometric scatterometer makes the normalization measurement in different conditions than the regular measurements. Also, there could be factors that I was unable to detect. However, the situation is reasonable for two reasons. First, to a limit, the precision and the accuracy can be uncoupled. Second, the industry is more interested in precision that in accuracy. The manufacturers want to know how well their structures are reproduced rather than the exact, absolute dimensions of these structures. Taking into account these two considerations, we can ignore up to a point the discrepancies shown in the column of the linewidth and focus on the precision columns. These columns indicate a satisfactory situation. Of course, a solution to the inaccuracy problem in Table 2 would be welcomed.

Predictions for a sample with small pitch and linewidth were calculated and are shown in Table 3 because it is of interest to see if scatterometry and the SAF method are still applicable when the microelectronics features are further decreased. The prediction for the optimum sensitivity is not as good as for the sample XL6, but the precision is still acceptable. The conventional configurations 1 and 2 from Table 3 correspond to the two lower corners of Fig. 7. Note that the optimum configuration is not much better than a conventional configuration. In this case the optimization procedure did not bring any improvement. In the case of two-parameter estimation, 95% probability corresponds to a $\Delta\chi$ of 2.44775.

Table 3. Numerical Results for the Sensitivity Analysis of the Sample P200 for Ellipsometric Scatterometer Measurements at $\lambda = 632.8$ nm

Label	Measurement Configuration $\phi(°)$	$P(°)$	$A(°)$	$\sigma(lw)$ (Å) (Theory)
1 (conventional)	0	0	0	5.0
2 (conventional)	0	90	90	1.4
3 (conventional)	90	0	0	1.5
4 (conventional)	90	90	90	8.4
5 (optimum)	0	−78	102	1.4

9. Conclusion

In this paper scatterometry was investigated from the point of view of sensitivity to provide methods to assess optimum measurement configurations. The primary contribution of this paper is the SAF, a formalism that analyzes the sensitivity of the fitting for scatterometry data in a straightforward and rigorous manner. The formalism is not restricted to scatterometry. It is presented in the most general form but was designed for scatterometry. The first important feature of SAF is the statistical analysis that provides an analytic expression to determine precision in terms of the partial derivatives of the measurable with respect to the parameters to be determined. This expression also involves the measurement precision, which has to be known *a priori*. Extrapolation of the error patterns found in certain measurement configurations into an empirical error function for all the possible measurement configurations is required. The most important and original aspect of SAF is the scanning of the sensitivity of the measurement configurations over the domain of the degrees of freedom of the experimental arrangement in search for the optimum configuration. Once this optimum configuration is found, a comparison with experimental results is undertaken. A qualitative agreement between the theoretical predictions for sensitivity of SAF and actual experimental results was found. The optimum configuration is sample dependent.

The SAF method was applied to the ellipsometric scatterometer, an improved version of the conventional scatterometer. The experimental arrangement contains polarizers with variable polarization angles that, together with the azimuth of the sample, constitute the degrees of freedom of the arrangement. The incidence angle is the varied parameter in the measurement curves generated by the scatterometer. In the examples studied, SAF revealed that scatterometry can provide the resolution required by the microelectronics industry, with the 632.8-nm He–Ne laser, provided that the optimum measurement configuration is utilized. Despite the good precision and reproducibility demonstrated by individual measurement configurations, some discrepancy among the results provided by them was noted and attributed to possible sources of systematic errors, intrinsic or extrinsic to the sample. Because the microelectronics industry is primarily interested in reproducibility rather than accuracy, the discrepancy was deemed tolerable to the extent it does not interfere severely with the estimation precision of the geometric parameters of the sample. Symmetry considerations allowed us to focus only on the nonredundant measurement configuration and considerably reduced the amount of work required by the optimization procedure.

References

1. L. E. Ocola, D. S. Fryer, G. Reynolds, A. Krasnoperova, and F. Cerina, "Scanning force microscopy measurements of latent image topography in chemically amplified resists," Appl. Phys. Lett. **68,** 717–719 (1996).
2. M. Postek and D. Joy, "Submicrometer microelectronics dimensional metrology: scanning electron microscopy," J. Res. Natl. Bur. Stand. **92,** 205–228 (1987).
3. A. Sicignano and M. Vaez-Iravani, "Precision metrology of integrated circuit critical dimensions using *in situ* differential scanning electron microscopy," Scanning **10,** 201–206 (1988).
4. I. Kallioniemi, J. Saarinen, and E. Oja, "Characterization of diffraction gratings in a rigorous domain with optical scatterometry: hierarchical neural-network model," Appl. Opt. **38,** 5920–5930 (1999).
5. S. Hava and M. Auslender, "Optical scatterometry evaluation of groove depth in lamellar silicon grating structures," Opt. Eng. **40,** 1244–1248 (2001).
6. J. R. McNeil, S. S. H. Naqvi, S. M. Gaspar, K. C. Hickman, and S. R. Wilson, "Optical scatterometry," in *Encyclopedia of Materials Characterization*, C. A. Evans, C. R. Brundle, and S. Wilson, eds. (Manning, Boston, 1992).
7. C. J. Raymond, J. R. McNeil, and S. S. H. Naqvi, "Scatterometry for CD measurements of etched structures," in *Metrology, Inspection, and Process Control for Microlithography X*, S. K. Jones, ed., Proc. SPIE **2725,** 720–728 (1996).
8. J. R. McNeil, S. S. H. Naqvi, S. M. Gaspar, K. C. Hickman, K. P. Bishop, L. M. Milner, R. H. Krukar, and G. A. Petersen, "Scatterometry applied to microelectronic processing," Microlithogr. World **1,** 16–22 (1992).
9. B. K. Minhas, S. L. Prins, S. S. H. Naqvi, and J. R. McNeil, "Toward sub-0.1-μm CD measurements using scatterometry," in *Metrology, Inspection, and Process Control for Microlithography X*, S. K. Jones, ed., Proc. SPIE **2725,** 729–739 (1996).
10. B. K. Minhas, S. A. Coulombe, S. S. H. Naqvi, and J. R. McNeil, "Ellipsometric scatterometry for the metrology of sub-0.1-μm-linewidth structures," Appl. Opt. **37,** 5112–5115 (1998).
11. S. A. Coulombe, P. C. Logofatu, B. K. Minhas, S. S. H. Naqvi, and J. R. McNeil, "Ellipsometric scatterometry for sub-0.1-mm CD measurements," in *Metrology, Inspection, and Process Control for Microlithography XII*, B. Singh, ed., Proc. SPIE **3332,** 282–293 (1998).
12. P. C. Logofatu and J. R. McNeil, "Sensitivity analysis of fitting for scatterometry," in *Metrology, Inspection, and Process Control for Microlithography XIII*, B. Singh, ed., Proc. SPIE **3677,** 177–183 (1999).
13. E. Collett, *Polarized Light: Fundamentals and Applications* (Marcel Dekker, New York, 1993).
14. R. M. A. Azzam and N. M. Bashara, "Polarization characteristics of scattered radiation from a diffraction grating by ellipsometry with applications to surface roughness," Phys. Rev. B **5,** 4721–4729 (1972).
15. R. M. A. Azzam and N. M. Bashara, "Generalized ellipsometry for surfaces with directional preference: application to diffraction gratings," J. Opt. Soc. Am **62,** 1521–1523 (1972).
16. R. M. A. Azzam, "Application of generalized ellipsometry to anisotropic crystals," J. Opt. Soc. Am. **64,** 128–133 (1974).

17. P. R. Bevington and D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, Boston, 1992).
18. G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists* (Academic, San Diego, Calif., 2001).
19. L. Li, "Symmetries of cross-polarization diffraction coefficients of gratings," J. Opt. Soc. Am. A **17,** 881–887 (2000).
20. P. C. Logofatu, "Sensitivity-optimized scatterometry," Ph.D. dissertation (University of New Mexico, Albuquerque, N. Mex., 2000).
21. P. C. Logofatu and J. R. McNeil, "Identity of the cross-reflection coefficients for symmetric surface-relief gratings," J. Opt. Soc. Am. A **16,** 1108–1114 (1999).

UV Scatterometry

Petre C. Logofătu*
University of New Mexico, Center for High Technology Materials, Albuquerque, NM 87106

ABSTRACT

In this article are demonstrated the nanoscale capabilities of scatterometry, an optical metrology technique for the determination of the grating parameters. Notably, for both ellipsometric-scatterometry and phase-modulation scatterometry we are now currently using shorter wavelength laser sources, namely 325 nm in UV. We added to the scatterometer the ability to perform ϕ-scanning measurements. This capability adds new degrees of freedom to the sensitivity optimization procedure.

Because the sensitivity analysis formalism leads us to the conclusion that sensitivity is provided mainly by anomalies, a rigorous analysis of a certain type of anomaly, the most likely to occur for our type of samples, was made using Modal Analysis. Our analysis allows for the prediction of the anomaly position. This procedure allows us to find the optimum measurement configuration much faster than the SAF formalism used in the past.

Keywords: scatterometry, sensitivity, grating anomalies, ellipsometry, phase modulation

1. INTRODUCTION

The development of nanolithography requires a nanoscale metrology as well for the characterization of the microscopic structures. Specifically, we use scatterometry, an optical metrology technique, for characterization of surface-relief gratings used as test structures. There is a long history of research dedicated to scatterometry at the University of New Mexico [1-14]. In Fig. 1 a typical test structure is presented and the sample parameters are defined. In Fig. 2 we have the schematics of an ellipsometric scatterometer, the experimental arrangement that we used in our measurements. There are two version in use of this scatterometer, one with variable incidence angle (θ) and another with variable azimuth (ϕ). Another two versions of the scatterometer are obtained if we add in the experimental arrangement a phase-modulator (represented with faded hues in Fig. 2). Then we have a phase-modulation scatterometer with incidence angle scanning and a phase-modulation scatterometer with azimuth scanning. Scatterometry

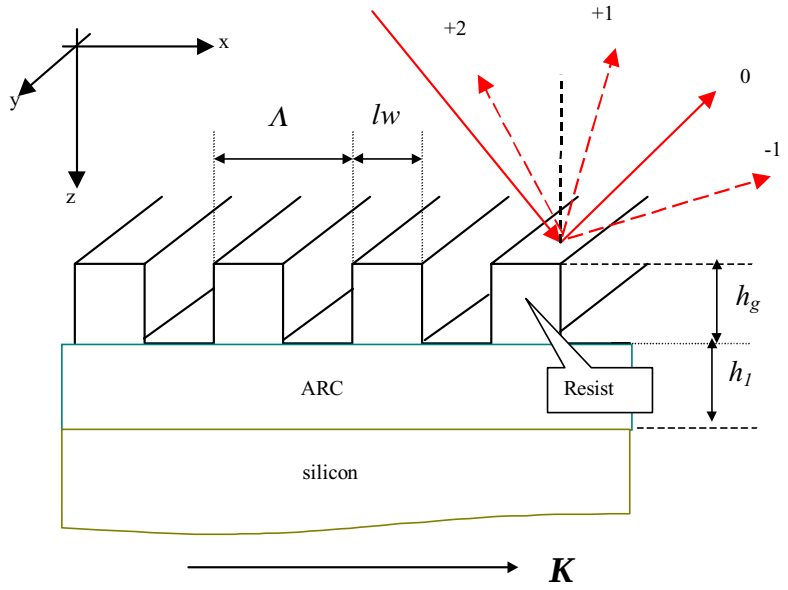


Fig. 1. Example of test structure suited for scatterometry studies. A unidimensional grating is placed on top of a stack of thin films ending with a silicon substrate. The materials in the figure are common for microelectronics samples The grating parameters in the figure are the linewidth *lw*, the grating height h_g, the pitch Λ, and the ARC height. Several diffraction orders are emitted from the sample after reflection due to the periodic nature of the grating.

* Correspondence: Email: logofatu@chtm.unm.edu; Telephone 505 272 7808; Fax 505 272 7801
Permanent address: National Institute of Laser, Plasma and Radiation Physics (INFLPR), Bucharest-Măgurele, PO-Box MG-36, 76900, Romania

determines from reflectivity measurements the parameters of the gratings. Among these geometrical parameters the most important is the critical dimension (CD), or the linewidth (lw) of the grating ridges. The determination is done by fitting reflectance type measurables to theoretical curves calculated with the Rigorous Coupled-Wave Analysis (RCWA) or Modal Analysis.

2. AZIMUTH-SCANNING SCATTEROMETER

In the past scatterometry measurements were done in our lab using only incidence angle (θ) scanning [7, 8, 13]. The sample holder (see Fig. 2) was rotating about its vertical axis and the detector's arm was moving with twice the speed (2 θ) about the same vertical axis of the sample holder so that the spot of the reflected specular order stays on the detector. Recently we added to the scatterometer the ability to rotate about the normal axis to its surface by attaching the sample holder to a rotation stage. This addition allows for making azimuth (ϕ) scanning measurements. There are some obvious advantages for this technique. One advantage is that the scatterometer is mechanically more stable. Only the sample holder is rotating as opposed to θ-scanning where both the sample holder and the detector were moving. Mechanical vibrations are an important source of random measurement errors and the drift of the spot's position on the detector is an important source of systematic errors. Both these types of errors are reduced by using ϕ-scanning. Another advantage is the fact that now we have more choices of measurement configurations and therefore the optimization procedure of the measurement configuration using the Sensitivity Analysis for Fitting (SAF) formalism[1, 5, 6] is likely to yield superior results compared to the past. We have a whole new class of ϕ-scanning measurement configurations and one of them may be more sensitive than the θ-scanning configurations that in the past were the only class of measurement configurations that could be investigated.

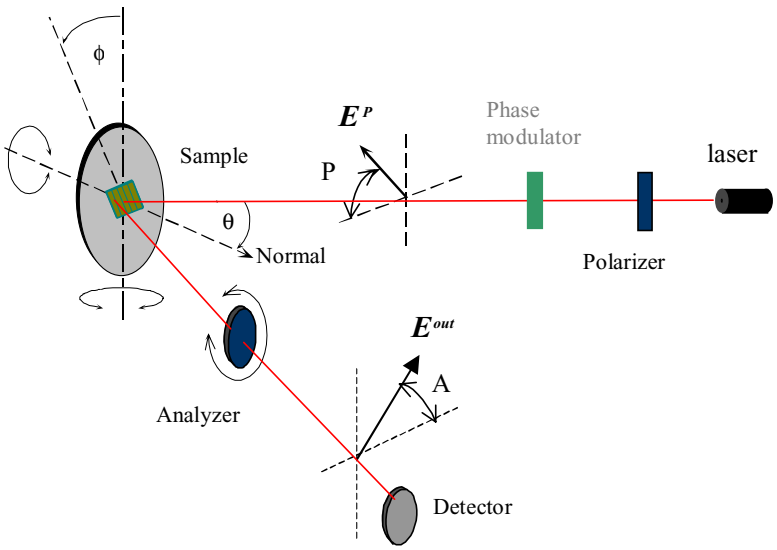
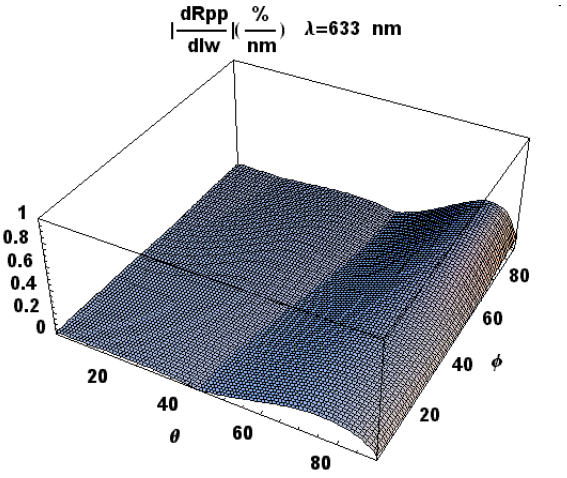
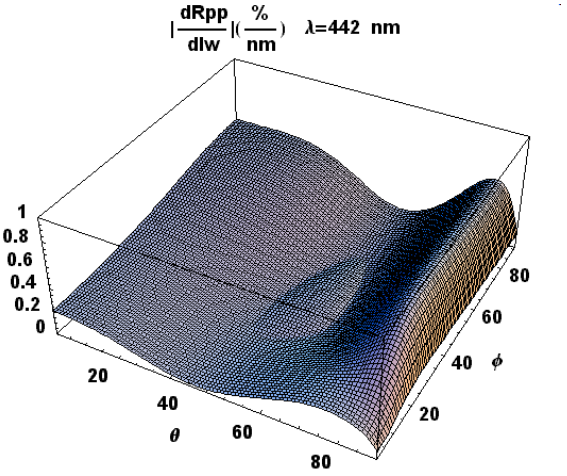


Fig. 2 Schematic description of the ellipsometric scatterometer. Two versions of this scatterometer are in use: the incident angle (θ) and the azimuth (ϕ) scanning scatterometers. The addition of a phase modulator (represented in faded hues) converts the experimental arrangement into the phase-modulation scatterometer.

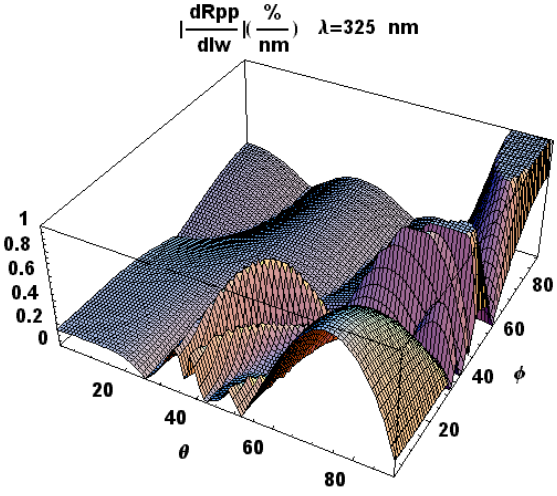
Besides our desire to reduce random and systematic measurement errors there is another reason why we added this new capability to the scatterometer. In converting the scatterometer to measurements at UV wavelengths one of the difficulties we encountered was the large size of the detectors able to measure UV light. Only photomultipliers tubes (PMT) fulfill all the requirements for UV scatterometry, such as quick reaction time, sensitivity and large frequency bandwidth. Their disadvantage is their large size, which makes difficult to move the detector's arm. We needed a measurement configuration in which the detector is immobile and this measurement configuration is ϕ-scanning.



(a)



(b)



(c)

Fig. 3 The sensitivity increases when the wavelength decreases.

3. UV SCATTEROMETRY

In the past we used successfully He-Ne lasers (λ=633 nm wavelength, red color light) as light sources for the scatterometry optical measurements. He-Ne lasers have many useful properties such as power stability and good coherence and they have low cost. These properties are necessary for good measurement precision. Other light sources generally do not have such high power stability and coherence and they are more expensive. For UV wavelengths domain there are supplementary problems such as lack of high quality optical components (polarizers and phase retarders) that are integral components of the scatterometer (see Fig. 2). However, light sources of shorter wavelengths have the advantage of making the measurables more sensitive with respect to changes of the gratings parameters. The determination precision of the grating parameters is made up by the measurement precision of the measurables and the sensitivity of the measurables with respect to changes of the grating parameters, which can be uncoupled to a certain extent [3]. Going to shorter wavelength causes the measurement precision to decrease, as is explained above and is shown quantitatively in Eq. 1 below.

$$\sigma^2 = 1.0\times10^{-8} + 1.3\times10^{-5}\,R^2\,,\ \lambda=633\text{ nm}, \tag{1.1}$$

$$\sigma^2 = 2.0\times10^{-6} + 1.0\times10^{-4}\,R^2\,,\ \lambda=442\text{ nm}, \tag{1.2}$$

$$\sigma^2 = 1.0\times10^{-5} + 5.5\times10^{-4}\ R^2\ ,\ \lambda=325\text{ nm}. \tag{1.3}$$

where R is the measurable, generally a reflectometric quantity. The sensitivity, however, increases, and at a certain degree of miniaturization of the gratings the growth in sensitivity compensates and surpasses the loss in measurement precision, as one can see in Fig. 3. The miniaturization trend in microlithography eventually forced us to convert the scatterometer to shorter wavelengths.

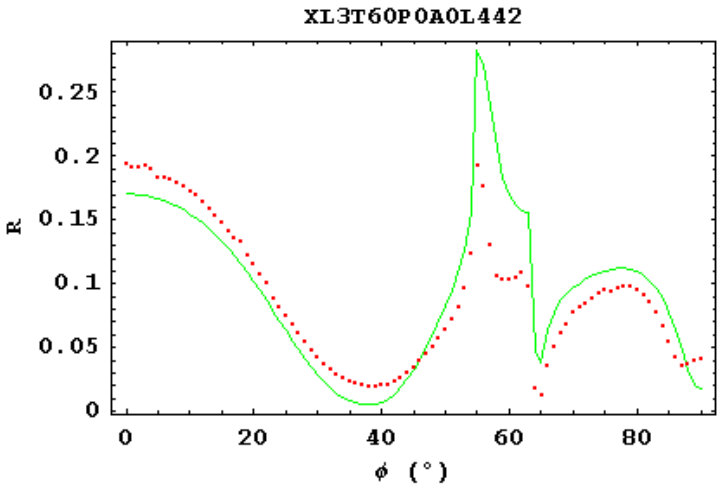


Fig. 4 Example of fitting of data obtained through UV scatterometry measurements (λ=325 nm). The measurement configuration is ϕ-scanning at θ=60°, input polarization P=0°, output polarization A=0°. The results of the fitting are lw=117.9 nm, h_g=566.6 nm, h_{ARC}=174.3 nm. The experimental data is plotted as a dotted curve and the best theoretical fit as continuous curve.

First we went down to 442 nm (blue color light), using a He-Cd laser. This conversion does not require dramatic changes in the experimental arrangement. The polarizers used for He-Ne, being broad range, could still be used at 442 nm. We had to purchase new phase retarders though, but their quality was comparable with the quality of the phase retarders for 632 nm.

It turns out, however, in the conditions of ever shrinking grating geometry, for samples with pitches below 200 nm, even a wavelength of 442 nm may not be enough. The condition for having good sensitivity, as it will be detailed in the next section, is to have grating anomalies, and these anomalies appear when the wavelength is at most twice the pitch. Therefore we had to go further on the wavelength scale down to the UV domain, namely to 325 nm, another line of the He-Cd laser. Here we have encountered supplementary difficulties. Both the polarizers and the phase retarders had to be changed and custom made for 325 nm and they don't have the high quality of the optical components for visible light and they are expensive. We had to replace also the silicon photodiode detector with a photomultiplier tube. The large size of the PMT requires keeping the detector arm immobile and this forced us to add azimuth rotation capability to the scatterometer. We have an illustration of UV scatterometry application in Fig. 4. The sample examined here was labeled "XL3" and the results of the fitting are shown in the caption.

4. GRATING ANOMALIES

One important concrete aspect in the sensitivity considerations is the presence of grating anomalies. There are various types of anomalies but they all cause a rapid variation of the light reflectance with respect to the grating parameters, which is another way of saying that they increase the sensitivity of the scatterometry measurements. Actually the main concrete reason for the loss of sensitivity when the geometry of microlithographic structures shrinks is precisely the absence of anomalies. The search for optimum sensitivity, done according to the SAF algorithm shows that the optimum sensitivity is reached when the θ-scan or ϕ-scan intersects as much as possible an anomaly, as seen in Fig. 5. By decreasing the wavelength down to the UV domain, the anomalies reappear and the sensitivity is boosted up. There is one specific type of anomaly with which we concerned ourselves, the last to disappear when the geometry is shrinking. This anomaly is the most likely to help in boosting the sensitivity in the conditions of small pitch samples. Moreover this anomaly cause a sharper variation of the reflectance versus the grating parameter than the anomaly caused by the emergence of the first reflection order. This is the same as saying that this anomaly has a stronger impact on the sensitivity than the first order emergence anomaly, as it can be seen in Fig. 5. In the past we used a simplified model to account for the existence of this anomaly, in which we used Effective Medium Theory approximation in the zeroth order. In this model the anomaly was due to a Fabry-Pérot resonance of the +1 diffracted order in the grating. This approximate model predicted well the angular position of the anomaly (θ, ϕ) but did predict correctly whether the anomaly appears or not.

A rigorous analysis of the anomaly can be made with the help of Modal Analysis. It turns out that this anomaly is indeed a Fabry-Perot resonance but of the +1 modal order, and only when the electric field between the ridges is not evanescent. The values of θ and ϕ at which the resonance occurs can be calculated from the dispersion equation of Modal Analysis, the equation that determines the allowed modes in the grating,

$$\cos\left[\kappa_1 lw\right]\cos\left[\kappa_2\left(\Lambda - lw\right)\right] - \frac{1}{2}\left(\frac{\kappa_1}{\varepsilon\,\kappa_2} + \frac{\varepsilon\,\kappa_2}{\kappa_1}\right)\sin\left[\kappa_1 lw\right]\sin\left[\kappa_2\left(\Lambda - lw\right)\right] = \cos\left(k_{x,n}\,\Lambda\right), \quad (2)$$

where κ_1 and κ_2 are the wave vectors of the mode n in the ridge medium (photoresist) and the medium between the ridges (air) respectively,

$$\kappa_1 = \left(\varepsilon_g k_0^2 - k_t^2\right)^{1/2}, \quad (3.a)$$

$$\kappa_2 = \left(\varepsilon_a k_0^2 - k_t^2\right)^{1/2}, \quad (3.a)$$

with $k_{x,n}$ being the component along the grating vector $\boldsymbol{K}$ in an uniform medium of the mode n and having the expression

$$k_{x,n} = k_0 \sin\theta\cos\phi - nK, \quad (4)$$

k_0 the wave vector in vacuum, ε_g the dielectric permittivity of the photoresist, ε_a the dielectric permittivity of the air, the grating vector

$$K = \frac{2\pi}{\lambda}, \quad (5)$$

and the quantity ε is defined so that

$$\varepsilon = \begin{Bmatrix} 1 \text{ for TE mode} \\ \varepsilon_g \big/ \varepsilon_a \text{ for TM mode} \end{Bmatrix}. \quad (6)$$

The quantity k_t is the transversal component of the mode n and has the expression

$$k_t^2 = k_z^2 + k_0^2 \sin^2\theta \sin^2\phi. \quad (7)$$

In order to have a Fabry-Pérot resonance we set the condition

$$2k_z h_g = 2\pi. \quad (8)$$

Introducing Eq. (8) in Eq. (2) we obtain two equations in two variables, θ, ϕ corresponding to the TE and TM cases. Solving these equations we obtain the two curves in Fig. 4, the red and blue ones. There is a supplementary condition that must be fulfilled in order to have resonance, namely the electric transversal field between ridges must be non-evanescent. This condition, together with the distribution of the fields inside the grating illustrated in Fig. 5, explains why the Fabry-Perot resonance occurs only for mode +1.
In actual measurements we adjust the measurement configuration of the scatterometer so that this anomaly occurs and hence we have optimum sensitivity.

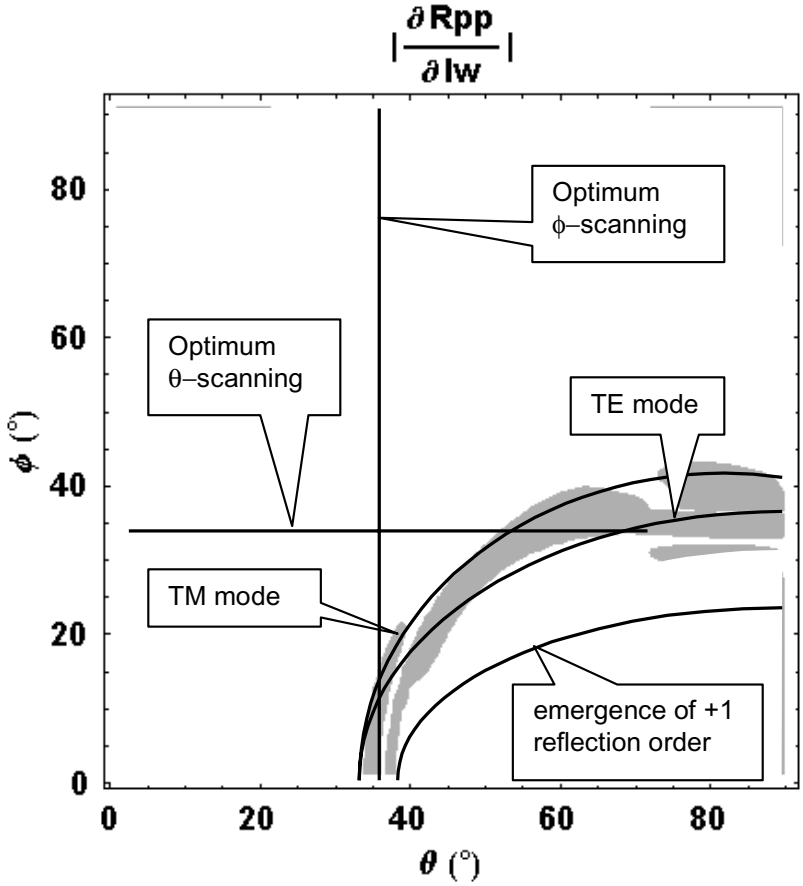


Fig. 5 The Fabry-Pérot anomaly contour calculated with the dispersion equation for sample XL6. The thick grey strip represents the peak areas of the partial derivative $\left|\partial R_{pp}/\partial lw\right|$. The anomaly due to the emergence of the +1 reflection order has much less impact on the sensitivity than the Fabry-Pérot anomaly. The importance of the anomaly is shown by the fact that the optimum θ-scanning and the optimum ϕ-scanning measurement configurations are those that intersects as much as possible the ridge of the anomaly.

5. CONCLUSION

In this paper are demonstrated the nanoscale capabilities of scatterometry, an optical metrology technique for the determination of the grating parameters. Lately we concentrated on improving and expanding the existing capabilities of scatterometry at University of New Mexico. Notably, for both ellipsometric-scatterometry and phase-modulation scatterometry we are now currently using shorter wavelength laser sources, namely we switched from He-Ne laser (633 nm) to He-Cd laser (442 nm and 325 nm). We are doing now what we call "blue and UV" scatterometry. Related to the con-

version to UV but also with an independent value of its own, we added to the scatterometer the ability to perform ϕ-scanning measurements. This capability adds new degrees of freedom to the sensitivity optimization procedure. Because the sensitivity analysis formalism leads us to the conclusion that sensitivity is provided mainly by anomalies, a rigorous analysis of a certain type of anomaly, the most likely to occur for our type of samples, was made using Modal Analysis. It turns out this is a Fabry-Pérot resonance of the +1 modal order when the fields between the ridges are not evanescent. Our analysis allows for the prediction of the anomaly position. This procedure allows us to find the optimum measurement configuration much faster than the SAF formalism used in the past.

REFERENCES

1. P. C. Logofatu, J. R. McNeil, "Sensitivity analysis of fitting for scatterometry," *Metrology, Inspection, and Process Control for Microlithography XIII*, *Proc. SPIE 3677*, pp. 177-183, 1999.

2. P. C. Logofatu, S. A. Coulombe, B. K. Minhas, J. R. McNeil, "Identity of the cross reflection coefficients for symmetric surface-relief gratings", *J. Opt. Soc. Am. A* **16**(5), pp. 1108-1114, 1999.

3. P. C. Logofatu, J. R. McNeil, "Measurement precision of optical scatterometry," *Metrology, inspection, and process control for microlithography XV*, *4344*, pp. 447-453, 2001.

4. P. C. Logofatu, J. R. McNeil, "Scatterometry: a metrology for subwavelength surface-relief gratings," *Metrology, inspection, and process control for microlithography XV*, *4344*, pp. 472-483, 2001.

5. P. C. Logofatu, "Phase-modulation scatterometry", *Appl. Opt.* **41**(34), pp. 7187-7192, 2002.

6. P. C. Logofatu, "Sensitivity analysis of grating parameter estimation", *Appl. Opt.* **41**(34), pp. 7179-7186, 2002.

7. J. R. McNeil, S. S. H. Naqvi, S. M. Gaspar, K. C. Hickman, S. R. Wilson, in *Encyclopedia of Materials Characterization* C. A. E. C. Richard Brundle, Shaun Wilson, Ed. (Manning Publishing Co., 1992).

8. J. R. McNeil, et al., "Scatterometry applied to microelectronic processing", *Microlithography World* **1**(15), pp. 16-22, 1992.

9. J. R. McNeil, S. A. Coulombe, P. C. Logofatu, C. J. Raymond, S. S. H. Naqvi, "Application of optical scatterometry to microelectronics," *Conference on Lasers and Electro-Optics, Cleo '98*, pp. 348-349, 1998.

10. B. K. Minhas, S. A. Coulombe, S. S. H. Naqvi, J. R. McNeil, "Ellipsometric-Scatterometry for the metrology of sub-0.1 μm linewidth structures", *Appl. Opt.* **37**(22), pp. 5112-5115, 1998.

11. S. S. H. Naqvi, J. R. McNeil, R. H. Krukar, K. P. Bishop, "Scatterometry and the simulation of diffraction-based metrology", *Microlithography World* **2**(3), pp. 5-16, 1993.

12. S. S. H. Naqvi, J. R. McNeil, R. H. Krukar, Z. R. Hatab, "Grating parameter estimation using scatterometry," *Miniature and Micro-Optics and Micromechanics*, *Proc. SPIE 1992*, pp. 170, 1993.

13. C. J. Raymond, M. R. Murnane, S. S. H. Naqvi, J. R. McNeil, "Metrology of subwavelength photoresist gratings using optical scatterometry", *J. Vac. Sci. Technol.* **13**(4), pp. 1484-1495, 1995.

14. C. J. Raymond, M. R. Murnane, S. L. Prins, S. S. H. Naqvi, J. R. McNeil, "Multi-parameter process metrology using scatterometry," *Optical Characterization Techniques for High-Performance Microelectronic Device Manufacturing II*, *Proc. SPIE 2638*, pp. 84-93, 1995.

Nanoheteroepitaxial growth of GaN on Si by organometallic vapor phase epitaxy

D. Zubia,[a] S. H. Zaidi, S. R. J. Brueck,[b] and S. D. Hersee
Center for High Technology Materials and Electrical and Computer Engineering Department, University of New Mexico, Albuquerque, New Mexico 87106

(Received 27 July 1999; accepted for publication 16 December 1999)

Nanoheteroepitaxy has recently been proposed as a technique for significantly extending the thickness of pseudomorphic growth in mismatched heterostructures. This letter reports the experimental application of nanoheteroepitaxy for the growth of GaN on patterned ⟨111⟩ oriented silicon-on-insulator substrates by organometallic vapor phase epitaxy. Transmission electron microscopy reveals that the defect concentration decays rapidly away from the heterointerface as predicted by nanoheteroepitaxy theory. The melting point of the nanoscale islands is found to be significantly reduced, enhancing substrate compliance and further reducing the strain energy in the GaN epitaxial layer. © *2000 American Institute of Physics.* [S0003-6951(00)02607-3]

Nanoheteroepitaxy[1] (NHE) exploits three-dimensional stress relief mechanisms that are available to nanoscale islands to reduce the strain energy in lattice-mismatched material systems. The NHE approach consists of patterning a substrate into a two-dimensional array of 10–300 nm sized nucleation islands and then selectively growing epitaxial material vertically on the islands (Fig. 1, point *a*). This is followed by lateral overgrowth to achieve coalescence (Fig. 1, point *b*). In contrast to current epitaxial lateral overgrowth,[2] and pendeo[3] approaches, where the patterning is on the scale of 1–10 μm, the nanoscale patterning in NHE allows enhanced substrate compliance mechanisms to operate. For example, in addition to strain partitioning similar to that found in planar compliant structures,[4] stress in the NHE sample decays exponentially away from the heterointerface with a characteristic decay length proportional to the diameter of the island. Strain partitioning and stress decay interact synergistically to significantly lower the strain energy in nanostructural lattice mismatched material systems. NHE theory predicts that mismatch dislocation formation can be eliminated from materials systems with a lattice mismatch in the range of 0%–4%. In material systems with larger lattice mismatch such as GaN on Si (20% lattice mismatch) defects are probably unavoidable, however as we show below, NHE results in a significant reduction in the local defect density in nanostructural GaN/Si islands.

Nanoheteroepitaxy was investigated using organometallic vapor phase epitaxy (OMVPE) growth of GaN on the patterned surface silicon layer of commercial ⟨111⟩ silicon-on-insulator (SOI) wafers manufactured by the separation by implantation of oxygen (SIMOX) process. The SOI wafers were patterned using interferometric lithography and reactive ion etching[5,6] to form a square two-dimensional array of silicon islands on top of a field of SiO_2 as illustrated in Fig. 1. The height of the islands was 40 nm while the diameter varied between 80 and 300 nm. Island periodicity was 360 and 900 nm. GaN was then selectively grown on the silicon islands using a three step sequence consisting of an *in situ* cleaning at a temperature of 878 °C for 5 min under a hydrogen atmosphere, a low temperature (543 °C) nucleating layer deposition, and finally main epilayer growth at high temperature (1031 °C). This yielded a GaN island thickness of 100–200 nm. The exposed SiO_2 layer in the SOI served as a convenient mask for the selective growth of GaN on silicon. The SOI structure allowed efficient implementation of the patterning and selective growth requirements of the NHE approach.

In addition to exploiting three-dimensional stress relief mechanisms due to geometrical effects, nanostructuring can also enable other stress relief effects due to the large surface-to-volume ratio of nanoscale structures. For example, in material systems in which the epitaxial material has a larger melting point than the substrate, compliance can be actively enhanced due to softening of the nanoscale substrate islands at growth temperatures well below the substrate bulk melting point. Evidence of substrate softening is shown in Fig. 2 which compares the size and shape of silicon nanometer sized islands (a) without and (b) with heat treatment at 1110 °C for 10 min. Figure 2 shows that mass transport has occurred due to melting in the heat-treated sample. This indicates that the melting point for these silicon islands has been reduced from its bulk value of 1412 °C to less than 1110 °C. The *in situ* cleaning temperature previously used

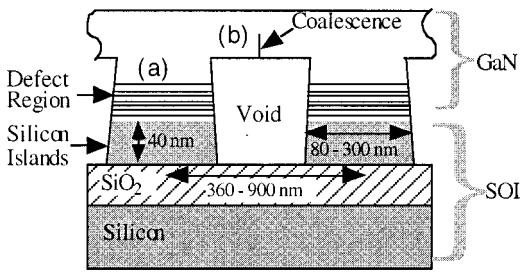


FIG. 1. Shows the patterned SOI substrate with islands created in the upper Si layer. The selective growth of GaN is also illustrated. A defected region is located at the heterointerface.

[a]Electronic mail: davzubia@chtm.unm.edu
[b]Also with the Department of Physics and Astronomy, University of New Mexico.

 © 2000 American Institute of Physics

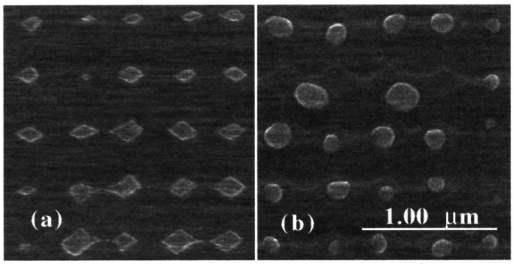


FIG. 2. Comparison of silicon morphologies on patterned SOI substrate before (a) and after (b) heating to 1110 °C. The change of shape indicates that the silicon islands have melted.

for GaN growth on sapphire[7] is 1100 °C, however, because of this observed melting we have now reduced the *in situ* clean temperature to 878 °C. A reduction of melting point is anticipated from recent results on nanophase materials which has shown that the materials properties (including the melting point) of nanometer scale particles are significantly modified.[8] In the case of NHE this reduced melting point can be utilized during growth to soften the silicon islands (i.e., reduce their elastic modulus) and enhance substrate compliance. We call this effect active compliance.

While not explicitly discussed in the original NHE model,[1] active compliance can be incorporated in the strain partitioning equations by adjusting the elastic compliance ratio, K, given by

$$K = Y_{\text{epi}}(1-\nu_{\text{sub}})/[Y_{\text{sub}}(1-\nu_{\text{epi}})], \tag{1}$$

where Y is Young's modulus and ν the Poisson's ratio in the epitaxial material and the substrate materials. At low temperatures, well below the actual melting point of the epitaxial and substrate materials, the Young's moduli of these materials are equal to their respective bulk values. For a higher growth temperature, near the reduced melting point of the substrate island, Y_{sub} will be reduced causing K to be increased above its bulk value, K_o. The impact of active compliance on strain partitioning is illustrated in Fig. 3, which shows the effect that increasing K has on the normalized strain, ϵ/ϵ_T, in the epitaxial and substrate layers, where ϵ_T is the total mismatch strain. It is not known at this time by how much the value of K can be increased beyond K_o, however, to illustrate the active compliance effect, an increased value of $5K_o$ was arbitrarily chosen for comparison in Fig. 3. The graph shows that active compliance enhances strain partitioning by increasing the partitioning at a given thickness and increasing the ultimate degree of strain partitioning. In addition, active compliance works to directly lower the total strain energy in the system when Y_{sub} is reduced because strain energy is proportional to the Young's modulus.[1]

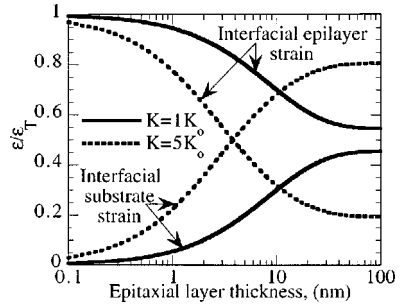


FIG. 3. Plots the normalized interfacial strain in the epilayer and substrate vs epilayer thickness. Without active compliance (solid lines) the final amount of strain in the epilayer and substrate are similar. With active compliance (dashed lines) much more of the total strain is transferred to the substrate.

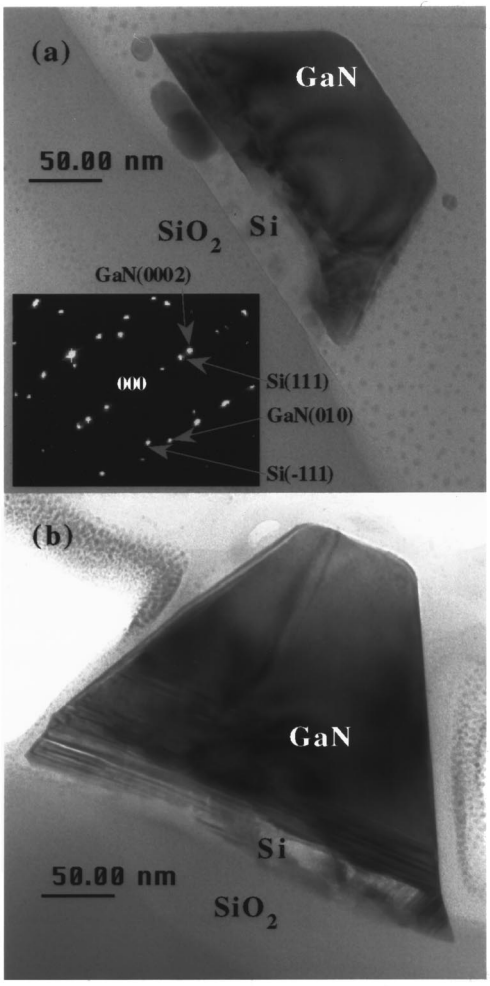


FIG. 4. Cross-sectional TEM micrographs of GaN grown on Si islands on the patterned SOI substrate. Both figures show defects at the heterointerface but high quality GaN away from the interface. The electron diffraction pattern reveals the epitaxial relationship between GaN and Si.

Island size is a critical parameter in NHE as it controls both the elastic modulus through the actual melting point of the substrate island and the characteristic decay length of the strain energy. During selective epitaxy, the island size and spacing will also control the effective growth rate at the island.

Cross-section transmission electron microscopy (XTEM) micrographs of GaN growth on two silicon islands, one 80 nm in diameter and the other 280 nm in diameter, are shown in Fig. 4. In both cases the defects are concentrated near the GaN/Si heterointerface and the defect density decays rapidly to vanish at approximately 20–50 nm from the interface. In Fig. 4(a), strain field contours can be observed in the defect-free region. The general shape of the strain contours indicates that the strain originates at the GaN/Si heterointerface. The contour lines broaden and separate, away from the heterointerface, indicating a decaying strain field in agreement with the predictions of NHE theory. Lateral growth of GaN, extending beyond the edges of the silicon island, is also ob-

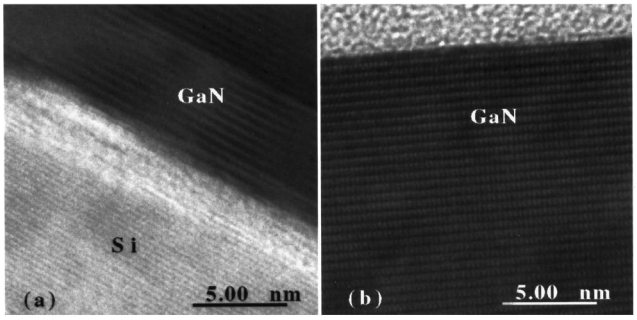


FIG. 5. High-resolution TEM micrographs of (a) GaN/Si heterointerface and (b) GaN away from the interface.

served. The electron diffraction pattern from the sample in Fig. 4(a) (inset) reveals the relative crystallographic orientation of GaN and Si as GaN[0001]‖Si[111]. Thus the silicon island is acting as an epitaxial template. Figure 4(b) shows a threading dislocation in the GaN, indicating that the mechanism that creates threading dislocations was not suppressed sufficiently to prevent the formation of the threading dislocation in this particular island. Figures 4(a) and 4(b) also show that GaN grows preferentially on the (111) surface of the silicon island, suggesting that selective growth via crystallographic orientation (i.e., use surface patterning instead of using an oxide mask) may be achievable for this material system.

High-resolution XTEM (Fig. 5) reveals stacking faults in the GaN near the interface [Fig. 5(a)] and a highly defected silicon layer at the interface. Segmented silicon lattice planes lying within the 2.0-nm-thick defected layer, in some areas penetrating as much as three lattice planes (0.95 nm) into the layer, indicate that the layer is composed of silicon, implying that much of the mismatch strain is being taken up by the silicon. Figure 5(b) shows that GaN is locally of high quality away from the interface.

In this experimental demonstration of nanoheteroepitaxy, results are presented on the selective-area OMVPE growth of GaN on a novel, patterned SOI substrate. A reduced melting point is observed for nanoscale silicon islands enabling active compliance effects that work to further enhance strain partitioning and reduce strain energy. Electron microscopy reveals that the defect concentration decays rapidly away from the GaN/Si heterointerface, in agreement with NHE theory. Beyond this defected region, the GaN appears to be of high quality. Current work is focusing on evaluating the quality of the material grown by the NHE approach including the coalescence region and other effects such as differential thermal expansion and will be discussed elsewhere.

Partial support for this work was provided by the Air Force Office of Scientific Research. XTEM images were acquired with a JEOL 2010 high-resolution analytical electron microscope at the University of New Mexico.

[1] D. Zubia and S. D. Hersee, J. Appl. Phys. **85**, 6492 (1999).
[2] S. Nakamura, J. Cryst. Growth **195**, 242 (1998).
[3] T. S. Zheleva, S. A. Smith, D. B. Thomson, T. Gehrke, K. J. Linthicum, P. Rajagopal, E. Carlson, W. M. Ashmawi, and R. F. Davis, MRS Internet J. Nitride Semicond. Res. **4S1**, G3.38 (1999).
[4] J. P. Hirth and A. G. Evans, J. Appl. Phys. **60**, 2372 (1986); Y. H. Lo, Appl. Phys. Lett. **59**, 2311 (1991).
[5] S. H. Zaidi, A. S. Chu, and S. R. J. Brueck, Mater. Res. Soc. Symp. Proc. **358**, 957 (1995).
[6] S. H. Zaidi, A. S. Chu, and S. R. J. Brueck, J. Appl. Phys. **80**, 6997 (1996).
[7] S. D. Hersee, J. Ramer, K. Zheng, C. F. Kranenberg, K. Malloy, M. Banas, and M. Goorsky, J. Electron. Mater. **24**, 1519 (1995).
[8] A. Goldstein, *Handbook of Nanophase Materials* (Marcel Dekker, New York, 1997), pp. 3 and 4.

Nanoheteroepitaxy: Nanofabrication route to improved epitaxial growth

D. Zubia,[a),b)] S. H. Zaidi,[c)] S. D. Hersee,[b)] and S. R. J. Brueck[b),d)]
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

(Received 1 June 2000; accepted 5 September 2000)

Nanoheteroepitaxy is a fundamentally new epitaxial approach that utilizes three-dimensional stress relief mechanisms available to nanoscale heterostructures to eliminate defects provided the island diameter is below a critical value $2l_c$. Analysis shows that $2l_c \sim (15-30)\times$ the critical thickness h_c. In the case of GaAs on Si ($\sim$4% misfit), $2l_c \sim 40$ nm. In material systems such as GaN on Si ($\sim$20% misfit), where the misfit is much larger and interfacial defects are unavoidable, the nanoheteroepitaxial structure is shown to reduce the formation and propagation of threading defects. Nanostructured substrate parameters that impact growth are discussed and interferometric lithography is introduced as a method for fabrication of large-area substrates for nanoheteroepitaxy. Si nanoisland diameters as small as 20 nm are demonstrated. Scanning and transmission electron microscopy data of GaN grown on Si (via organometallic vapor phase epitaxy) shows reduced threading defects in nanostructured samples compared to growth on planar substrates. Photoluminescence intensity data of nanostructured samples is enhanced by $\sim 100\times$ as compared to planar-growth samples. © *2000 American Vacuum Society.* [S0734-211X(00)11206-5]

I. INTRODUCTION

The heterogeneous integration of compound semiconductors and devices onto Si substrates is a long sought goal that, if achieved, would have a major impact on the functionality of integrated circuits by combining improved speed, noise, and optical performance with the massive density of Si very large scale integration. However one of the major obstacles to achieving this goal has been the lack of a metallurgically compatible (lattice parameter, chemical interactions, and thermal expansion) compound semiconductor with suitable optical and electronic properties and a mature processing technology. Primarily due to its processing maturity and superior optical and electronic properties, much research has focused on developing a strategy for integrating GaAs with Si although this system has a lattice mismatch of $\sim$4%, thermal expansion mismatch of $\sim$60%, and a polarity mismatch.[1] By far the dominant issue in the monolithic integration of GaAs on Si substrates has been the formation and propagation of large densities of line and areal defects, principally dislocations, stacking faults, and cracks, arising from the lattice mismatch.[2] While material quality and device performance has improved over the years, a persistent problem has been poor long-term reliability due to instability of nonradiative defects that propagate during room temperature cw operation, especially in devices such as lasers where there is significant heat generation during device operation.[3,4]

Recently developed heterogeneous integration strategies that address mismatch issues have resulted in a renewed research emphasis on integrating compound semiconductors with Si. For example, compliant[5,6] substrates mitigate strain energy buildup by partitioning the strain between the epilayer and the substrate. Epitaxial lateral overgrowth[7,8] reduces defect density by changing growth direction and eliminating the atomic connection between the two materials. Nanoheteroepitaxy (NHE) is a fundamentally new approach that takes advantage of three-dimensional stress relief mechanisms available to nanometer sized heterostructures to reduce strain energy.[9,10] In NHE, growth is selectively initiated on a dense array of nanosized nuclei [Fig. 1(b)]. After a thin layer is grown, typically less than 100 nm, the epilayer becomes strain free [see Fig. 1(b)]. At this point, growth conditions are modified to encourage lateral growth and coalescence of islands into a continuous epitaxial film. Theory shows that for lattice mismatches up to 4.5% (this range includes many of the important III–V on Si combinations including GaAs), the strain energy can be below that required to form defects for readily achievable nucleus dimensions. Even in more highly mismatched systems, such as GaN on Si (20% misfit), the local presence of free surfaces in the novel NHE structure impacts the character of the defects and leads to improved material quality.

II. THEORY OF NHE

The additional stress relief mechanisms available to nanoscale islands as compared to a conventional planar epilayer are illustrated in Fig. 1. In a conventional planar structure [Fig. 1(a)], the epilayer can elastically deform vertically (i) to relieve part of the mismatch stress. However, because only part of the stress is relieved, the strain energy in the epilayer increases linearly with increasing epilayer thickness up to the point where it becomes energetically favorable to create interfacial misfit dislocations (v). In contrast, in nanoscale islands [Fig. 1(b)] the substrate and epilayer are

[a)]Electronic mail: davzubia@chtm.unm.edu
[b)]Also with: Department of Electrical and Computer Engineering, University of New Mexico.
[c)]Present address: Gratings, Inc., 2655 Pan American Freeway, Suite A, Albuquerque, NM 87107.
[d)]Also with Department of Physics and Astronomy, University of New Mexico.

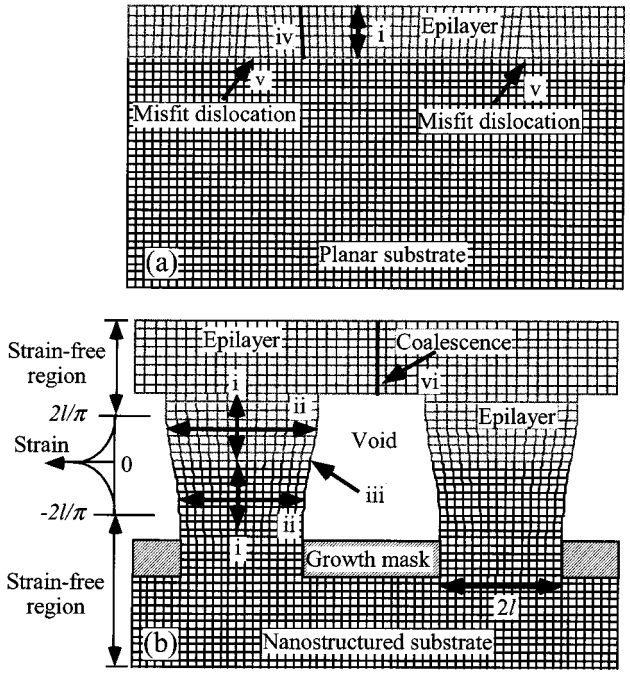


FIG. 1. Schematic comparison of conventional (large-area) heteroepitaxy (a) and nanoheteroepitaxy (b). Nanoheteroepitaxy allows 3D strain relief mechanisms that are unavailable in large-area growth.

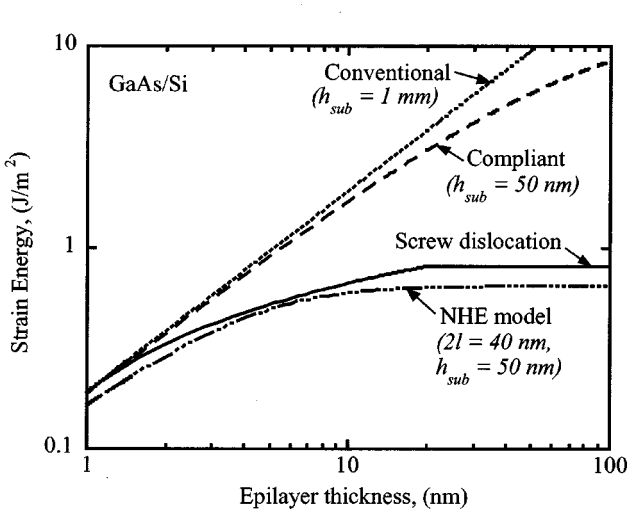


FIG. 2. Strain energy as a function of epilayer thickness in the GaAs/Si material system for conventional growth, compliant substrate growth, and NHE growth. Only the NHE growth (40 nm diam islands) is below the screw dislocation nucleation threshold (also shown) for all film thicknesses.

able to deform vertically (i) and laterally (ii). Moreover, the strain decays exponentially away from the interface in both materials. This helps to reduce both the strain in the epilayer and the strain energy in each nanoisland. In many cases, strain-free material is achieved after only ~100 nm of growth. If mismatch dislocations are formed in the nanoisland, they can glide or climb to an island edge (iii), assuming the glide plane is suitably oriented. In conventional planar heteroepitaxy, dislocations are unlikely to reach the sample edge and instead nucleate threading defects (iv).

A. Calculation of the strain energy

The areal strain energy associated with a film of thickness h_{epi} in conventional planar epitaxy is given by[11]

$$E_{\text{CON}}=\frac{Y}{1-\nu}\varepsilon_T^2 h_{\text{epi}}, \tag{1}$$

where Y is Young's modulus, ν is Poisson's ratio, and $\varepsilon_T = 2|a_{\text{epi}}-a_{\text{sub}}|/(a_{\text{epi}}+a_{\text{sub}})$ is the lattice mismatch. a_{epi} and a_{sub} are the equilibrium lattice parameters of the epilayer and substrate. Due to strain partitioning (and neglecting any bending stresses) the areal strain energy in planar compliant substrates is the sum of the strain energies in the epilayer and substrate, given by[5]

$$E_{\text{COM}}=\frac{Y_{\text{epi}}}{1-\nu_{\text{epi}}}\varepsilon_{\text{ipi}}^2 h_{\text{epi}}+\frac{Y_{\text{sub}}}{1-\nu_{\text{sub}}}\varepsilon_{\text{sub}}^2 h_{\text{sub}}, \tag{2}$$

where ε_{epi} and ε_{sub} are the partitioned strain in the epilayer and substrate and h_{sub} is the thickness of the substrate. ε_{epi} and ε_{sub} are functions of the elastic moduli and thickness of the epilayer and substrate.[12]

The areal strain energy in NHE structures is the sum of the strain energies in the epilayer and substrate due to strain partitioning, as in the compliant substrate case. However, the strain energy calculation must also take into account strain decay in the NHE structure [see Fig. 1(b)]. The areal strain energy in each nanoisland was previously[9] shown to be

$$E_{\text{NHE}}=\frac{Y_{\text{epi}}}{1-\nu_{\text{epi}}}\varepsilon_{\text{epi}_0}^2 h_{\text{epi}}^{\text{eff}}+\frac{Y_{\text{sub}}}{1-\nu_{\text{sub}}}\varepsilon_{\text{sub}_0}^2 h_{\text{sub}}^{\text{eff}}, \tag{3}$$

where $\varepsilon_{\text{eip}_0}$ and $\varepsilon_{\text{sub}_0}$ are the strain in the epilayer and substrate materials at the heterointerface, respectively, and are called the interfacial partitioned strains. The parameters $h_{\text{epi}}^{\text{eff}}$ and $h_{\text{sub}}^{\text{eff}}$ in Eq. (3) account for the strain decay and describe the effective thickness of material that is contributing to the integrated strain energy in the epilayer and substrate, respectively.

A plot of strain energy as a function of epilayer thickness for different substrate configurations is given in Fig. 2 for the case of GaAs growth on silicon. Also shown in Fig. 2 is the energy associated with the formation of a screw dislocation. The curve labeled ''Conventional'' is for a 1 mm thick planar substrate. Thin planar substrates that enable compliance and strain partitioning effects are represented by the curve labeled ''Compliant.'' These substrates were assumed to have a compliant layer thickness of 50 nm. The nanoheteroepitaxial configuration is represented by the curve labeled ''NHE model.'' The energy associated with a screw dislocation increases as the logarithm of the epilayer thickness up to the point where the epilayer thickness is equal to the radius of the island ($l=20$ nm). Beyond this point, the screw dislocation energy is constant and controlled by l because the edge of the sample is the nearest free surface and the distance to this free surface controls the screw dislocation energy. The strain energy in the thick planar substrate ($h_{\text{sub}}=1$ mm) increases linearly with increasing epilayer thickness. Notice that even at an epilayer thickness of h_{epi}

$=1$ nm, the strain energy in the conventional substrate is equal to or greater than the screw dislocation energy. This means that it is energetically favorable to create dislocations in such epilayers. Although thin compliant substrates ($h_{\text{sub}}=50$ nm) are designed to reduce the strain energy, the energy reduction in these substrates is insufficient to eliminate dislocations. Moreover, at epilayer thicknesses below 2 nm, the thin substrate is incompliant and behaves much like a conventional substrate.

In contrast, the strain energy in the NHE configuration ($2l=40$ nm, $h_{\text{sub}}=50$ nm) is below the screw dislocation energy at all epilayer thicknesses, making it energetically unfavorable for screw dislocations to form. For thick NHE films, the strain energy reaches an asymptote, below the screw dislocation energy, as the film relaxes to its normal lattice constant. Throughout the thickness range, the strain energy of the NHE sample is less than the screw dislocation energy, indicating that GaAs of arbitrary thickness can be coherently grown on silicon, provided the surface is nanostructured with islands having a diameter of $\lesssim 40$ nm.

Equation (3) can be greatly simplified for the case when h_{epi} and h_{sub} are $\gg l$ and $K=1$. This simplified equation can then be compared to criteria for coherent growth on planar substrates such as the criteria put forth by People and Bean[11] for the critical thickness h_c, at which it becomes energetically favorable for dislocations to be formed in strained epitaxial layers. [$K=1$ implies that $Y_{\text{epi}}/(1-\nu_{\text{epi}})=Y_{\text{sub}}/(1-\nu_{\text{sub}})=Y/(1-\nu)$]. In this limit, the interfacial partitioned strains reduce to

$$\varepsilon_{\text{epi}_0}=\varepsilon_{\text{sub}_0}=\frac{\varepsilon_T}{2}. \tag{4}$$

The prefactors $[1-\text{sech}(kl)]$ in $h_{\text{epi}}^{\text{eff}}$ and $h_{\text{sub}}^{\text{eff}}$ normally vary from 0.78 to 0.91 corresponding to Poisson's ratio varying from 0.1 to 0.5 and can be taken as ~ 1 as a worst case approximation.[9] In addition, the factors $[1-\exp(-\pi h/l)]$ reduce to unity for h_{epi} and h_{sub} large. This simplifies the effective thicknesses to

$$h_{\text{epi}}^{\text{eff}}=h_{\text{sub}}^{\text{eff}}=\frac{l}{\pi}. \tag{5}$$

Equation (3) can now be written in the simplified form

$$E_{\text{NHE}}=\left(\frac{Y}{1-\nu}\right)\varepsilon_T^2\frac{l}{2\pi}. \tag{6}$$

Equation (6) shows that the strain energy in the NHE sample is finite and proportional to the island radius l. Equation (6) describes the asymptote that the NHE approaches on the right edge of Fig. 2.

The energy balance approach taken by People and Bean[11] for determining the onset of misfit dislocation states ''that interfacial misfit dislocations will be generated when the areal strain energy exceeds the self-energy of an isolated dislocation.'' Equating Eq. (6) and area energy density associated with an isolated screw dislocation[9] yields

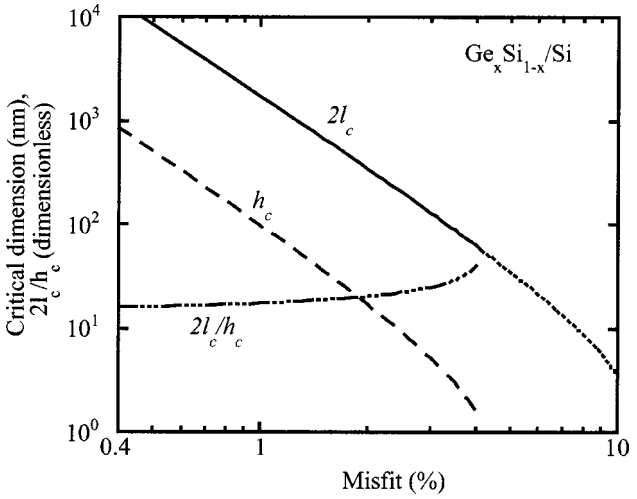


FIG. 3. Theoretical critical layer thickness h_c and critical island diameter $2l_c$ of Ge_xSi_{1-x} alloys grown on silicon vs misfit. Also shown is the ratio $2l_c/h_c$.

$$\frac{l_c}{2\pi}=\left(\frac{1-\nu}{1+\nu}\right)\left(\frac{1}{20\pi\sqrt{2}}\right)\left(\frac{b^2}{a}\right)\left(\frac{1}{\varepsilon_T^2}\right)\ln\left(\frac{l_c}{b}\right), \tag{7}$$

where b is the magnitude of the Burger's vector, a is the lattice parameter, and l has been replaced by l_c to represent the critical island radius beyond which it is energetically favorable to generate dislocations. Figure 3 compares $2l_c$ to h_c as a function of misfit using the Ge_xSi_{1-x} on Si material system as an example. Also shown is the ratio of $2l_c/h_c$ which is obtained by dividing Eq. (7) by h_c:

$$\frac{2l_c}{h_c}=4\pi\frac{\ln(l_c/b)}{\ln(h_c/b)}. \tag{8}$$

Both $2l_c$ and h_c decrease as the lattice mismatch increases. However, $2l_c$ is between 15 and 30 times larger than h_c depending on the misfit. Figure 3 shows how the lateral dimensions accessible by lithography can be used to control the strain energy and, thereby, the quality of growth in the vertical direction. This gives a wholly new degree of freedom in heterostructure growth.

B. Defect management

The dashed line in Fig. 3 is an extrapolation of the $2l_c$ curve and it indicates that for material systems with lattice mismatch greater than ~8%, the critical diameter drops below ~10 nm and will be increasingly difficult to achieve for large-area substrates within current lithographic and process capabilities. However, one crucial aspect of NHE theory is that the strain decays exponentially away from the interface into both the epilayer and substrate with characteristic length $2l/\pi$ [see Fig. 1(b)]. This means that all of the strain energy is contained in a region near the interface. Material grown beyond this region is essentially stress-free inhibiting the formation of defects in epitaxial layers not only during growth, but also when device fabrication and operation stress the material. This is potentially important for high performance

TABLE I. Aspects of nanostructure patterning controllable by interferometric lithography.

Nanoisland parameter	NHE impact	IL approach	Accessible range
Pattern periodicity	Film coalescence	Incident angles $\{p=\lambda/[2\sin(\theta)]\}$, nonlinear IL and integration of optical and interferometric lithographies	>75 nm
Size	Island growth, strain partitioning, active compliance	Vary exposure flux, use photoresist and other process nonlinearities.	to 20 nm
Pattern symmetry	Film coalescence	Number and orientation of beams	Rectangular, hexagonal,...
Localized growth areas	Thermal expansion strain, integration with Si circuitry	Integration of conventional optical and IL, hard masks.	Fully arbitrary to optics resolution limits

and long term reliability of devices under high power or high current density applications. Moreover, in cases where interfacial defects cannot be avoided (e.g., GaN on Si), the geometrical structure of NHE provides free surfaces close to the interface suitable for interfacial dislocation termination. (This is in contrast to a planar structure which provides a free surface essentially only at the top of the epilayer.) This helps to inhibit the formation of threading defects since dislocations must terminate upon themselves in a loop, on another dislocation, or at a free surface.[13] These defect management effects are clearly observed in the growth of GaN on nanostructured ⟨111⟩ Silicon on insulator (SOI) as presented in Sec. IV.

III. NANOSTRUCTURED SUBSTRATE PARAMETERS FOR NHE

Clearly, island diameter is an important parameter in NHE. However, other parameters such as island spacing or pitch are also expected to play an important role. Table I lists four nanostructured substrate parameters along with how they can be controlled using interferometric lithography (IL),[14–17] an inexpensive, large-area, highly uniform patterning technology.

Island spacing or pitch is also a critical parameter for NHE since it impacts the growth rate at the island and determines the lateral dimension needed for coalescing the islands into a continuous layer. We have observed for example that reducing the pitch from 900 to 360 nm shortened the growth time and improved coalescence of GaN grown on nanostructured SOI substrates. The limiting pitch for conventional IL is $\sim\lambda/2$. To date, most of our experiments have been carried out at wavelengths of 364 nm (Ar ion laser) or 355 nm [third harmonic of yttrium–aluminum–garnet (YAG)] with a limiting period of ~180–200 nm. This wavelength range was chosen to be compatible with well-developed, commercially available *i*-line photoresists. With the advent of availability of 193 nm resists developed for ArF based photolithography, we are extending our experiments to ~125 nm pitches with a fifth harmonic YAG source at 213 nm. Nonlinear processes[18] to further extend lithographic capabilities to higher density (smaller pitch) patterns have also been demonstrated. The basic idea follows from the realization that the fundamental limit of IL is on the pitch, not on the linewidth or critical dimension (CD). First a periodic array of small features with $CD<p/4$ with p the pitch is transferred into a sacrificial layer (e.g., a Si_3N_4 film). In a second lithographic step, an identical pattern is defined with a phase shift of $p/2$ and transferred into the sacrificial layer. The result is a pattern with one-half the pitch, doubling the density.

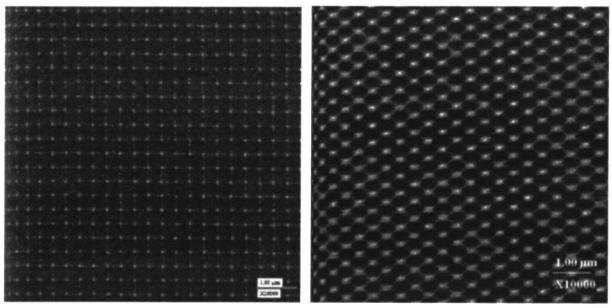


FIG. 4. Square and hexagonal arrays of ~20 nm diam SOI nanoisland samples (360 nm pitch) fabricated by interferometric lithography.

Pattern symmetry may also play an important role in coalescence. For example, a hexagonal symmetry is expected to yield superior coalesce of wurtzite GaN crystals compared to a rectangular symmetry. Pattern symmetry can be controlled by varying the exposure/development and pattern transfer parameters. Complex two-dimensional (2D) patterns can be generated in a single exposure with multiple input beams. Specifically, three-beam and three exposure geometries lead to hexagonal symmetries and four and five beam geometries give 2D hole arrays.[19] Combined with multiple resist tonalities, image reversal, multilayer resists, and the possibilities created with additional sacrificial hard-mask layers, there is a very large flexibility to provide a wide variety of nanostructure geometries. Figure 4 shows nanostructured SOI substrates with rectangular and hexagonal symmetries with ~20 nm diam islands.

The ability to define localized growth areas will address issues related to thermal expansion mismatch as well as integration with silicon circuitry. Isolated areas of nanostructures can be generated by combining IL with conventional optical lithography to define larger areas.

A major advantage of IL is its scalability and low cost manufacturability. In contrast to e-beam lithography approaches, which are able to reach these same scales, IL offers a critical parallel-printing, large-area capability. Applying the large-field, parallel-printing advantages of optical lithography, IL easily provides macroscopic sample areas ($\gtrsim$1–100 cm^2) with a high density of nanostructures. This will be critical both for optimization of NHE across a broad parameter space and for eventual device fabrication.

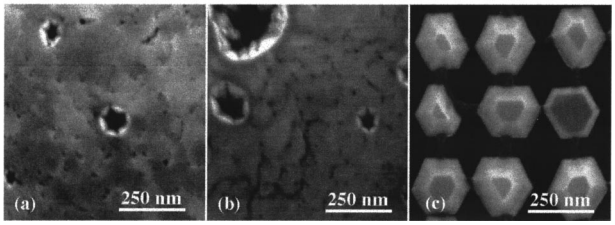


FIG. 5. Plan view SEM images of ~140 nm thick GaN epitaxial samples grown on: (a) planar ⟨111⟩Si; (b) planar ⟨111⟩ SOI; and (c) on a ~70 nm diam island nonostructured ⟨111⟩ SOI sample.

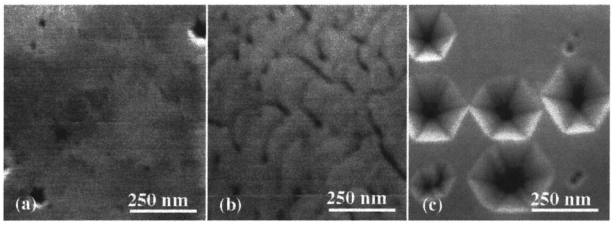


FIG. 6. Plan view SEM image of ~400 nm thick GaN samples grown on: (a) planar ⟨111⟩Si; (b) planar ⟨111⟩ SOI; and (c) nanostructured ⟨111⟩ SOI samples.

IV. RESULTS

An initial investigation of NHE was carried out using organometallic vapor phase epitaxy growth of GaN on nanostructured ⟨111⟩ SOI wafers manufactured by the separation by implantation of oxygen process. For comparative purposes, GaN was also grown on planar ⟨111⟩ silicon and planar ⟨111⟩ SOI. The general procedure consisted of first nanostructuring the surface silicon layer of the SOI substrate. This was followed by cleaning the substrates to remove impurities and native oxide. All three types of substrates were then loaded into the growth chamber. GaN was then grown on the samples using a three-step sequence consisting of: (1) an *in situ* clean at a temperature of 878 °C for 5 min under a hydrogen/nitrogen atmosphere; (2) a low temperature (543 °C) nucleating layer deposition;[20] and (3) the main epilayer growth at high temperature (1031 °C). The samples were characterized using scanning-electron microscopy (SEM), transmission-electron microscopy (TEM), and photoluminescence (PL).

Figure 5 shows plan view SEM micrographs of GaN grown to a nominal thickness of 140 nm on the three types of substrates. The SEM image of GaN grown on the planar ⟨111⟩ Si [Fig. 5(a)] clearly shows etch pits and dark spots indicative of crystal defects. The GaN grown on the planar ⟨111⟩ SOI [Fig. 5(b)] also has etch pits and a connected network of dark lines indicative of grain boundaries. In contrast, Fig. 5(c) shows an array of truncated six-sided pyramid shaped crystals typical of wurtzite GaN. No grain boundaries or etch pits are observed indicating that the array is composed of high quality single crystals. The diameter of the crystals is approximately 300 nm.

GaN was grown to a nominal thickness of 400 nm in order to coalesce the array of GaN crystals into a continuous layer. Figure 6(a) is a plan view SEM of GaN grown on planar ⟨111⟩ Si while Fig. 6(b) shows GaN grown on planar ⟨111⟩ SOI. These SEMs look similar to those of the corresponding 140 nm samples. In contrast, Fig. 6(c) shows evidence that significant (although still incomplete) coalescence has occurred after 400 nm of growth for the NHE sample. Figure 7 shows cross-sectional diffraction-contrast images of the three types of samples after 400 nm of growth. Figure 7(a) clearly shows threading dislocations in the GaN layer while no dislocations are observable in the silicon substrate. The GaN grown on the planar ⟨111⟩ SOI is also highly defected, however the defect character is different as shown in Fig. 7(b). The defects in the SOI sample appear to be misoriented grain boundaries with possible cubic GaN nanophases at the interface. Figure 7(c) is a cross-sectional image of the coalescence region. Small looped defects can be observed near the coalescence region. While grain boundaries were found in other coalescence regions, this result is

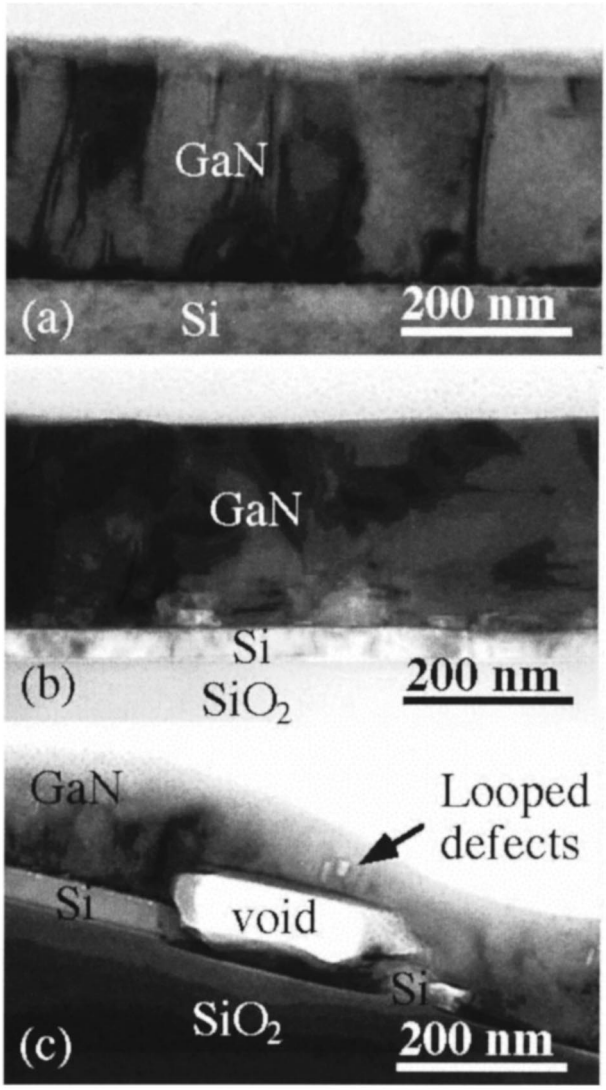


FIG. 7. Cross-section diffraction-contrast images of the same samples as shown in Fig. 6.

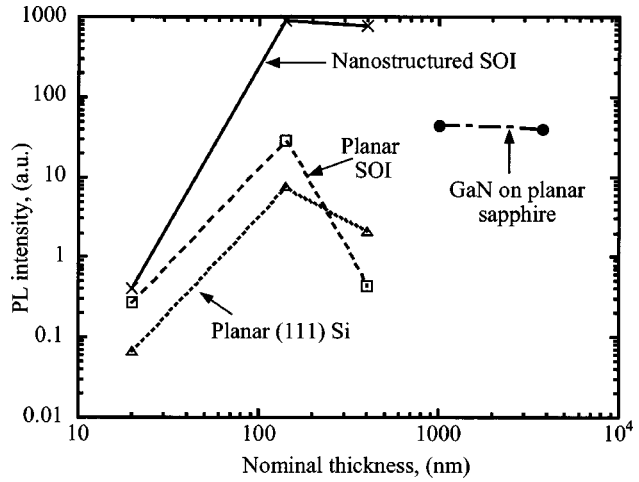


FIG. 8. Room temperature, band gap PL intensity of GaN grown on the substrates indicated as a function of GaN thickness. Note both the much larger intensity for the NHE sample and the decay in the PL intensity for the bulk film cases indicating an increase in nonradiative defects for the thicker growths. The PL intensity for a GaN on sapphire sample is also shown.

important because it establishes that it is possible to achieve coalescence without grain boundaries or defects propagating through the epilayer.

Room temperature PL intensity results at the ~365 nm band edge are plotted in Fig. 8 for GaN grown on nanostructured and planar substrates. The 244 nm wavelength pump power density was ~700 mW/cm^2. The data show that the PL intensity on the nanostructured substrate was not only brighter (by almost 2 orders of magnitude), but its brightness remained fairly constant with film thickness while the intensity on the planar ⟨111⟩ silicon and ⟨111⟩ SOI substrates was much weaker and decreased with increasing thickness above 140 nm. PL spectra also showed almost no mid-gap PL (''yellow band'') for the NHE samples while this band was much more intense in the bulk silicon and SOI samples. A sample with surface relief such as the NHE sample is expected to be brighter than a sample with a smooth planar surface because the scattering from the structure reduces the impact of total internal reflection for a high-index material such as GaN. However, these effects should be decreased for the thicker, more coalesced, sample. The fact that the intensity remained relatively constant in the thickness range from 140 to 400 nm in the NHE samples suggests that the brighter intensity is not solely attributable to these scattering effects but also reflects a high quantum efficiency for the NHE sample.

The decrease in the intensity for the planar ⟨111⟩ silicon and SOI substrates is possibly due to increased defect density and accompanying nonradiative recombination in these materials due to the increased stresses characteristic of growth on misfit planar structures. In contrast, the fact that the intensity in the nanostructured sample is much stronger and remains relatively constant indicates that this material is of higher quality and does not become defected as the thickness increases. This is supported by SEM and TEM data. The PL data support nanoheteroepitaxial theory, which states that the material is essentially stress free beyond the characteristic length of decay. In addition, the PL intensity of the nanostructured sample was also ~25 times brighter as compared to that of a thick GaN film grown on planar sapphire[21] as is also shown in Fig. 8.

V. SUMMARY

We have introduced a novel heterostructure growth technique, nanoheterepitaxy, which shows great promise in gaining fundamental insights into nanocrystalline growth and advancing toward the grand challenge of integrating additional functionality with silicon circuitry. The critical island diameter $2l_c$ has been established to be $\sim(15-30)h_c$. The proximity of the free surface at the island edge makes the nucleation and propagation of defects fundamentally different from the traditional growth case where the top of the growing film is the closest free surface.

Interferometric lithography is proposed as a flexible and cost-effective technique for fabricating the large-area uniform nanoisland substrate arrays required by NHE. SOI island diameters down to 20 nm on a 360 nm pitch are demonstrated using IL and nonlinear develop/etch processes.

Initial experiments of NHE growth of GaN on SOI shows an improvement in material quality versus growth on planar substrates. The SEM and TEM results show a dramatic reduction in threading defect concentration and show promise for coalescence of the GaN nanoislands into a continuous film. PL measurements show that the emission from the NHE samples is much stronger (~100×) and exhibits much less midgap, defect-related emission.

These preliminary results are very encouraging and suggest that NHE may play a role in advancing toward this semiconductor-materials challenge. Much more work is necessary to completely assess the potential of this new technique over a variety of heterostructure material systems, and to extend these results from early materials studies to device investigations.

ACKNOWLEDGMENTS

Partial support for this work was provided by the Air Force Office of Scientific Research and by the Multidisciplinary University Research Initiative of the Department of Defense. The authors thank Wenjun Fan for fabricating the 20 nm SOI nanoisland samples.

[1]S. F. Fang, K. Adomi, S. Iyer, H. Morkoc, H. Zabel, C. Choi, and N. Otsuka, J. Appl. Phys. **68**, R31 (1990).
[2]H. Kroemer, Mater. Res. Soc. Symp. Proc. **67**, 3 (1986).
[3]T. Egawa, Y. Murata, T. Jimbo, and M. Umeno, IEEE Photonics Technol. Lett. **9**, 87 (1997).
[4]D. G. Deppe, N. Holonyak, Jr., D. W. Nam, K. C. Hsieh, G. S. Jackson, R. J. Matyi, H. Shichijo, J. E. Epler, and H. F. Chung, Appl. Phys. Lett. **51**, 637 (1987).
[5]Y. H. Lo, Appl. Phys. Lett. **59**, 2311 (1991).
[6]Y. H. Lo and Z. H. Zhu, Mater. Res. Soc. Symp. Proc. **510**, 81 (1998).
[7]O. Parillaud, E. Gil-Lafon, B. Gerard, P. Etienne, and D. Pribat, Appl. Phys. Lett. **68**, 2654 (1996).
[8]D. Kapolnek, S. Keller, R. Vetury, R. D. Underwood, P. Kozodoy, S. P. DenBaars, and U. K. Mishra, Appl. Phys. Lett. **71**, 1204 (1997).

[9]D. Zubia and S. D. Hersee, J. Appl. Phys. **85**, 6492 (1999).
[10]D. Zubia, S. H. Zaidi, S. R. J. Brueck, and S. D. Hersee, Appl. Phys. Lett. **76**, 858 (2000).
[11]R. People and J. C. Bean, Appl. Phys. Lett. **47**, 322 (1985).
[12]J. P. Hirth and A. G. Evans, J. Appl. Phys. **60**, 2372 (1986).
[13]J. P. Hirth and J. Lothe, *Theory of Dislocations* (Wiley, New York, 1982).
[14]E. H. Anderson, C. M. Horowitz, and H. I. Smith, Appl. Phys. Lett. **43**, 874 (1983).
[15]S. H. Zaidi and S. R. J. Brueck, J. Vac. Sci. Technol. B **11**, 658 (1993).
[16]J. P. Spallas, A. M. Hawryluk, and D. R. Kania, J. Vac. Sci. Technol. B **13**, 1973 (1995).
[17]X. Chen, Z. Zhang, S. R. J. Brueck, R. A. Carpio, and J. S. Petersen, Proc. SPIE **3048**, 309 (1997).
[18]S. H. Zaidi and S. R. J. Brueck, Proc. SPIE **3676**, 371 (1999).
[19]X. Chen, S. H. Zaidi, S. R. J. Brueck, and D. J. Devine, J. Vac. Sci. Technol. B **14**, 3339 (1996).
[20]S. D. Hersee, J. Ramer, K. Zheng, C. Kranenberg, K. Malloy, M. Banas, and M. Goorsky, J. Electron. Mater. **24**, 1519 (1995).
[21]J. C. Ramer, Ph.D. thesis, University of New Mexico, 1997.

Initial Nanoheteroepitaxial Growth of GaAs on Si(100) by OMVPE

D. ZUBIA,[1,2,4] S. ZHANG,[1,2] R. BOMMENA,[1,2] X. SUN,[1,2] S.R.J. BRUECK,[1,2,3] and S.D. HERSEE[1,2]

1.—University of New Mexico, Center for High Technology Materials, Albuquerque, NM 87106.
2.—Also University of New Mexico, Department of Electrical and Computer Engineering.
3.—Also University of New Mexico, Department of Physics and Astronomy.
4.—e-mail: davzubia@chtm.unm.edu.

Initial growth studies of GaAs on an array of Si islands nanostructured on (100) oriented silicon-on-insulator substrates show that growth occurs through a mixture of selective-area and 3D growth modes. An optimum initiation growth temperature must tune the growth conditions to the geometry of the seed array so that selective-area control is maintained while defect density is minimized. The optimum temperature for a square array of Si islands, 500 nm in pitch, and 100 nm to 280 nm in diameter, is ~600°C. This temperature yields single-crystal nucleation on each Si island while maintaining selective-area growth mode control. Transmission electron microscope (TEM) analysis of optimized and non-optimized grown GaAs/Si heterostructures show that they accommodate 0.4–0.7% strain. Further reduction in stacking-fault defects attributed to side wall growth may be possible through masking of side wall or annealing.

Key words: Heteroepitaxy, GaAs on Si, growth initiation, lattice mismatch

INTRODUCTION

Nanoheteroepitaxy has recently been proposed as a technique for significantly reducing defect density in lattice mismatched heteroepitaxy.[1,2] Nanoheteroepitaxy incorporates a number of defect reduction techniques including: substrate patterning, lateral epitaxial overgrowth,[3–5] strain partitioning in thin compliant substrates,[6,7] and epitaxial necking,[8] which exploits free surfaces to terminate dislocations. In addition, nanoheteroepitaxy exploits three-dimensional stress relief mechanisms that are available only in nanoscale heterostructures, where the strain energy density decays rapidly away from the interface.[9] For example, nanoheteroepitaxy theory predicts that growth of GaAs on Si (~4% misfit) should support a coherent interface provided the island size is below the critical island diameter ($2l_c$) of 40 nm.[2] Nanoheteroepitaxy addresses highly lattice mismatched and polarity mismatched systems that are expected to grow in Stranski-Krastanov or Volmer-Weber mode. One example is the growth of GaAs on Si which becomes three-dimensional (3D) at an early stage of film growth (<7 nm) to minimize the strain energy.[10] 3D growth is usually undesirable because it leads to morphological problems and one approach to limit its effects has been to employ a low-temperature buffer layer on planar,[11,12] patterned,[13] and porous[14] substrates to achieve smooth continuous films. This approach has been successfully applied to GaN heteroepitaxy,[15] where atomic-force microscopy measurements show that a low temperature buffer is also 3D but on a very fine scale. This finer scale of 3D growth is presumably related to the reduced surface mobility of the growth species at the lower temperature. However, a tradeoff appears to exist between low defect density and smooth film morphology,[10] where the finer scale 3D growth results in a smoother film but can cause a higher defect density as misorientation between the grains will lead to dislocations and low angle grain boundaries when they coalesce. In nanoheteroepitaxy, the seed diameter is in the nanoscale, in the range of the domain size of crystals grown at normal temperatures. Therefore, it may be possible to grow an ordered array of single-crystal and, perhaps, coherent GaAs crystals on Si using nanoheteroepitaxy. This paper describes observations of the initial stages of nanoheteroepitaxial growth of GaAs on nanoscale-patterned silicon-on-insulator substrates.

(Received January 21, 2001; accepted March 24, 2001)

EXPERIMENTAL

Epitaxial GaAs was grown on nanostructured silicon islands patterned on (100) oriented silicon-on-insulator substrates using organometallic vapor-phase epitaxy. Four stages of the growth process were identified as (1) growth initiation, (2) lateral growth, (3) coalescence, and (4) planarization. This study emphasizes stages (1), (2) and, in particular, addresses the effect that growth initiation has on the quality of GaAs during lateral growth.

The surface silicon layer of (100) oriented silicon-on-insulator substrates was patterned using interferometric lithography, a maskless, large-area patterning technology. For details on interferometric lithography please refer to Ref. 16. The initial thickness of the surface silicon layer was 200 nm. The thickness of the oxide layer was 370 nm. The surface silicon layer was thinned to 110 nm by an oxidation and hydrofluoric acid (HF) etching sequence. A square array of silicon islands was then patterned on the surface silicon layer. The pitch of the array was 500 nm and the diameter of the islands varied between 100 and 280 nm from batch to batch. Figure 1 contains a plan view SEM image of a nanostructured (100)silicon-on-insulator (SOI) wafer. The inset in Fig. 1 is a schematic cross-sectional view of the nanostructured (100)SOI illustrating the different layers in the substrate. Unintentional variations in processing caused the shape of the islands to vary from diamond to circular (see Fig. 2); however, for consistency the lateral extent or size of the islands is referred to as the diameter.

Arsine and trimethylgallium were used as precursors for group V and III sources, respectively. Reactor parameters common to all stages were the reactor pressure (100 torr), the carrier gas (hydrogen), and the total gas flow rate (5.9 slm). The arsine flow rate was kept constant at 35 sccm in all growth stages (1) to (4).

Outgassing and oxide desorption, which is normally performed at a temperature of 900–1000°C [11] was limited to 860°C for 10 min in hydrogen due to gross morphological changes that occur in nanostructured[17] and porous[18] silicon samples at temperatures above ~1000°C and 750°C, respectively. The temperature was then reduced in preparation of the growth initiation stage. In order to avoid antiphase domain boundaries, an arsenic terminated silicon surface[12] was created by exposing silicon to arsine at a rate of 35 sccm for 2 min immediately prior to growth of GaAs. In this stage, four different growth temperatures were tested between 430°C to 730°C to study the effects of growth of this stage on material grown in later stages. The trimethylgallium flow rate was between 1 and 2.4 sccm. This gave a V/III molar ratio of ~580 (this ratio is calculated based on the input flows of the constituents. The reduced decomposition rate of arsine at the lower growth temperature no doubt reduced the actual V/III ratio during the lower temperature growths.) The initial growth rate (stage 1) was ~15 nm/min. During lateral growth

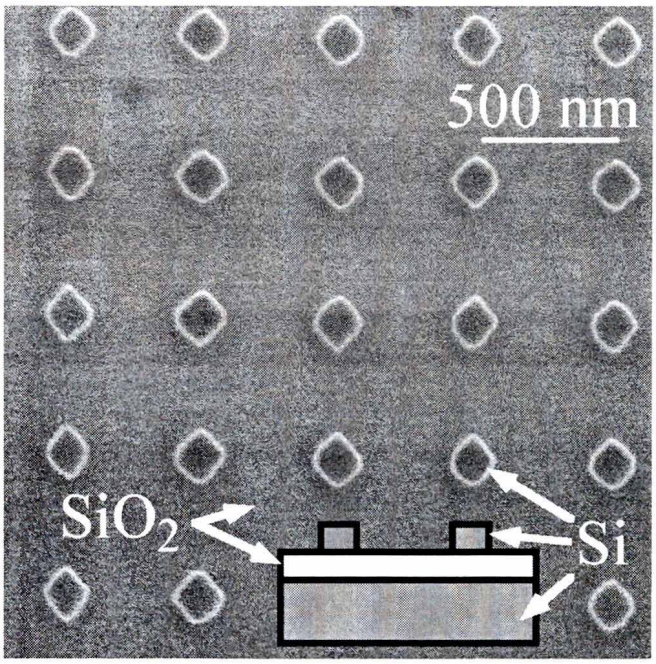


Fig. 1. Plan view SEM image of Si array of nanoislands patterned on (100) oriented silicon-on-insulator. Inset is cross-sectional schematic of nanostructured (100) silicon-on-insulator substrate.

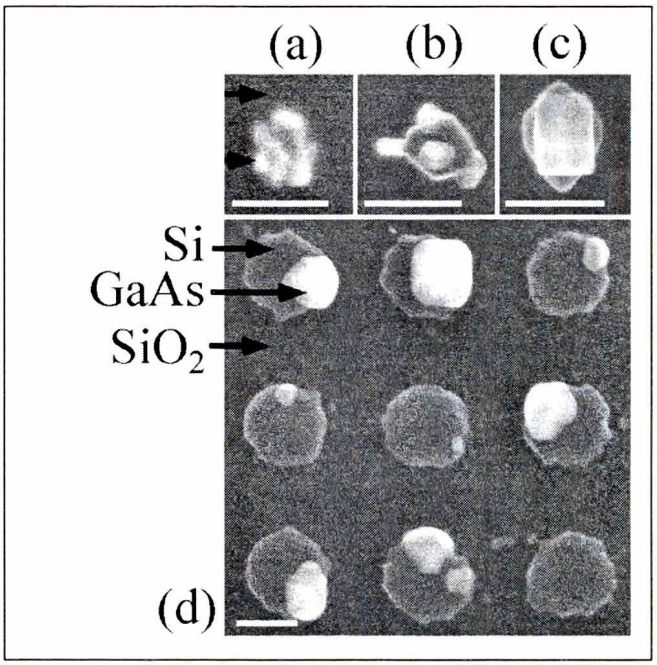


Fig. 2. Plan view SEM images of GaAs nucleation on silicon islands at a growth initiation temperature of (a) 433°C, (b) 519°C, (c) 605°C, and (d) 734°C. All scale bars indicate 200 nm.

(stage 2) the temperature was 605°C and the trimethylgallium flow rate was 8 sccm.

A Hitachi S-800 field-emission, scanning electron microscope (SEM) operating at 20 kV and a JOEL 2010 transmission electron microscope (TEM) operating at 200 kV were used to characterize the GaAs films.

RESULTS

Figure 2 shows the normal incidence SEM images

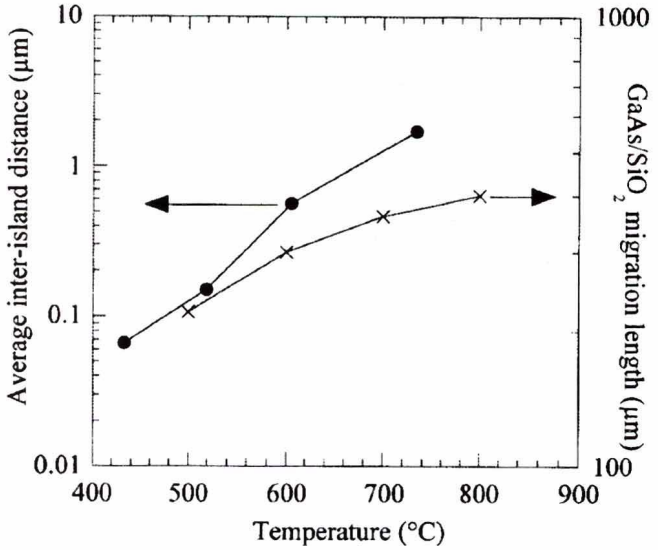


Fig. 3. Average inter-island distance of GaAs nuclei on planar Si and migration length of GaAs on SiO_2 versus temperature.

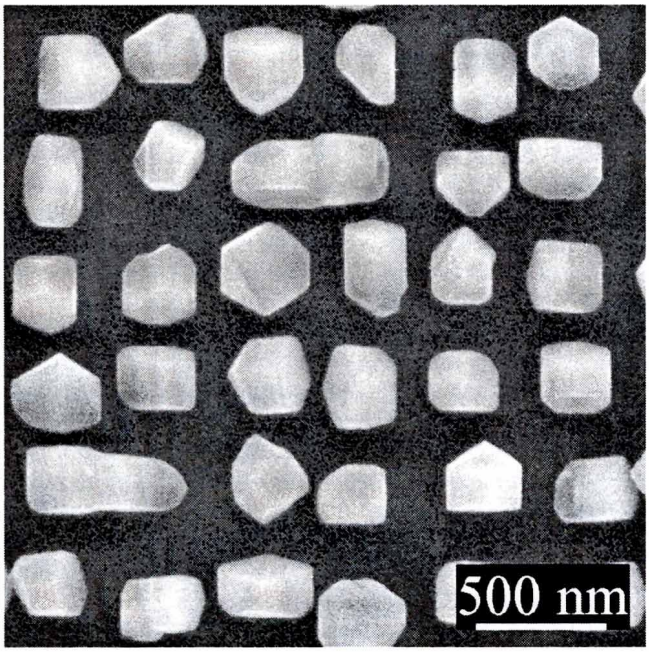


Fig. 4. Plan view SEM image of an array of GaAs crystals deposited on nanostructured (100) oriented silicon-on-insulator. Approximately 100 nm of material was deposited at a temperature of 605°C. Coalescence of two pairs of crystals occurred during GaAs growth and is not due to the patterning.

of GaAs nucleation behavior after approximately 15 nm of deposition on nanostructured silicon islands at substrate temperatures of T_s = 433, 519, 605, and 734°C. At a low substrate temperature of 433°C, multiple (5 to 8) GaAs crystals with grain size approximately 25 nm in diameter nucleate on each silicon island as shown in Fig. 2a. As the temperature increases the number of nuclei per silicon island decreases, until at 605°C only one GaAs crystal nucleates per silicon island (Fig. 2c). However if the temperature is too high [734°C nucleation occurs only on some silicon islands (see Fig. 2d)]. It appears that there is indeed a tendency to 3D growth; however, the size and spacing of the nuclei is influenced by surface mobility, which is controlled largely by surface diffusion and therefore growth temperature. Figure 3 plots data from related experiments, where the average inter-island distance of GaAs nuclei was measured for growth on planar Si(100) as a function of temperature. The temperature dependence of island spacing is quite strong and indicates that nucleation is diffusion controlled.[10] This figure also plots the surface migration length of GaAs on SiO_2 from Ref. 19. Clearly, the surface migration length of GaAs on SiO_2 is for all temperatures much larger than the pitch of the nanostructured islands (500 nm) so the growth species has an adequate migration length for selective growth as indicated in Fig. 2. In contrast, the average inter-island distance for GaAs on Si is of the same order as the silicon island diameter and pitch. This is consistent with the results of Fig. 2, where, at low growth temperature, the GaAs has a low mobility on Si and multiple nuclei are observed on each substrate island. While at high temperature, the GaAs mobility is much higher and the growth species may have too much mobility. For example, at 734°C, the average inter-island distance is 1700 nm and growth selectivity is lost because the large surface migration length leaves some of the islands denuded. This suggests that there is an optimum temperature for achieving selective growth in the nanoheteroepitaxial (NHE) process. Under optimum conditions the growth species will have adequate mobility to reach a Si island and to create a single nucleus on each Si island, but they will not have excessive mobility that can leave some Si islands without any GaAs at all. Based on Figs. 2 and 3, this optimum temperature is approximately 605°C. Figure 4 shows an array of GaAs NHE nuclei grown on silicon islands after ~100 nm of deposition at 605°C. Some coalescing can be observed in the image. Well-aligned rectangular faceting can also be observed indicative of epitaxy although several crystal habits are present.

Cross-sectional TEM analysis of NHE GaAs nuclei on Si islands reveals that the temperature during growth initiation has a strong effect on the defect structure. Figure 5 shows cross-sectional, diffraction contrast images taken along the [011] zone axis. This figure compares samples that were all grown at 605°C during the lateral stage (stage 2) but had a different temperature during the growth initiation stage (stage 1). All samples had ~100 nm of GaAs deposition. Figure 5a (stage 1 at 433°C) and 5b (stage 1 at 519°C) were typical of several GaAs/Si islands observed that had a growth initiation at either 433°C or 519°C. In the island in Fig. 5a stacking faults can be observed. Notice that moiré fringes in the Si area, caused by the superposition of the GaAs and Si lattices, can also be observed. The average period, τ, of the moiré pattern was measured at 5.33 nm near the right-hand side of the Si island. At the top GaAs/Si interface the average period was 5.79 nm. Using a simple analysis, it is speculated that the periods correspond to lattice mis-

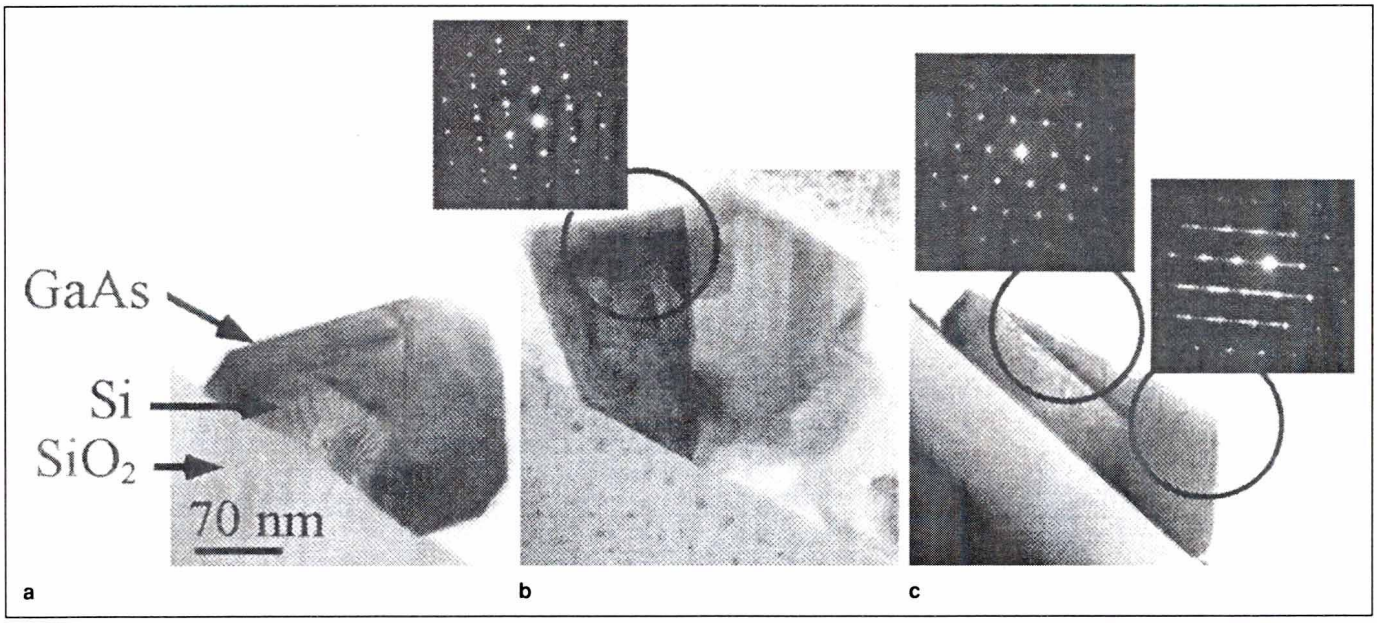


Fig. 5. Cross-section diffraction-contrast images of GaAs/Si heterostructures after 100 nm of GaAs deposition at 605°C during the lateral growth stage but with a growth initiation temperature of (a) 433°C, (b) 519°C, and (c) 605°C. Inset diffraction patterns in (b) and (c) were taken from the areas indicated by circles. No silicon island is observed in (b) because the thinned sample was from an area adjacent to the island. All scale bars indicate 70 nm.

matches of $\varepsilon = d_{Si[011]}/\tau = 3.6\% \pm 0.1\%$ at the edge and $3.3\% \pm 0.1\%$ at the interface, where $d_{Si[011]} = 0.192$ nm is the lattice plane spacing of Si in the [011] direction. Thus the materials are accommodating 0.4–0.7% strain. The diffraction pattern in Fig. 5b was taken approximately from the area indicated by the circle and contains two sets of diffraction spots indicating that twinning has occurred. In addition a low-angle grain boundary runs diagonally across the island as revealed by difference in contrast in the GaAs. In contrast, the material with a growth initiation temperature of 605°C (Fig. 5c) had no low-angle grain boundaries although stacking faults can still be observed. Figure 5c contains two diffraction patterns. One pattern taken approximately from the defected area contains streaks, confirming that the defects are stacking faults. Another diffraction pattern taken approximately from the GaAs/Si interface area contains two sets of spots, one created by the GaAs and the other by the Si. Analysis of the diffraction pattern taken from the interface shows a 0.8° misalignment between GaAs[100] and Si[100] directions. Measurement of the distance from the origin, 000, to the 002 spots show a ~3.5% lattice mismatch agreeing well with the moiré fringe in Fig. 5a. Finally, inspection of Fig. 5a and c show that GaAs grows both on the top Si(100) surface and on the side walls of the silicon islands. The image in Fig. 5c was typical of many GaAs/Si islands observed, showing that most of the defects originate from the side walls.

CONCLUSIONS

Four stages of the growth process were identified as (1) growth initiation, (2) lateral growth, (3) coalescence, and (4) planarization. Growth occurs through a mixture of selective-area growth (driven by the differential surface mobility of GaAs on SiO_2 versus Si), and 3D growth of GaAs on Si (controlled by surface diffusion and strain energy considerations). Due to the high surface mobility of GaAs on SiO_2, GaAs grows selectively on the nanostructured silicon islands and not on the SiO_2 mask at all temperatures tested. The temperature of the initial growth controls the surface diffusion length and ideally should be optimized so that single crystal nucleation occurs on each silicon island. If the temperature is too high, single crystal nucleation occurs only on a fraction of the total silicon islands with other islands left denuded. If the temperature is too low, multiple-crystal nucleation will occur on each island leading to grain boundaries and twinning. The optimum temperature for the nanostructured substrates utilized in this study is approximately 605°C.

The GaAs/Si heterostructures are partially strained to 0.4–0.7%, although the data does not reveal whether strain relaxation occurred through misfit dislocations or three-dimensional strain relief mechanisms. TEM images show that GaAs grows both on the top Si(100) and side wall surfaces and indicate that stacking faults originate from the side walls. Covering the side walls with a mask to prevent growth might be helpful in eliminating the stacking faults. Alternatively, annealing might also reduce the defects; however, due to gross morphological changes, the annealing temperature must be kept below 850°C.

ACKNOWLEDGEMENTS

Partial support for this work was provided by the Air Force Office of Scientific Research and by the Multidisciplinary University Research Initiative of the Department of Defense. Part of this research was conducted at the Transmission Electron Microscopy

Laboratory in the Department of Earth and Planetary Sciences of the University of New Mexico, which is supported by NSF, NASA, and the State of New Mexico.

REFERENCES

1. D. Zubia and S.D. Hersee, *J. Appl. Phys.* 85, 6492 (1999).
2. D. Zubia, S.H. Zaidi, S.D. Hersee, and S.R.J. Brueck, *J. Vac. Sci. Technol. B* 18, 3514 (2000).
3. O. Parillaud, E. Gil-Lafon, B. Gerard, P. Etienne, and D. Pribat, *Appl. Phys. Lett.* 68, 2654 (1996).
4. D. Kapolnek, S. Keller, R. Vetury, R.D. Underwood, P. Kozodoy, S.P. Denbaars, and U.K. Mishra, *Appl. Phys. Lett.* 71, 1204 (1997).
5. O.-H. Nam, M.D. Bremser, T.S. Zheleva, and R.F. Davis, *Appl. Phys. Lett.* 71, 2638 (1997).
6. Y.H. Lo, *Appl. Phys. Lett.* 59, 2311 (1991).
7. P.D. Moran, D.M. Hansen, R.J. Matyi, J.G. Cederberg, L.J. Mawst, and T.F. Kuech, *Appl. Phys. Lett.* 75, 1559 (1999).
8. T.A. Langdo, C.W. Leitz, M.T. Currie, E.A. Fitzgerald, A. Lochtefeld, and D.A. Antoniadis, *Appl. Phys. Lett.* 76, 3700 (2000).
9. S.A. Chaparro, Y. Zhang, and J. Drucker, *Appl. Phys. Lett.* 76, 3534 (2000).
10. D.K. Biegelsen, F.A. Ponce, A.J. Smith, and J.C. Tramontana, *J. Appl. Phys.* 61, 1856 (1987).
11. J.W. Lee, J.P. Salerno, R.P. Gale, and J.C.C. Fan, *Mater. Res. Soc. Symp. Proc.* 91, 33 (1987).
12. S.F. Fang, A.S. Iyer, H. Morkoc, H. Zabel, C. Choi, and N. Otsuka, *J. Appl. Phys.* 68, R31 (1990).
13. R.J. Matyi, H. Shichijo, and H.L. Tsai, *J. Vac. Sci. Technol. B* 6, 699 (1988).
14. S. Hasegawa, K. Maehashi, H. Nakashima, T. Ito, and A. Hiraki, *J. Cryst. Growth* 95, 113 (1989).
15. S.D. Hersee, J.C. Ramer, and K.J. Malloy, *MRS Bulletin* 22, 45 (1997).
16. X. Chen, S.H. Zaidi, S.R.J. Brueck, and D.J. Devine, *J. Vac. Sci. Technol. B* 14, 3339 (1996).
17. D. Zubia, S.H. Zaidi, S.R.J. Brueck, and S.D. Hersee, *Appl. Phys. Lett.* 76, 858-860 (2000).
18. T.L. Lin, L. Sadwick, K.L. Wang, Y.C. Kao, R. Hull, C.W. Nieh, D.N. Jamieson, and J.K. Liu, *Appl. Phys. Lett.* 51, 814 (1987).
19. K. Kiruma, T. Haga, and M. Miyazaki, *J. Cryst. Growth* 102, 717 (1990).

Initial Nanoheteroepitaxial Growth of GaAs on Si(100) by OMVPE

D. ZUBIA,[1,2,4] S. ZHANG,[1,2] R. BOMMENA,[1,2] X. SUN,[1,2] S.R.J. BRUECK,[1,2,3] and S.D. HERSEE[1,2]

1.—University of New Mexico, Center for High Technology Materials, Albuquerque, NM 87106.
2.—Also University of New Mexico, Department of Electrical and Computer Engineering.
3.—Also University of New Mexico, Department of Physics and Astronomy.
4.—e-mail: davzubia@chtm.unm.edu.

Initial growth studies of GaAs on an array of Si islands nanostructured on (100) oriented silicon-on-insulator substrates show that growth occurs through a mixture of selective-area and 3D growth modes. An optimum initiation growth temperature must tune the growth conditions to the geometry of the seed array so that selective-area control is maintained while defect density is minimized. The optimum temperature for a square array of Si islands, 500 nm in pitch, and 100 nm to 280 nm in diameter, is ~600°C. This temperature yields single-crystal nucleation on each Si island while maintaining selective-area growth mode control. Transmission electron microscope (TEM) analysis of optimized and non-optimized grown GaAs/Si heterostructures show that they accommodate 0.4–0.7% strain. Further reduction in stacking-fault defects attributed to side wall growth may be possible through masking of side wall or annealing.

Key words: Heteroepitaxy, GaAs on Si, growth initiation, lattice mismatch

INTRODUCTION

Nanoheteroepitaxy has recently been proposed as a technique for significantly reducing defect density in lattice mismatched heteroepitaxy.[1,2] Nanoheteroepitaxy incorporates a number of defect reduction techniques including: substrate patterning, lateral epitaxial overgrowth,[3–5] strain partitioning in thin compliant substrates,[6,7] and epitaxial necking,[8] which exploits free surfaces to terminate dislocations. In addition, nanoheteroepitaxy exploits three-dimensional stress relief mechanisms that are available only in nanoscale heterostructures, where the strain energy density decays rapidly away from the interface.[9] For example, nanoheteroepitaxy theory predicts that growth of GaAs on Si (~4% misfit) should support a coherent interface provided the island size is below the critical island diameter ($2l_c$) of 40 nm.[2] Nanoheteroepitaxy addresses highly lattice mismatched and polarity mismatched systems that are expected to grow in Stranski-Krastanov or Volmer-Weber mode. One example is the growth of GaAs on Si which becomes three-dimensional (3D) at an early stage of film growth (<7 nm) to minimize the strain energy.[10] 3D growth is usually undesirable because it leads to morphological problems and one approach to limit its effects has been to employ a low-temperature buffer layer on planar,[11,12] patterned,[13] and porous[14] substrates to achieve smooth continuous films. This approach has been successfully applied to GaN heteroepitaxy,[15] where atomic-force microscopy measurements show that a low temperature buffer is also 3D but on a very fine scale. This finer scale of 3D growth is presumably related to the reduced surface mobility of the growth species at the lower temperature. However, a tradeoff appears to exist between low defect density and smooth film morphology,[10] where the finer scale 3D growth results in a smoother film but can cause a higher defect density as misorientation between the grains will lead to dislocations and low angle grain boundaries when they coalesce. In nanoheteroepitaxy, the seed diameter is in the nanoscale, in the range of the domain size of crystals grown at normal temperatures. Therefore, it may be possible to grow an ordered array of single-crystal and, perhaps, coherent GaAs crystals on Si using nanoheteroepitaxy. This paper describes observations of the initial stages of nanoheteroepitaxial growth of GaAs on nanoscale-patterned silicon-on-insulator substrates.

(Received January 21, 2001; accepted March 24, 2001)

EXPERIMENTAL

Epitaxial GaAs was grown on nanostructured silicon islands patterned on (100) oriented silicon-on-insulator substrates using organometallic vapor-phase epitaxy. Four stages of the growth process were identified as (1) growth initiation, (2) lateral growth, (3) coalescence, and (4) planarization. This study emphasizes stages (1), (2) and, in particular, addresses the effect that growth initiation has on the quality of GaAs during lateral growth.

The surface silicon layer of (100) oriented silicon-on-insulator substrates was patterned using interferometric lithography, a maskless, large-area patterning technology. For details on interferometric lithography please refer to Ref. 16. The initial thickness of the surface silicon layer was 200 nm. The thickness of the oxide layer was 370 nm. The surface silicon layer was thinned to 110 nm by an oxidation and hydrofluoric acid (HF) etching sequence. A square array of silicon islands was then patterned on the surface silicon layer. The pitch of the array was 500 nm and the diameter of the islands varied between 100 and 280 nm from batch to batch. Figure 1 contains a plan view SEM image of a nanostructured (100)silicon-on-insulator (SOI) wafer. The inset in Fig. 1 is a schematic cross-sectional view of the nanostructured (100)SOI illustrating the different layers in the substrate. Unintentional variations in processing caused the shape of the islands to vary from diamond to circular (see Fig. 2); however, for consistency the lateral extent or size of the islands is referred to as the diameter.

Arsine and trimethylgallium were used as precursors for group V and III sources, respectively. Reactor parameters common to all stages were the reactor pressure (100 torr), the carrier gas (hydrogen), and the total gas flow rate (5.9 slm). The arsine flow rate was kept constant at 35 sccm in all growth stages (1) to (4).

Outgassing and oxide desorption, which is normally performed at a temperature of 900–1000°C [11] was limited to 860°C for 10 min in hydrogen due to gross morphological changes that occur in nanostructured[17] and porous[18] silicon samples at temperatures above ~1000°C and 750°C, respectively. The temperature was then reduced in preparation of the growth initiation stage. In order to avoid antiphase domain boundaries, an arsenic terminated silicon surface[12] was created by exposing silicon to arsine at a rate of 35 sccm for 2 min immediately prior to growth of GaAs. In this stage, four different growth temperatures were tested between 430°C to 730°C to study the effects of growth of this stage on material grown in later stages. The trimethylgallium flow rate was between 1 and 2.4 sccm. This gave a V/III molar ratio of ~580 (this ratio is calculated based on the input flows of the constituents. The reduced decomposition rate of arsine at the lower growth temperature no doubt reduced the actual V/III ratio during the lower temperature growths.) The initial growth rate (stage 1) was ~15 nm/min. During lateral growth

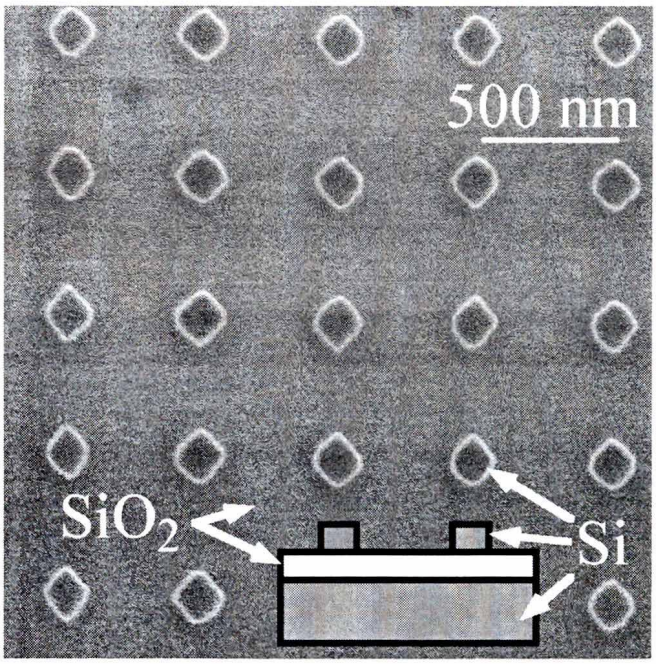


Fig. 1. Plan view SEM image of Si array of nanoislands patterned on (100) oriented silicon-on-insulator. Inset is cross-sectional schematic of nanostructured (100) silicon-on-insulator substrate.

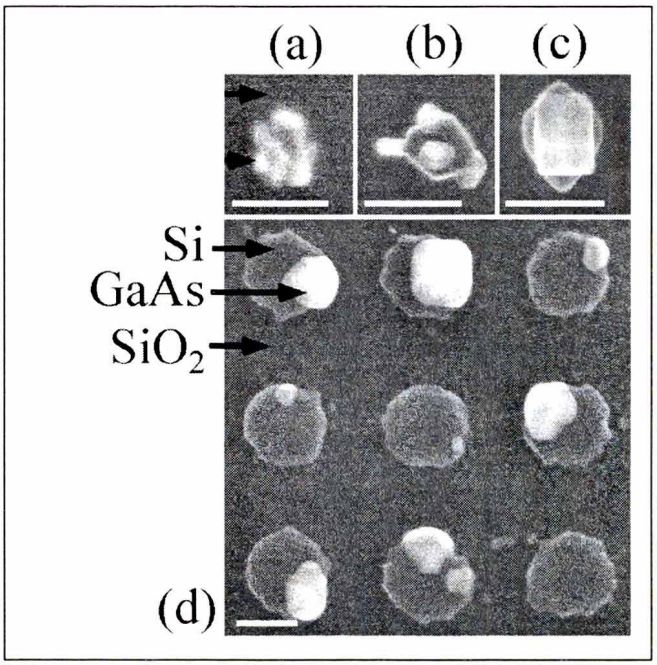


Fig. 2. Plan view SEM images of GaAs nucleation on silicon islands at a growth initiation temperature of (a) 433°C, (b) 519°C, (c) 605°C, and (d) 734°C. All scale bars indicate 200 nm.

(stage 2) the temperature was 605°C and the trimethylgallium flow rate was 8 sccm.

A Hitachi S-800 field-emission, scanning electron microscope (SEM) operating at 20 kV and a JOEL 2010 transmission electron microscope (TEM) operating at 200 kV were used to characterize the GaAs films.

RESULTS

Figure 2 shows the normal incidence SEM images

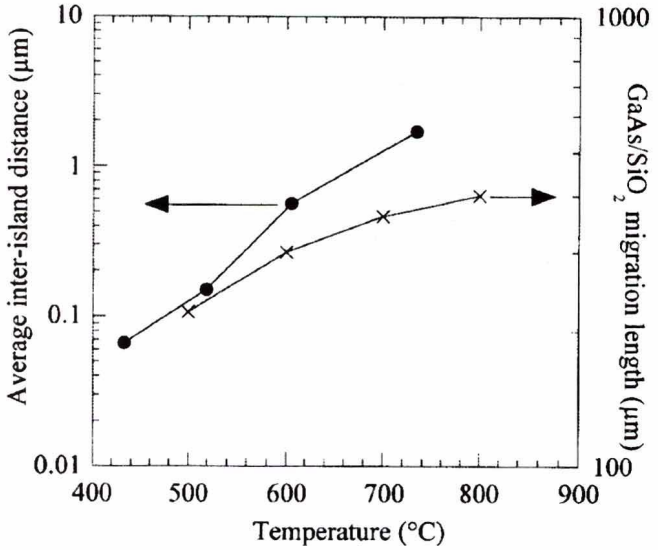


Fig. 3. Average inter-island distance of GaAs nuclei on planar Si and migration length of GaAs on SiO_2 versus temperature.

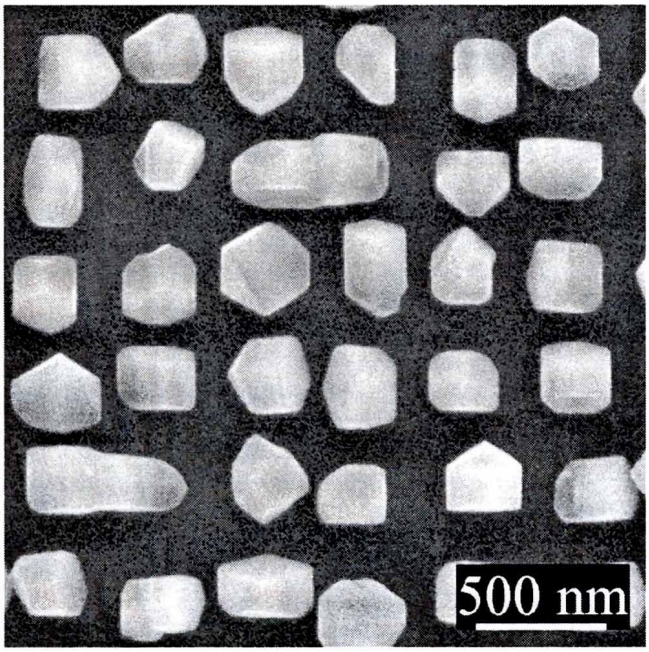


Fig. 4. Plan view SEM image of an array of GaAs crystals deposited on nanostructured (100) oriented silicon-on-insulator. Approximately 100 nm of material was deposited at a temperature of 605°C. Coalescence of two pairs of crystals occurred during GaAs growth and is not due to the patterning.

of GaAs nucleation behavior after approximately 15 nm of deposition on nanostructured silicon islands at substrate temperatures of T_s = 433, 519, 605, and 734°C. At a low substrate temperature of 433°C, multiple (5 to 8) GaAs crystals with grain size approximately 25 nm in diameter nucleate on each silicon island as shown in Fig. 2a. As the temperature increases the number of nuclei per silicon island decreases, until at 605°C only one GaAs crystal nucleates per silicon island (Fig. 2c). However if the temperature is too high [734°C nucleation occurs only on some silicon islands (see Fig. 2d)]. It appears that there is indeed a tendency to 3D growth; however, the size and spacing of the nuclei is influenced by surface mobility, which is controlled largely by surface diffusion and therefore growth temperature. Figure 3 plots data from related experiments, where the average inter-island distance of GaAs nuclei was measured for growth on planar Si(100) as a function of temperature. The temperature dependence of island spacing is quite strong and indicates that nucleation is diffusion controlled.[10] This figure also plots the surface migration length of GaAs on SiO_2 from Ref. 19. Clearly, the surface migration length of GaAs on SiO_2 is for all temperatures much larger than the pitch of the nanostructured islands (500 nm) so the growth species has an adequate migration length for selective growth as indicated in Fig. 2. In contrast, the average inter-island distance for GaAs on Si is of the same order as the silicon island diameter and pitch. This is consistent with the results of Fig. 2, where, at low growth temperature, the GaAs has a low mobility on Si and multiple nuclei are observed on each substrate island. While at high temperature, the GaAs mobility is much higher and the growth species may have too much mobility. For example, at 734°C, the average inter-island distance is 1700 nm and growth selectivity is lost because the large surface migration length leaves some of the islands denuded. This suggests that there is an optimum temperature for achieving selective growth in the nanoheteroepitaxial (NHE) process. Under optimum conditions the growth species will have adequate mobility to reach a Si island and to create a single nucleus on each Si island, but they will not have excessive mobility that can leave some Si islands without any GaAs at all. Based on Figs. 2 and 3, this optimum temperature is approximately 605°C. Figure 4 shows an array of GaAs NHE nuclei grown on silicon islands after ~100 nm of deposition at 605°C. Some coalescing can be observed in the image. Well-aligned rectangular faceting can also be observed indicative of epitaxy although several crystal habits are present.

Cross-sectional TEM analysis of NHE GaAs nuclei on Si islands reveals that the temperature during growth initiation has a strong effect on the defect structure. Figure 5 shows cross-sectional, diffraction contrast images taken along the [011] zone axis. This figure compares samples that were all grown at 605°C during the lateral stage (stage 2) but had a different temperature during the growth initiation stage (stage 1). All samples had ~100 nm of GaAs deposition. Figure 5a (stage 1 at 433°C) and 5b (stage 1 at 519°C) were typical of several GaAs/Si islands observed that had a growth initiation at either 433°C or 519°C. In the island in Fig. 5a stacking faults can be observed. Notice that moiré fringes in the Si area, caused by the superposition of the GaAs and Si lattices, can also be observed. The average period, τ, of the moiré pattern was measured at 5.33 nm near the right-hand side of the Si island. At the top GaAs/Si interface the average period was 5.79 nm. Using a simple analysis, it is speculated that the periods correspond to lattice mis-

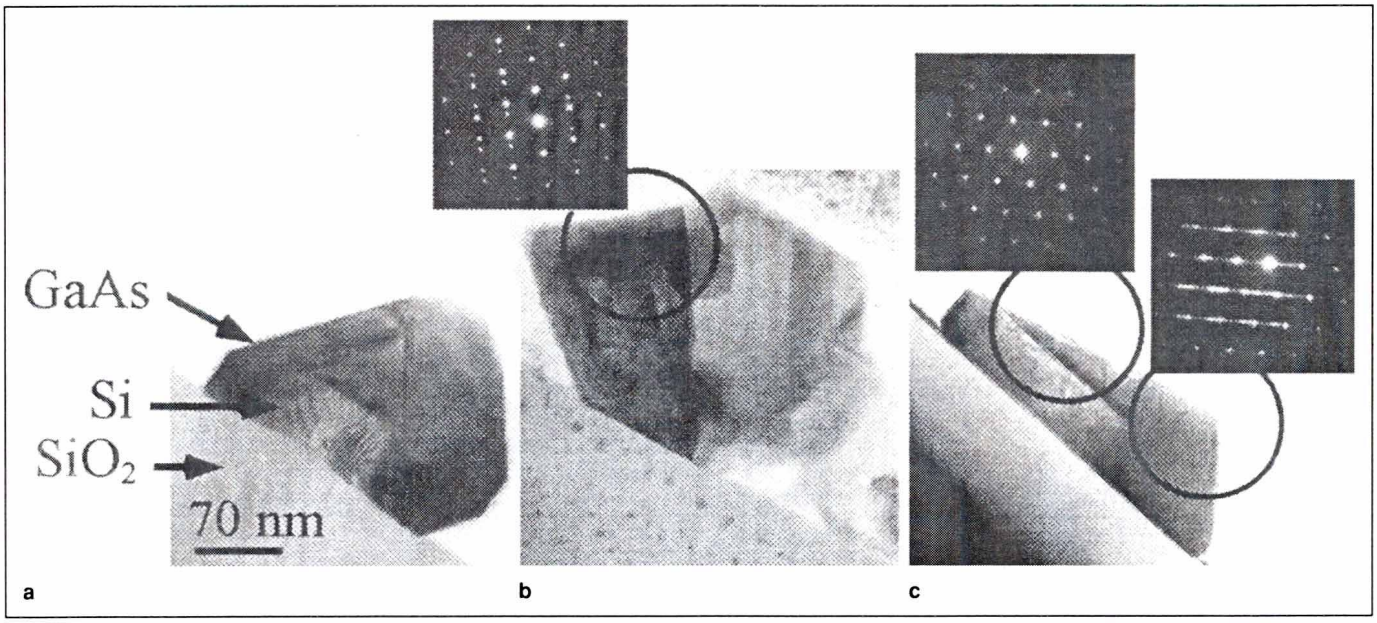


Fig. 5. Cross-section diffraction-contrast images of GaAs/Si heterostructures after 100 nm of GaAs deposition at 605°C during the lateral growth stage but with a growth initiation temperature of (a) 433°C, (b) 519°C, and (c) 605°C. Inset diffraction patterns in (b) and (c) were taken from the areas indicated by circles. No silicon island is observed in (b) because the thinned sample was from an area adjacent to the island. All scale bars indicate 70 nm.

matches of $\varepsilon = d_{Si[011]}/\tau = 3.6\% \pm 0.1\%$ at the edge and $3.3\% \pm 0.1\%$ at the interface, where $d_{Si[011]} = 0.192$ nm is the lattice plane spacing of Si in the [011] direction. Thus the materials are accommodating 0.4–0.7% strain. The diffraction pattern in Fig. 5b was taken approximately from the area indicated by the circle and contains two sets of diffraction spots indicating that twinning has occurred. In addition a low-angle grain boundary runs diagonally across the island as revealed by difference in contrast in the GaAs. In contrast, the material with a growth initiation temperature of 605°C (Fig. 5c) had no low-angle grain boundaries although stacking faults can still be observed. Figure 5c contains two diffraction patterns. One pattern taken approximately from the defected area contains streaks, confirming that the defects are stacking faults. Another diffraction pattern taken approximately from the GaAs/Si interface area contains two sets of spots, one created by the GaAs and the other by the Si. Analysis of the diffraction pattern taken from the interface shows a 0.8° misalignment between GaAs[100] and Si[100] directions. Measurement of the distance from the origin, 000, to the 002 spots show a ~3.5% lattice mismatch agreeing well with the moiré fringe in Fig. 5a. Finally, inspection of Fig. 5a and c show that GaAs grows both on the top Si(100) surface and on the side walls of the silicon islands. The image in Fig. 5c was typical of many GaAs/Si islands observed, showing that most of the defects originate from the side walls.

CONCLUSIONS

Four stages of the growth process were identified as (1) growth initiation, (2) lateral growth, (3) coalescence, and (4) planarization. Growth occurs through a mixture of selective-area growth (driven by the differential surface mobility of GaAs on SiO_2 versus Si), and 3D growth of GaAs on Si (controlled by surface diffusion and strain energy considerations). Due to the high surface mobility of GaAs on SiO_2, GaAs grows selectively on the nanostructured silicon islands and not on the SiO_2 mask at all temperatures tested. The temperature of the initial growth controls the surface diffusion length and ideally should be optimized so that single crystal nucleation occurs on each silicon island. If the temperature is too high, single crystal nucleation occurs only on a fraction of the total silicon islands with other islands left denuded. If the temperature is too low, multiple-crystal nucleation will occur on each island leading to grain boundaries and twinning. The optimum temperature for the nanostructured substrates utilized in this study is approximately 605°C.

The GaAs/Si heterostructures are partially strained to 0.4–0.7%, although the data does not reveal whether strain relaxation occurred through misfit dislocations or three-dimensional strain relief mechanisms. TEM images show that GaAs grows both on the top Si(100) and side wall surfaces and indicate that stacking faults originate from the side walls. Covering the side walls with a mask to prevent growth might be helpful in eliminating the stacking faults. Alternatively, annealing might also reduce the defects; however, due to gross morphological changes, the annealing temperature must be kept below 850°C.

ACKNOWLEDGEMENTS

Partial support for this work was provided by the Air Force Office of Scientific Research and by the Multidisciplinary University Research Initiative of the Department of Defense. Part of this research was conducted at the Transmission Electron Microscopy

Laboratory in the Department of Earth and Planetary Sciences of the University of New Mexico, which is supported by NSF, NASA, and the State of New Mexico.

REFERENCES

1. D. Zubia and S.D. Hersee, *J. Appl. Phys.* 85, 6492 (1999).
2. D. Zubia, S.H. Zaidi, S.D. Hersee, and S.R.J. Brueck, *J. Vac. Sci. Technol. B* 18, 3514 (2000).
3. O. Parillaud, E. Gil-Lafon, B. Gerard, P. Etienne, and D. Pribat, *Appl. Phys. Lett.* 68, 2654 (1996).
4. D. Kapolnek, S. Keller, R. Vetury, R.D. Underwood, P. Kozodoy, S.P. Denbaars, and U.K. Mishra, *Appl. Phys. Lett.* 71, 1204 (1997).
5. O.-H. Nam, M.D. Bremser, T.S. Zheleva, and R.F. Davis, *Appl. Phys. Lett.* 71, 2638 (1997).
6. Y.H. Lo, *Appl. Phys. Lett.* 59, 2311 (1991).
7. P.D. Moran, D.M. Hansen, R.J. Matyi, J.G. Cederberg, L.J. Mawst, and T.F. Kuech, *Appl. Phys. Lett.* 75, 1559 (1999).
8. T.A. Langdo, C.W. Leitz, M.T. Currie, E.A. Fitzgerald, A. Lochtefeld, and D.A. Antoniadis, *Appl. Phys. Lett.* 76, 3700 (2000).
9. S.A. Chaparro, Y. Zhang, and J. Drucker, *Appl. Phys. Lett.* 76, 3534 (2000).
10. D.K. Biegelsen, F.A. Ponce, A.J. Smith, and J.C. Tramontana, *J. Appl. Phys.* 61, 1856 (1987).
11. J.W. Lee, J.P. Salerno, R.P. Gale, and J.C.C. Fan, *Mater. Res. Soc. Symp. Proc.* 91, 33 (1987).
12. S.F. Fang, A.S. Iyer, H. Morkoc, H. Zabel, C. Choi, and N. Otsuka, *J. Appl. Phys.* 68, R31 (1990).
13. R.J. Matyi, H. Shichijo, and H.L. Tsai, *J. Vac. Sci. Technol. B* 6, 699 (1988).
14. S. Hasegawa, K. Maehashi, H. Nakashima, T. Ito, and A. Hiraki, *J. Cryst. Growth* 95, 113 (1989).
15. S.D. Hersee, J.C. Ramer, and K.J. Malloy, *MRS Bulletin* 22, 45 (1997).
16. X. Chen, S.H. Zaidi, S.R.J. Brueck, and D.J. Devine, *J. Vac. Sci. Technol. B* 14, 3339 (1996).
17. D. Zubia, S.H. Zaidi, S.R.J. Brueck, and S.D. Hersee, *Appl. Phys. Lett.* 76, 858-860 (2000).
18. T.L. Lin, L. Sadwick, K.L. Wang, Y.C. Kao, R. Hull, C.W. Nieh, D.N. Jamieson, and J.K. Liu, *Appl. Phys. Lett.* 51, 814 (1987).
19. K. Kiruma, T. Haga, and M. Miyazaki, *J. Cryst. Growth* 102, 717 (1990).

Nanoscale selective growth of GaAs by molecular beam epitaxy

Seung-Chang Lee,[a] Kevin J. Malloy,[a] and Steven. R. J. Brueck[a,b]
Center for High Technology Materials, University of New Mexico, 1313 Goddard, SE, Albuquerque, NM87106

Selective growth of GaAs on a nanoscale SiO_2-patterned GaAs(001) substrate by molecular beam epitaxy is reported. Reduction of the lateral dimensions of a SiO_2 pattern below the surface diffusion length of an incident Ga atom results in preferential migration from the SiO_2 surface to stable bonding configurations at nearby open GaAs substrate surfaces. This intrinsic selectivity is achieved under high growth temperature with low growth rate where surface migration on the SiO_2 surface is highly extended. A large-area nanoscale SiO_2 pattern is realized by interferometric lithography. A 330-nm period two-dimensional array of GaAs disks having a height of 40 nm and a diameter of about 240 nm, selectively grown on GaAs(001) substrates is presented. Variation of the nanoscale selective growth mode depending on the diameter of GaAs disks is also observed.

Journal of Applied Physics (to be published)

I. Introduction

Selective growth potentially will have a strong impact on device and circuit applications. Since Cho and Reinhart first demonstrated shadow-masked, isolated-area growth,[1] several attempts to realize selective growth in molecular beam epitaxy (MBE) have been reported.[1-9] Except for the work using mechanical masks,[1,2] these experiments rely on patterning of a substrate or of a deposited masking material and can be divided into two categories:[3-9] those dependent on process technology and those relying on material properties such as surface mobility of incident atoms and their reaction with a mask.

The first category includes direct deposition on a mask-patterned substrate with[3] or without[4] a subsequent lift-off step or a self-masking substrate taking an advantage of topological shadowing in MBE.[5] The second category utilizes variation in the physical or chemical properties of a material system consisting of a mask, an epilayer, and a substrate. An example of the latter category is providing several facets on a substrate upon which an incident atom has different sticking coefficient and surface mobility. These approaches are not perfectly selective but lead to a lateral thickness variation. Several facet-dependent selective growths on planar and nonplanar substrates have been reported.[6-8] Selective group IV MBE using chemical reactions between an incident Si or Ge atom and a SiO_2 mask to produce SiO or GeO volatilized from the mask can be included in this category.[9]

A new selective growth mechanism basically different from those previously mentioned is available for a mask-patterned substrate if the lateral dimension of the mask is reduced to less than a surface diffusion length of the incident atom. Atoms incident on the mask material will then migrate to nearby unmasked substrate regions where favorable bonding sites are available. That is, reduction of a mask dimension below the surface diffusion length on the mask material can result in a nanoscale selective growth (NSG) mode.

SiO_2 has been widely used as a mask material, but the surface diffusion length, $L_{D,SiO2}$, of an Ga atom incident on SiO_2 surface has not been well studied. The surface diffusion length on a GaAs surface, $L_{D,GaAs}$ critically depends on the growth conditions, ranging from 20 - to 30-nm up to a few μm.[4,7,10,11] The surface diffusion length of a Ga atom is different for epitaxial GaAs-on-GaAs and non-epitaxial GaAs-on-SiO_2. Comparing the adsorption mechanisms for these two cases, we speculate that a Ga atom will have a higher surface mobility but a shorter surface lifetime on a SiO_2 than on a GaAs surface. If nanoscale lithography is applied to SiO_2 patterning, $L_{D,SiO2}$ can be experimentally characterized and the possibility of NSG also can be examined. Recently, there has been remarkable progress in optical and electron-beam lithographies. Among them, interferometric lithography (IL) is very promising as a low-cost, reliable, and scalable technology for deep sub-100 nm scales.[12,13] In this work, we report the first demonstration of NSG of GaAs on a SiO_2-patterned substrate prepared by IL. We present a 330 nm-period 2-dimensional (2-D) array of GaAs disks with a height of 40 nm and diameters of about 240 nm, selectively grown on a SiO_2-patterned GaAs(001) substrate by MBE. We also report the dependence of NSG on the diameter of the GaAs disks.

---

a. Also Electrical and Computer Engineering Department.

b. Also Physics and Astronomy Department.

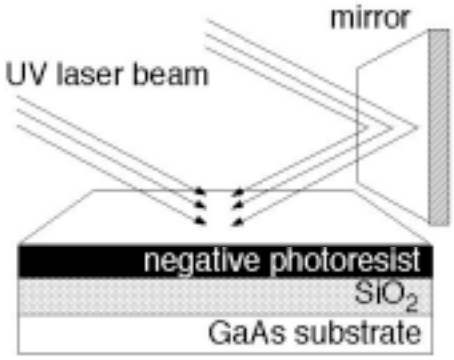


Figure 1 A schematic diagram of interferometric lithography.

II. Nanoscale pattern processing and MBE growth

As shown in Fig. 1, the basic structure for processing consists of a negative photoresist (PR) film, a 40 nm-thick SiO_2 film, and a GaAs(001) substrate. A single longitudinal and transverse mode, frequency tripled YAG laser at a 355-nm wavelength was used as an ultraviolet (UV) light source. In IL, as illustrated in Fig. 1, a direct and a mirror-reflected UV beam are used to produce a 330-nm period interference pattern on the PR film. To make a 2-D array of holes, a double exposure with a right angle rotation of the sample was performed. Fig. 2(a) and (b) are a 45°-tilted and a cross-sectional scanning electron microscope (SEM) image of the 2-D array of holes generated on the PR film. In Fig. 2(b), standing wave effects produce a sinusoidal profile along the vertical direction in the PR film.

After development, the sample was dry-etched to transfer the PR pattern of Fig. 2 (a) through the SiO_2 film. Figure 3 (a) shows a 45°-tilted SEM image of a 2-D array of holes transferred onto the SiO_2 film. These holes provide access for Ga atoms to the substrate surface for epitaxial growth. In this work, our concern is the lateral dimensions of the SiO_2 mask. While the diameter of the holes, d, ranges from 200 - 280 nm, the lateral width, L_M, between the holes in the SiO_2 mask ranges from ~ 90 nm at a minimum to a ~ 230 nm at a maximum as indicated by white arrows in Fig. 3 (a). These values are within the range reported for $L_{D,GaAs}$.[4,7,10,11]

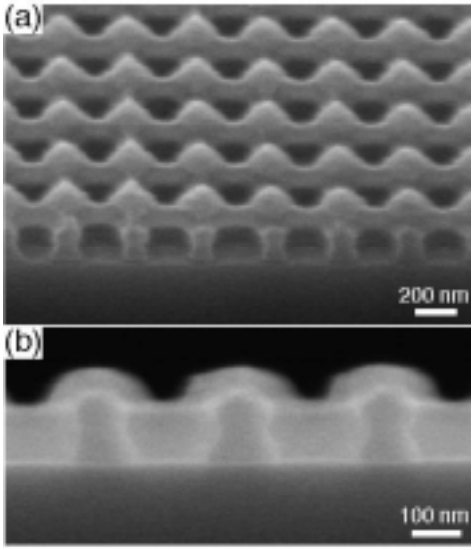


Figure 2(a) A 45°tilted and (b) a cross-sectional SEM image of a 2-D array of holes generated in a negative PR film.

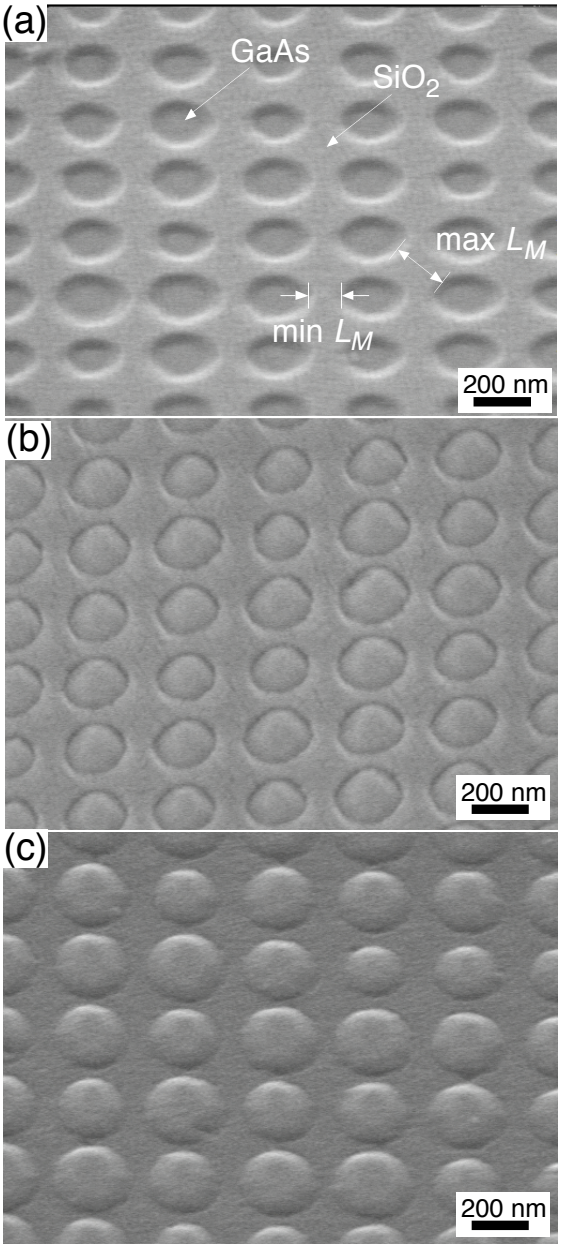


Figure 3 45°-tilted SEM images of (a) a 2-D array of holes transferred through the SiO_2 films by dry etching, (b) the MBE as-grown sample, and (c) a 2-D array of GaAs disks obtained by diluted HF treatment for removal of the SiO_2 pattern. In (a), white arrows indicate the maximum and the minimum L_M

On the SiO_2 pattern of Fig. 3 (a), MBE growth corresponding to a 30-nm thick homogeneous GaAs deposition was performed. In this work, to enhance the surface diffusion length of Ga atoms, low growth rate at high growth temperature was employed. The flux of Ga atoms was controlled to achieve a growth rate of 0.25 monolayer (ML) s^{-1}, which was separately calibrated by reflection high energy electron diffraction (RHEED) oscillation. The growth temperature was set to 630°C. This is very close to, but does not exceed,

the Ga desorption temperature.[14] The growth temperature was monitored by optical pyrometry. Although about 50% of the surface of the patterned substrate was covered by SiO_2, the thermocouple and the pyrometer showed consistent temperature control compared with a bare GaAs substrate. This means that the real temperatures were close to 630°C. In consideration of this high growth temperature of GaAs deposition, the As/Ga flux ratio was adjusted to about 40 in beam equivalent pressure.

III. Results and discussion

Figure 3(b) shows a 45°-tilted SEM image of the as-grown sample. This figure provides important experimental evidence of selective growth. First of all, from the comparison of Fig. 3(b) with Fig. 3(a), we find a change of topography. This is because the holes were filled by GaAs whereas the SiO_2 regions do not show any apparent deposition. If the deposition had occurred unselectively over the whole sample, the topography of the SiO_2 pattern of Fig. 3(a) would have been maintained during growth. No ring patterns appeared in the *in situ* RHEED, implying no formation of polycrystalline GaAs on the SiO_2 surface. These observations show that selective growth of GaAs was epitaxially achieved only inside the 2-D array of holes shown in Fig. 3(a) of which the bottoms are GaAs(001) surfaces.

Dipping the as-grown sample in a diluted HF solution strips the SiO_2 mask leaving only the selectively deposited GaAs disks on the substrate. Figure 3(c) shows a 45°-tilted SEM image of 2-D array of GaAs disks. Figure 4 presents a top-SEM image of a single GaAs disk. In Fig. 4, the GaAs disk has a circular shape without any apparent facets in the growth plane. Side-view SEM shown in Fig. 5, however, reveals several different cross sections such as a circular section [(a)], a triangle [(b)], and a trapezoid [(c) and (d)]. As is discussed in detail below, this variability reflects changes in the hole diameter resulting from pattern nonuniformity. Among these, as shown in Fig. 5(c) which has a bottom diameter similar to that of Fig. 4, a trapezoidal profile with a rounded top edge is the most frequently observed. The disk height of Fig. 5(c) is approximately 40 - 45 nm, almost 30 - 50% greater than the 30-nm thickness predicted by the deposition time and the Ga flux. This trapezoidal side-view implies that Ga atoms incident near the boundary between

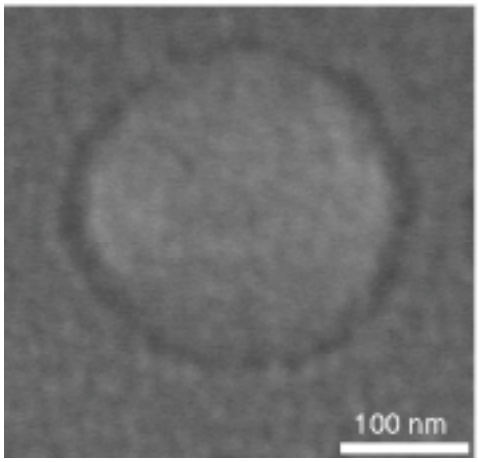


Figure 4 A top-view SEM image of a single GaAs disk of $d \sim$ 240 nm.

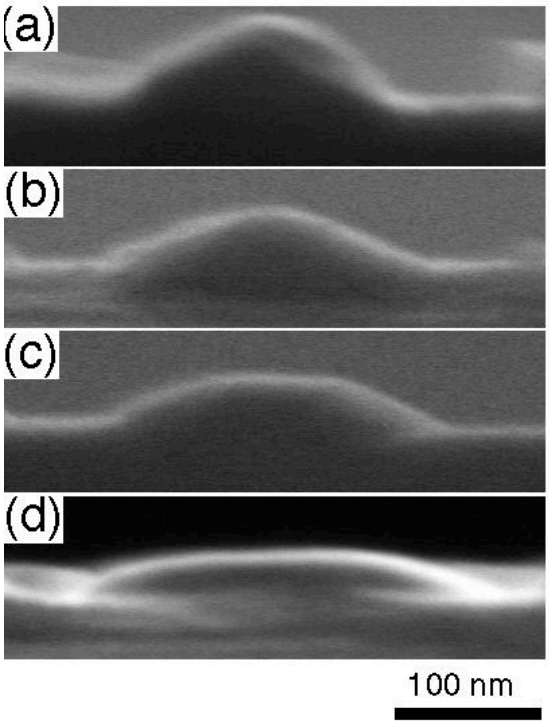


Figure 5 Side-view SEM images of GaAs disks having different diameters of (a) ~ 180 nm, (b) and (c) ~ 240 nm, and (d) ~ 330 nm.

GaAs and SiO_2 surfaces migrate to a (001) plane at the top surface of the disk to find a more stable adsorption site as shown in Fig. 5(c).[6] In Fig. 4, a dark ring with a width of about 10-nm appears on the periphery of the disk. This wasn't observed in the side-view SEM image of Fig. 5(c) but can be seen as a shadow of a shallow groove around each disk in Fig. 3(c). The origin of these rings is not clear but they may be due to some loss of As atoms from the substrate surface very close to the edge of the SiO_2 mask. Such a region is not exposed to an As-rich environment and can lose As atoms at high growth temperature.

We also investigated the behavior of Ga atoms on a unpatterned SiO_2 surface during growth. This is very important because the sticking coefficient of a Ga atom on a SiO_2 surface at the given growth condition must be examined. Figure 6 shows a 45°-tilted SEM image of a unpatterned SiO_2 surface for the same growth conditions as the SiO_2-patterned substrate. Rather than a uniform deposition, isolated islands of GaAs are formed and are randomly distributed on the SiO_2 surface.

However, it should be noted that Fig. 6 gives not a real but an *apparent* sticking coefficient, $S_{Ga\text{-}SiO2}$, of a Ga atom on a SiO_2 surface. Here, we define $S_{Ga\text{-}SiO2}$ at a *given* growth condition as a ratio of the area covered by GaAs in a unit area of SiO_2 to the unit area. The island generation observed in Fig. 6 provides seed formation for polycrystalline GaAs growth. The growth of GaAs on the unpatterned SiO_2 surface is quasi-isotropic. In other words, once a seed atom finds a favorable site on a SiO_2 surface, growth proceeds not only vertically but also laterally. Lateral growth, however, is not entirely due to adsorption of Ga atoms on a SiO_2 surface but partly due to preferred bonding between itinerant Ga atoms and Ga or As atoms already captured at the seed

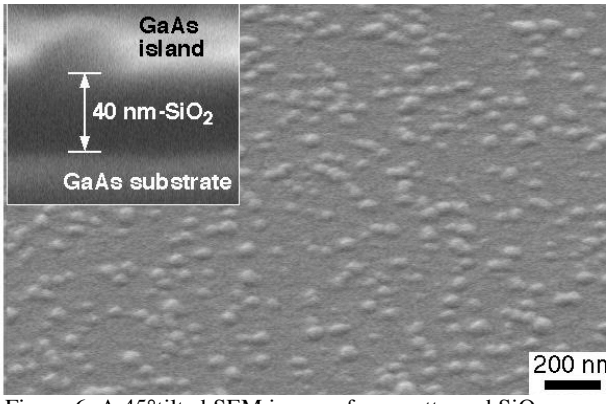


Figure 6 A 45°tilted SEM image of a unpatterned SiO_2 surface treated with same deposition under the same growth condition (see the text). The inset is a side-view SEM image of a GaAs island formed on the unpatterned SiO_2 mask.

site. Because of this, if growth is continued further, the SiO_2 surface would eventually be covered by GaAs and $S_{Ga\text{-}SiO2}$ would seem to be close to unity regardless of the initial sticking coefficient. That is, the apparent $S_{Ga\text{-}SiO2}$ depends on deposition time as well as temperature and can be very different from the initial sticking coefficient. To measure the initial sticking coefficient, submonolayer deposition with extremely low incident Ga flux is required so that an interatomic spacing of Ga atoms on SiO_2 surface is greater than $L_{D\text{-}SiO2}$ to avoid an encounter between a newly incident atom and pre-existing ones. In this non-epitaxial deposition, however, apparent sticking coefficient is more meaningful since experimental extraction of initial sticking coefficient is very difficult.

In the inset SEM image of Fig. 6, the height of the GaAs island is about 30 nm which is very close to the thickness expected from deposition time and Ga flux. In Fig. 6, however, distribution of GaAs islands is sparse, implying a low $S_{Ga\text{-}SiO2}$ for 30 nm-thick deposition. The island density is about 100 ± 20 μm^{-2} with an average spacing between GaAs islands of about 100 nm. Since the diameter of an island, shown in the inset to Fig. 5, is about 50 nm, $S_{Ga\text{-}SiO2} \sim 0.19$. By multiplying $S_{Ga\text{-}SiO2}$ by 0.67 from the consideration of the ellipsoidal shape of an GaAs island [the volume of a GaAs island $= 2\pi a^2 b/3$ where $a \sim 25$ nm and $b \sim 30$ nm], the fraction of incident Ga atoms which stick on the unpatterned SiO_2 surface for 30 nm-thick deposition is about 0.13. This implies a significant loss of incident Ga atoms. In spite of the low $S_{Ga\text{-}SiO2}$, approximately 13% of the incident Ga atoms still remain on the unpatterned SiO_2 surface to form GaAs islands. Comparing Fig. 3(b) with Fig 6 shows a very important direct evidence confirming NSG. In Fig. 3(b), no GaAs islands appear on the patterned SiO_2 mask. According to our definition, the apparent $S_{Ga\text{-}SiO2}$ on a nanoscale patterned SiO_2 surface is zero for these growth conditions. Seed formation similar to Fig. 6 also could occur on the patterned SiO_2 region. But, the nanoscale SiO_2 pattern having lateral dimensions less than $L_{D,SiO2}$ suppresses GaAs seed formation by providing Ga atoms a chance to migrate and bond to Ga or As atoms on a nearby GaAs surface. Generally, a low sticking coefficient implies a short surface lifetime. As previously mentioned, however, incident Ga atoms can have a high surface mobility on the SiO_2 surface and some of them travel a distance long enough to reach the nearest GaAs surface before desorption. Therefore, this comparison confirms that the selective growth originated from the migration of Ga atoms across nanoscale SiO_2 patterns. Since the direction of movement of an incident Ga atom on a SiO_2 surface is random, $L_{D,SiO2}$ is comparable to or greater than the maximum L_M for our growth conditions.

Although some of Ga atoms incident on the patterned SiO_2 region move to adjacent GaAs surfaces, the volume of deposited GaAs indicates that not all Ga atoms incident on the SiO_2 surface are captured. As a rough calculation, the height of a GaAs disk, h, of Fig. 5(c) is at most 40 - 50 nm which is not too different from the volume of a 30-nm thick cylindrical disk with the same bottom area expected from homogeneous unpatterned growth. This can be explained by a low $S_{Ga\text{-}SiO2}$. In addition, the ratio of the bottom area of a GaAs disk shown in Fig. 4 and Fig. 5(c) to a single period area, γ, is only about 0.42. By controlling the dry etch time, however, d can be varied from 180 nm to 330 nm or more so that γ can change from 0.23 to 0.79 or more. Here, d greater than the period is obtained by SiO_2-overetching leading to complete removal of a SiO_2 film between nearest neighbor holes. igure 7 shows 45°-tilted SEM images obtained from as-grown samples having different hole diameters but the same growth conditions. Figure 5(a) shows a side-view SEM image of the as-grown sample taken after HF treatment which has a diameter similar to those of the GaAs disks in Fig. 7 (a). Reduction of d means a decrease of γ and, as a result, an increased contribution of Ga atoms migrating from the SiO_2 surface to the volume of the GaAs

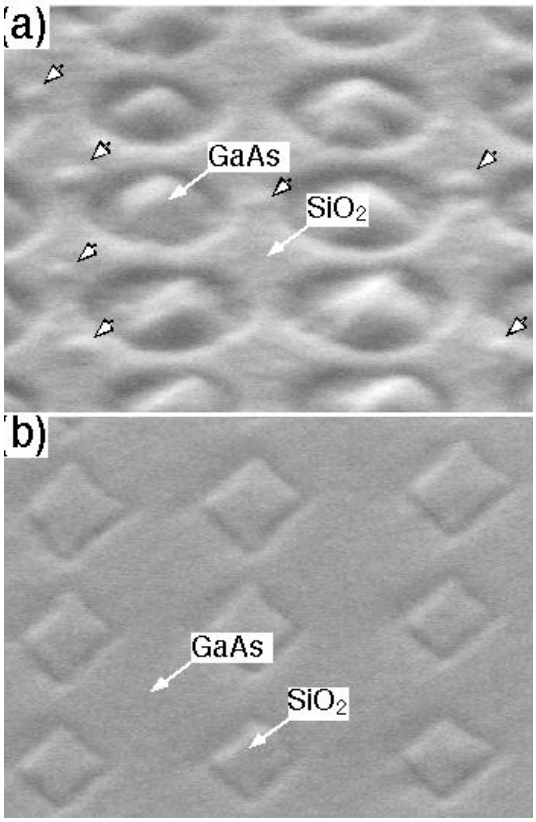


Figure 7 45°tilted SEM images obtained from the as-grown samples having d of (a) about 180 nm and (b) more than 330 nm.

disk. For $d \sim 180$ nm shown in Fig. 7 (a) and Fig. 5 (a), γ approaches 0.23. As d is decreased, the dominant profile in side view SEM varies from a trapezoid to a circular section as shown in Fig. 5. However, the overall shape of GaAs disks in Fig. 7 (a) is not spherical. The change of shape seems to be result from distortions of the holes in the SiO_2 film which generally occurs in underetching for small d. This could be induced by different etch rate in the center region which is faster than that near the periphery of a hole, resulting in U-shaped profile of the SiO_2 film of which the thickness close to the opening is too thin to be controlled properly. These distortions can cause irregular facet generation along the periphery or may prevent GaAs disks from achieving a (001) top-surface facet. In Fig. 7 (b), the SiO_2 film was overetched and mask regions between the nearest neighbor holes disappeared. Therefore, the remaining SiO_2 mask forms a 2-D array of approximately square posts rather than a 2-D array of holes as seen in Fig. 3(b) and Fig. 7(a).

Figure 8 shows the observed relation between d and h for GaAs disks measured from side-view SEM. The horizontal dotted line represents the nominal thickness estimated from Ga flux and deposition time. As shown in Fig. 5, h increases with decreasing d. When $d \sim 180$ nm or $\gamma \sim 0.23$, h is about 65 nm. This represents a relative increase of volume of a GaAs disk with decreasing γ, implying an additional contribution of Ga atoms migrating from SiO_2 regions to the volume of the GaAs disk. In a rough calculation from Fig. 8, about a 40 % increase in the volume of a GaAs disk was observed when $\gamma \sim 0.23$ using the same ellipsoidal scheme. For a circular-sectioned GaAs disk of $h \sim 65$ nm and $d \sim 180$ nm, about 12 % of the Ga atoms incident on the patterned SiO_2 surface in an unit period area then must stick to the patterned SiO_2 surface and migrate to an open GaAs surface, to explain the 40% volume increase. The volume calculation of a GaAs disk includes crude assumptions such as perfect ellipsoidal shape, but, 12% is close to the fraction of Ga atoms that remain on the unpatterned SiO_2 surface. The relative increment of the volume of a GaAs disk with decreasing d, therefore, also supports NSG induced by preferential migration of Ga atoms.

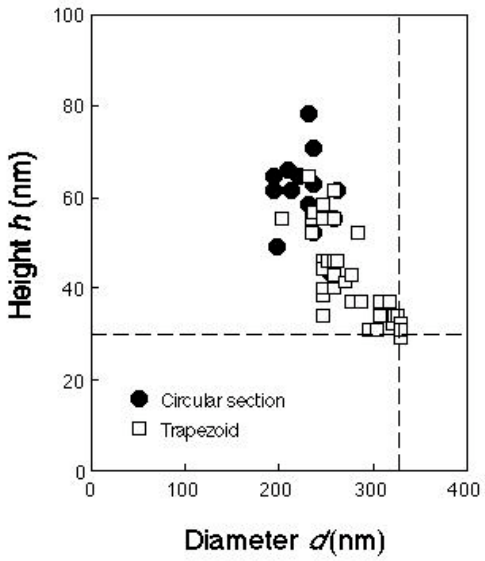


Figure 8 Relation between the diameter (*d*) and the height (*h*) of GaAs disks measured from side-view SEM. The vertical and the horizontal dotted line represent the pattern period and the nominal thickness of GaAs estimated from Ga flux and deposition time, respectively

Figure 7(a) provides the upper limit of $L_{D,SiO2}$. As indicated by white arrows, some islands similar to the GaAs islands formed on unpatterned SiO_2 surface were observed although their areal density is not enough to affect the contribution of migrating Ga atoms in the volume calculation. The formation of GaAs island on a patterned SiO_2 surface means that its L_M is no longer short enough for NSG. The low density implies initiation of island formation for the given L_M. A diameter of $d \sim 180$ nm corresponds to maximum $L_M \sim 290$ nm. At growth temperature of 630°C with growth rate of 0.25 MLs^{-1} therefore, 230 nm $\leq L_{D,SiO2} \leq$ 290 nm on a 330-nm period 2-D SiO_2-patterned GaAs(001) substrate. This value is much less than $L_{D,SiO2} \sim 4$ - 6 μm in [110] direction extrapolated for similar growth conditions which is generally less than that in [1-10] direction.[11] This small $L_{D,SiO2}$ explains the absence of selective growth mode in μm-scale SiO_2 mask patterns. On the other hand, this is more than twice the average spacing of GaAs islands on the unpatterned SiO_2 surface. Once the Ga atoms stick on and move from the patterned SiO_2 surface to open GaAs surface, they are no longer available to play a role in seeding a GaAs island for additional arriving atoms. This means that the steady state of NSG during growth consists of a continuous flow of Ga atoms from the patterned SiO_2 surface to the open GaAs surface keeping the patterned SiO_2 surface clean and without seed atoms. This explains the realization of NSG in spite of $L_{D,SiO2}$ which is greater than the average spacing between GaAs islands observed from Fig. 6.

As previously mentioned, trapezoidal, triangular, and circular-sectional shapes were observed in side-view SEM. The side-views shown in Fig. 5(a) and (c) correspond to Fig. 7(a) and Fig. 3(b), respectively. Of the various side-view shapes, the circular-sectional shape of Fig. 5(a) seems to result when imperfect dry etching gives noncircular hole patterns for small diameter holes. In case of the mid-size GaAs disks shown in Fig. 5(c), the pattern is very close to circles as seen in Fig. 4 and no particular facets are developed on the side-walls. Also, the trapezoidal shape is dominant in this case. However, a few triangular GaAs disks in side-view (or conical GaAs), as shown in Fig. 5(b), were observed among the trapezoidal ones. This implies that the migration of Ga atoms from the region near the boundary between GaAs and SiO_2 surfaces to a (001) top surface of the open GaAs region generates trapezoidal shape in side-view at the initial stage of growth and is continued until the (001) top surface disappears. If growth is continued, then most of trapezoidal GaAs disks will be converted to triangular ones. It is interesting to imagine what would happen if the growth is continued further. We deposited about 100 nm of GaAs on a SiO_2-patterned substrate similar to Fig. 3(a). It should be noted that the thickness of 100 nm is about 2.5 times greater than the thickness of the SiO_2 film. In this case, most of GaAs disks have a form of a cylindrical shape based on the hole pattern with a rounded conical top. No lateral growth was observed for this 100-nm deposition. If rectangular or stripe type patterns which can produce some particular facets during growth are employed, a nanoscale lateral growth mode for MBE possibly could be realized.

In this work, we enhanced $L_{D,SiO2}$ by using a low Ga flux with high growth temperature. These conditions, however, induce a low $S_{Ga\text{-}SiO2}$ different from the sticking coefficient of Ga atoms on a GaAs surface which is almost unity under our growth conditions.[14] Basically, the low $S_{Ga\text{-}SiO2}$ results in an incomplete selective growth mode which results in GaAs islands on the unpatterned SiO_2 surface. This also provides a favorable situation for migration of Ga atoms. By fabricating a SiO_2 mask with L_M smaller than $L_{D,SiO2}$, perfectly selective growth is achieved on a nanoscale SiO_2-patterned substrate.

NSG is radically different from previously reported selective growth techniques in several aspects. First of all, the process is very simple. Neither special wafer treatments nor sophisticated lithography technologies are required. A single SiO_2 film on a substrate patterned by IL is sufficient. Migration enhanced epitaxy may be helpful for enhanced L_{DSiO2}. Control of the interference pattern in IL will allow exploration of a wide range of patterns extending from the microscale to the nanoscale regime.

The most significant potential of NSG is in its application to heteroepitaxy. It is known that the critical thickness of a lattice-mismatched heterostructure is enhanced if the growth area is reduced.[15-18] In heteroepitaxy, an unselectively deposited homogeneous epilayer can be replaced by selectively grown 2-D array of nanoscale disks offering both strain relief and quantum confinement. In this work, we presented NSG with a hole diameter of about ~ 240 nm. But, IL and refined processing can generate patterns of sub-100-nm-diameter holes, approaching the lateral dimensions of typical self-assembled quantum dots.[19] NSG, therefore, suggests the possibility of growing various kinds of epilayers beyond their conventional critical thicknesses and of process-arranged quantum dots with well-controlled size and spatial distribution. Although there are still several problems to be overcome such as optimization of growth temperature, improvement of lithography resolution, and reliable processing, NSG presents a new frontier for strained-layer MBE technology. Presently, NSG of $In_xGa_{1-x}As$ and InAs on a GaAs substrate with sub-100-nm diameter hole patterns is underway.

IV. Summary and conclusion

Homoepitaxial nanoscale selective growth (NSG) on a SiO_2-patterned GaAs(001) substrate realized by interferometric lithography (IL) has been reported. On a SiO_2 surface for which the lateral dimension is less than a Ga atom surface diffusion length, the Ga atom either desorbs or migrates to nearby GaAs surfaces and is incorporated into a local epitaxial growth. With NSG, a 330-nm period, two-dimensional array of GaAs disks having a height of 40 nm and a diameter of about 240 nm has been successfully grown on a GaAs(001) substrate by MBE. This selectivity is achieved under the growth temperature of 630°C with the growth rate of 0.25 ML s^{-1} where surface migration length of a Ga atom on SiO_2 surface is extended up to 230 - 290 nm. Conclusively, this work demonstrates that selective growth is possible in MBE by understanding preferential migration of a Ga atom for mask dimensions less than the surface diffusion length.

Acknowledgment: The authors wish to acknowledge the support of DARPA and of ARO/MURI, and thank A. Stinz for assistance in MBE growth.

References

[1] A. Y. Cho and F. K. Reinhart, Appl. Phys. Lett. **21**, 355 (1972).
[2] W. T. Tsang and M. Illegems, Appl. Phys. Lett. **31**, 301 (1977).
[3] G. M. Metze, A. B. Cornfeld, P. E. Laux, T. C. Ho, and K. P. Pande, Appl. Phys. Lett. **57**, 2576 (1990).
[4] A. Y. Cho and W. C. Ballamy, J. Appl. Phys. **46**, 783 (1975).
[5] S. Nagata and T. Tanaka, J. Appl. Phys. **48**, 940 (1977).
[6] W. T. Tsang and A. Y. Cho, Appl. Phys. Lett. **30**, 293 (1977).
[7] E. Kapon, M. C. Tarmago, And D. M. Hwang, Appl. Phys. Lett. **50**, 347 (1987).
[8] A. Madhukar, K. C. Pakkumar, and P. Chen, Appl. Phys. Lett. **62**, 1547 (1993).
[9] S. J. Yun, S. C. Lee, B. W. Kim, and S. W. Kang, J. Vac. Sci. Technol. **B12**, 1167 (1994).
[10] S. Nilsson, E. Van Gieson, D. J. Arent, H. P. Meier, W. Walter, and T. Forster, Appl. Phys. Lett. **55**, 972 (1989).
[11] M. Hata, T. Isu, A. Watanabe, and Y. Ktayama, J. Vac. Sci. Technol. **B8**, 692 (1990).
[12] L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll, Appl. Optics, **17**, 1165 (1978).
[13] S. R. J. Brueck, S. H. Zaidi, X. Chen, and Z. Zhang, Microelec. Engineering, **41/42**, 145 (1998).
[14] R. Fischer, L. Klem, T. J. Drummond, R. E. Thorne, W. Kopp, H. Morkoc, and A. Y. Cho, J. Appl. Phys. **54**, 2508 (1983).
[15] E. A. Fizgerald, G. P. Watson, R. E. Proano, D. G. Ast, P. D. Kirchner, G. D. Pettit, and J. M. Woodall, J. Appl. Phys. **65**, 2220 (1989).
[16] A. Fischer and H. Richter, Appl. Phys. Lett. **61**, 2656 (1992).
[17] S. Luryi and E. Suhir, Appl. Phys. Lett. **49**, 140 (1986).
[18] D. Zubia, S. H. Zaidi, S. R. J. Brueck, and S. D. Hersee, Appl. Phys. Lett. **76**, 858 (2000).
[19] A. Stintz, G. T. Liu, A. L. Gray, R. Spillers, S. M. Delgado, and K. J. Malloy, J. Vac. Sci. Technol. **B18**, 1496 (2000).

Molecular-beam epitaxial growth of one-dimensional rows of InAs quantum dots on nanoscale-patterned GaAs

S. C. Lee,[a),b)] L. R. Dawson,[a)] K. J. Malloy,[a)] and S. R. J. Brueck[a),c)]
Center for High Technology Materials, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

(Received 22 June 2001; accepted for publication 7 August 2001)

Single-dot-wide rows of InAs quantum dots (QDs) aligned along a $[01\bar{1}]$ direction on a 180-nm-period nanoscale-patterned (nanopatterned) GaAs(100) substrate are reported. The nanopatterned substrate is realized by interferometric lithography along with the selective growth mode of GaAs. Orientation-dependent migration and incorporation of In atoms from (111)*A* to (100) facets on the nanopatterned substrate localizes QD formation exclusively along a 30–40-nm-wide (100) facet defined by neighboring (111)*A*-type facets within each period. These aligned QDs form face-to-face multi-QDs analogous to multi-quantum-well structures, in a one-dimensional configuration. Spatially controlled formation of QDs with an improved size uniformity on the nanopatterned substrate is presented. © *2001 American Institute of Physics.*
[DOI: 10.1063/1.1409947]

Recent progress in the field of semiconductor nanostructures shows strong promise of realizing the characteristics required for next-generation electronic and optoelectronic devices and for basic research in low-dimensional physics. In particular, self-assembled quantum dots (QDs) have been intensively studied as a result of their fabrication-free formation leading to high-quality, low-dimensional nanoscale structures.[1]

In spite of the advantages resulting from the self-assembled formation mechanism, the inherent size fluctuation and random spatial distribution are critical issues that are not yet adequately controlled. Several attempts relying on patterned growth have employed formation of steps or particular facets, to enable orientation-dependent migration and incorporation (ODMI) of In atoms.[2–9]

ODMI is mass transport during growth from one facet to another one caused by different adsorption kinetics corresponding to alternative surface bonding configurations. For molecular beam epitaxy (MBE), it is known that Ga and In atoms have a property of ODMI from a $(n11)A$-type facet to a (100) facet ($n=1$ or 3), resulting in a faster growth rate on a (100) facet than on a $(n11)A$-type facet when both are exposed to identical beam fluxes.[10–13] ODMI, then, can also affect QD formation. Although several investigations of ODMI on QD formation with a patterned substrate have been reported, the patterns reported so far have had scales beyond the lateral dimension of typical self-assembled QDs, L_{SQ}. Generally, $L_{SQ} \sim 20-30$ nm for InAs QDs on a GaAs substrate.[1] Recent developments in lithography now allow the possibility of large-area nanoscale patterns with critical dimensions approaching L_{SQ}. Especially, large-area interferometric lithography (IL) is emerging as a low-cost, reliable, and scalable technology for sub-100-nm scales.[14]

In this work, we demonstrate a nanopatterned GaAs(100) substrate fabricated with a 180-nm-period 1D stripe pattern by IL followed by selective MBE growth to form the (111)*A*-type facets. Each period of the pattern consists of a 30–40-nm-wide (100) facet along a $[01\bar{1}]$ direction sandwiched between (111)*A*-type facets. We report growth of 1D rows of InAs multi-QDs aligned on these (100) facets by ODMI. We also present photoluminescence (PL) of these InAs multi-QDs and compare it with PL of QDs randomly generated on an unpatterned substrate.

The critical steps are: (1) to prepare a GaAs substrate patterned with nanoscale (111)*A* and (100) facets; and (2) to grow InAs QDs self-aligned to the pattern by MBE. Figure 1 shows the major steps of the process. A 180-nm-period 1D pattern is generated on a 25-nm-thick SiO_2 film by IL and dry etching. The pattern was aligned along the $[01\bar{1}]$ direction of the substrate. Figure 2 presents a 45°-tilted scanning electron microscopy (SEM) image of a SiO_2-patterned substrate corresponding to Fig. 1(a). The width of each SiO_2 stripe shown in the inset of Fig. 2 is, less than 50 nm, comparable to L_{SQ}.

Formation of the nanopatterned substrate is accomplished by MBE growth of GaAs on the SiO_2-patterned substrate under selective growth conditions with a zero sticking coefficient of a Ga atom on a SiO_2 surface as evidenced by the lack of any GaAs growth on an unpatterned SiO_2 film. For certain growth conditions such as high growth temperature at extremely low growth rate, the adsorption behavior of a Ga atom on the SiO_2 surface results in a zero sticking coefficient while still keeping the sticking coefficient on the GaAs surface close to unity. If the growth temperature is above 615 °C when the growth rate is less than 0.1 monolayer (ML) s^{-1}, we found that the sticking coefficient of a Ga atom on an unpatterned SiO_2 surface becomes zero. To build up (100) and (111)*A* facets on the SiO_2-patterned substrate, selective growth, instead of chemical etching, was employed in this work. As shown in Fig. 1(b), selective growth results in the formation of an isosceles triangular cross section on an open GaAs stripe separated by SiO_2 stripes. This

[a)]Also at the Department of Electrical and Computer Engineering.
[b)]Electronic mail: sclee@chtm.unm.edu
[c)]Also at the Department of Physics and Astronomy.

 © 2001 American Institute of Physics
Downloaded 10 Oct 2001 to 64.106.37.51. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

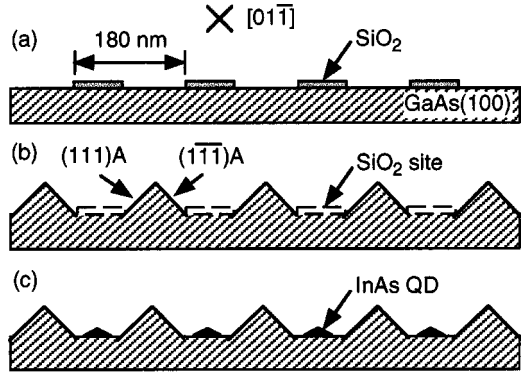


FIG. 1. Schematic diagram of the process consisting of (a) SiO_2 patterning by IL and dry etching, (b) formation of isosceles triangular-shaped GaAs by selective growth, and (c) formation of QDs self-aligned on a nanopatterned substrate prepared in (b) at cross-sectional view.

is because Ga atoms incident near the boundary between GaAs and SiO_2 surfaces move to an inner side of the open GaAs stripe to find a more stable adsorption site.[10] The resulting sidewalls developed by this movement are (111)*A*-type facets. Therefore, selective growth produces (111)*A*-type facets keeping the (100) facet protected by the SiO_2 stripe mask.

To completely form (111)*A*-type facets, deposition corresponding to 42 nm of GaAs was performed on the SiO_2-patterned substrate of Fig. 2 at the growth conditions mentioned above. Dipping the as-grown sample in HF simply removes the SiO_2 stripe mask and the cross-sectional SEM image shown in Fig. 3(b) reveals an isosceles triangular cross section as predicted. In Fig. 3(b), the widths of nanoscale (111)*A* sidewalls and (100) bottom facets are about 90–110 and 30–40 nm, respectively. These distances are much less than the reported surface migration length of an In atom under similar growth conditions and enable ODMI across the entire (111)*A*-type facets.[11,12] The 45°-tilted SEM image shown in Fig. 3(a) reveals that sidewalls along the pattern are not perfect (111)A type because of nonuniformities of the SiO_2 pattern but are sufficient to activate ODMI as seen below.

On the nanopatterned substrate shown in Fig. 3, deposition of InAs was performed at 510 °C with a nominal growth rate of 0.05 ML s^{-1} so that InAs QDs are formed only on (100) facets by ODMI as shown in Fig. 1(c). The total

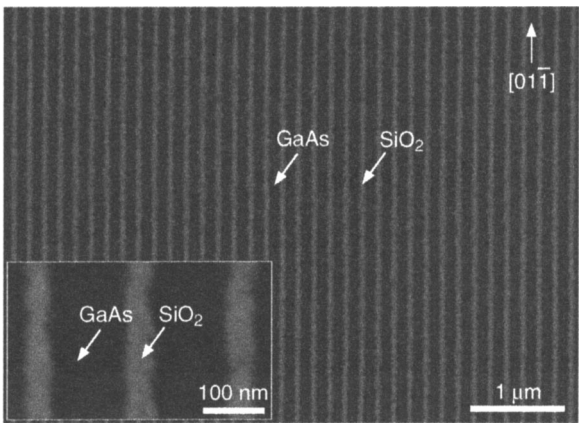


FIG. 2. 45°-tilted SEM image of a SiO_2-patterned GaAs(100) substrate. The inset shows 30–40-nm-wide SiO_2 mask stripes comparable to L_{SQ}.

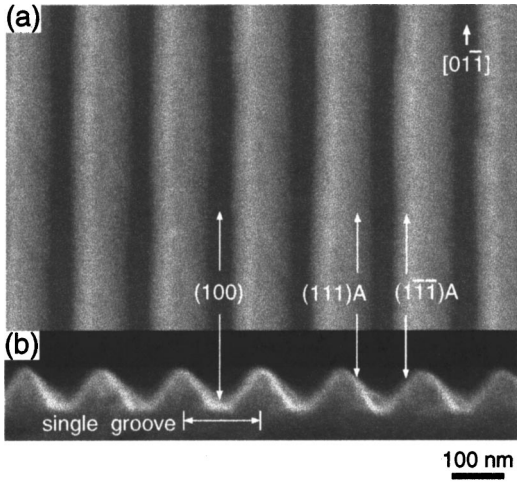


FIG. 3. (a) 45°-tilted and (b) cross-sectional SEM image of a nanopatterned GaAs(100) substrate prepared by selective growth.

amount of InAs supplied corresponds to about 1.5–2.5 ML. Figure 4 shows atomic-force microscopy (AFM) images obtained from the as-grown sample. As shown in Fig. 4(a), InAs QDs were generated but aggregated only along the bottom faces. No QDs were located on the sidewalls. This confirms that mass transport of In atoms from (111)*A* to the (100) facet due to ODMI occurs during growth.

The dot density across the entire area measured from Fig. 4(a) is about $1.2\pm0.2\times10^{10}\,\mathrm{cm}^{-2}$. This value is about 40%–50% of dot densities of random-generated QDs on a unpatterned substrate grown under the same growth conditions reported in patterned growth. ODMI effectively restricts the area available for QD formation to the (100) facet within each period. In Figs. 4(a) and 4(b), most of the QDs seem to be squeezed by adjacent sidewalls. These QDs are apparently very close to each other and build up 1D rows of multi-QDs analogous to multi-quantum-well structures. Figure 4(b) reveals that most of the aligned QDs of Fig. 4(a) have AFM-convolved lateral dimensions in the range of 30–33 nm which is highly uniform compared with typical random-generated QDs on an unpatterned substrate. In the sideviews shown in Figs. 4(c) and 4(d), the aligned QDs have a height of about 5–7 nm, almost the same as that of a typical QD grown on a unpatterned substrate.[1]

Room-temperature PL of the aligned QDs on a nanopatterned substrate and of randomly generated QDs on an unpatterned substrate, grown at the same growth conditions, are presented in Fig. 5. These PL spectra were measured with He–Ne laser excitation (633 nm) and a 77-K-cooled Ge detector. In Fig. 5, the aligned QDs on the nanopatterned substrate have a lower integrated PL intensity than the randomly generated QDs on the unpatterned substrate, supporting the lower dot density on the nanopatterned substrate previously mentioned. PL spectra from the aligned QDs and the randomly generated QDs of Fig. 5 have peak wavelengths at 1559 and 1564 nm or 0.795 and 0.793 eV, respectively. It is known that air-exposed InAs QDs are under lower compression than capped ones and have a PL peak around these values and also have a relatively large full width at half maximum (FWHM) compared with capped QDs.[15] However, it should be noted that the FWHM of the PL spectrum of QDs on a nanopatterned substrate is about 97 meV, which is

Downloaded 10 Oct 2001 to 64.106.37.51. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

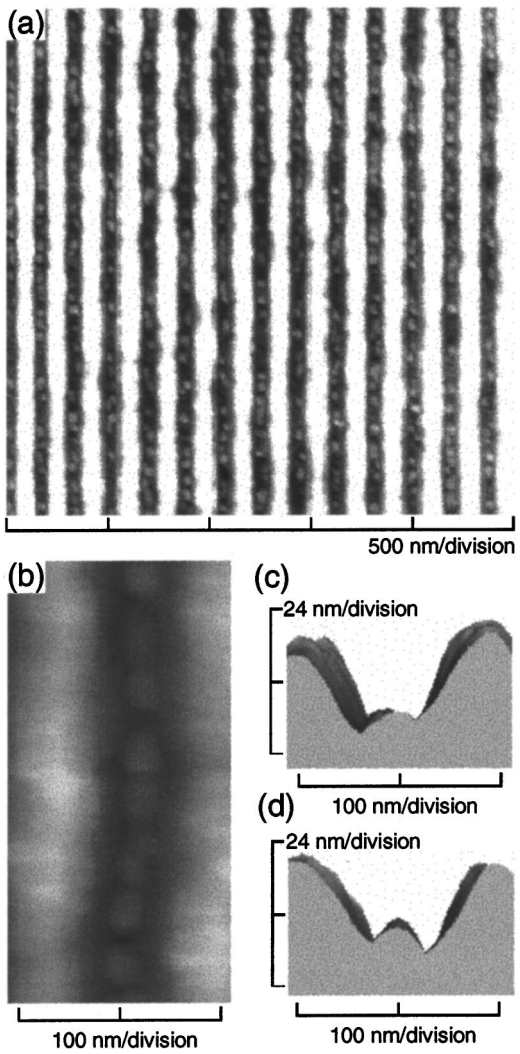


FIG. 4. AFM images of aligned QDs (a) with 2.5 μm scan range, and (b) with its 5× magnification. Sideviews of the AFM image of (b) seen from (c) the top and (d) the bottom.

26% less than the FWHM of 131 meV of the PL for randomly generated QDs on an unpatterned substrate. Reduction of the FWHM generally implies improved size uniformity and confirms the AFM result of Fig. 4.

Although most of the aligned QDs in the 1D single row of Fig. 4(a) appear adjacent, the tip-limited resolution of the AFM does not allow precise measurement of nearest-neighbor distances. It is not clear whether they are close enough to expect coupling between adjacent QDs.

The reduced FWHM of Fig. 5 suggests an advantage for aligned QDs grown on a nanopatterned substrate in device applications lowering the threshold for laser diodes. Figures 4 and 5, therefore, show important experimental results that ODMI spatially forces QD formation to be aligned along 30–40-nm-wide (100) facets and that those aligned QDs form 1D multi-QDs on a nanopatterned substrate with improved size uniformity.

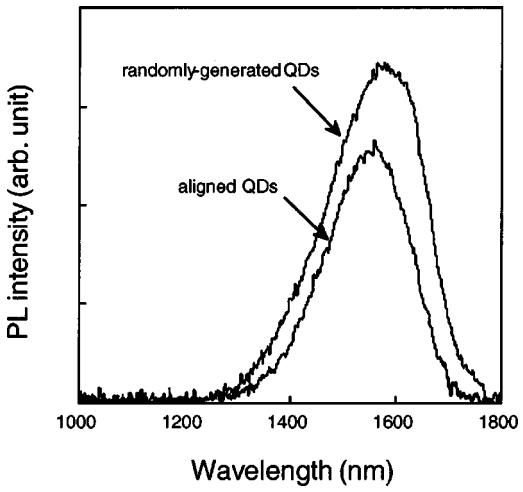


FIG. 5. Room-temperature PL spectra of the aligned QDs on a nanopatterned substrate and the randomly generated QDs on an unpatterned substrate.

In this work, we have reported 1D single-dot-wide rows of InAs QDs aligned along the $[01\bar{1}]$ direction on a 180-nm-period nanopatterned GaAs(001) substrate prepared by IL and selective MBE. QD formation forced by ODMI is exclusively observed along a 30–40-nm-wide (100) facet defined by neighboring (111)*A*-type facets within each period. These aligned QDs form face-to-face multi-QDs in a 1D configuration suggesting the possibility of dot-to-dot coupling. They also show improved size uniformity in AFM, which is confirmed by a 26% reduced FWHM of the PL spectrum compared with that of randomly generated QDs.

The authors wish to acknowledge the support of DARPA and of ARO/MURI.

[1] For review, see *Self-Assembled InGaAs/GaAs Quantum Dots*, edited by M. Sugawara (Academic, San Diego, CA, 1999).
[2] D. S. Mui, D. Leonard, L. A. Coldren, and P. M. Petroff, Appl. Phys. Lett. **66**, 1620 (1995).
[3] Y. Sugiyama, Y. Sakuma, S. Muto, and N. Yokoyama, Appl. Phys. Lett. **67**, 256 (1995).
[4] W. Seifert, N. Carlsson, A. Petersson, L.-E. Wernersson, and L. Samuelson, Appl. Phys. Lett. **68**, 1684 (1996).
[5] M. Araki, Y. Hanada, H. Fuhikura, and L. Hasegawa, Jpn. J. Appl. Phys., Part 1 **36**, 1763 (1997).
[6] S. Jeppesen, M. S. Miller, D. Hessman, B. Kowalski, I. Maximov, and L. Samuelson, Appl. Phys. Lett. **68**, 2228 (1996).
[7] A. Konkar, A. Madhukar, and P. Chen, Appl. Phys. Lett. **72**, 220 (1998).
[8] T. Ishikawa, T. Nishimura, S. Kohmoto, and K. Asakawa, Appl. Phys. Lett. **76**, 167 (2000).
[9] H. Lee, J. A. Johnson, M. Y. He, J. S. Speck, and P. M. Petroff, Appl. Phys. Lett. **78**, 105 (2001).
[10] W. T. Tsang and A. Y. Cho, Appl. Phys. Lett. **30**, 293 (1977).
[11] S. Nilsson, E. V. Gieson, D. J. Arent, H. H. Meier, W. Walter, and T. Foster, Appl. Phys. Lett. **55**, 972 (1989).
[12] D. J. Arent, S. Nilsson, Y. D. Galeuchet, H. P. Meier, and W. Walter, Appl. Phys. Lett. **55**, 2611 (1989).
[13] M. Hata, T. Isu, A. Watanabe, and Y. Katayama, J. Vac. Sci. Technol. B **8**, 692 (1990).
[14] S. R. J. Brueck, S. H. Zaidi, X. Chen, and Z. Zhang, Microelectron. Eng. **41/42**, 145 (1998).
[15] H. Saito, K. Nishi, and S. Sugou, Appl. Phys. Lett. **73**, 2742 (1998).

Downloaded 10 Oct 2001 to 64.106.37.51. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

Introduction to the Feature Section on Growth of Heterostructure Materials on Nanoscale Substrates

HETEROSTRUCTURED materials are important—and now often indispensable—enablers of a wide range of electronic and optoelectronic semiconductor devices. Heterojunctions play a dominant role in quantum effect devices by providing carrier and photon confinement, enabling tunneling transport, and providing a broad gain spectrum.

The papers in this Feature Section provide a snapshot of some new approaches to heteroepitaxy that exploit micro- and nanoscale features, created by both "bottom-up" self-assembly and "top-down" lithographic techniques. These approaches promise further significant device improvements by relaxing the traditional restrictions imposed by defect formation in non-lattice-matched material systems and they will eventually allow highly mismatched semiconductors to be integrated monolithically.

Drucker discusses the growth of self-assembled Ge:Si quantum dots, investigating both the morphological and optical/electronic properties as a function of growth parameters. Ichikawa investigates ~10-nm nanostructures, in the same material system, that are defined using ultrathin SiO_2 technology. Buttard *et al*. also investigate the Ge:Si system using molecular hydrophobic bonding of ultra-thin Si layers to provide self-assembled nanoscale dislocation networks, confined to the interfacial region, leading to a periodic strain field that extends to the growth interface and controls the nucleation of the Ge quantum dots. Roskowski *et al*. extend the understanding of pendeo-epitaxy for the GaN:SiC system to achieve marked reductions in the dislocation density in this technologically important system. Hersee *et al*. investigate the use of nanoheteroepitaxy—growth on lithographically defined nanoscale seeds, taking advantage of both strain partitioning between the growing film and the nanoscale substrate and the local presence of a free surface at the nanoscale boundaries, to modify defect formation in the GaAs:Si and GaN:Si systems.

These approaches are evolving and maturing. As they move from materials science studies to device structures, they promise to improve the performance of existing devices and to offer new capabilities and new functionality in electronics and optoelectronics.

STEVEN R. J. BRUECK, *Guest Editor*
University of New Mexico
Center for High Technology Materials
Albuquerque, NM 87106 USA

STEPHEN D. HERSEE, *Guest Editor*
University of New Mexico
Center for High Technology Materials
Albuquerque, NM 87106 USA

DAVID ZUBIA, *Guest Editor*
University of Texas
Electrical Engineering Department
El Paso, TX 79968 USA

Publisher Item Identifier 10.1109/JQE.2002.801646.



Steven R. J. Brueck (S'63–M'71–SM'89–F'93) was born in New York City in 1944. He received the B.S. degree from Columbia University, NY, in 1965 and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1967 and 1971, respectively, all in electrical engineering.

From 1971 to 1985, he was a Member of the Technical Staff at MIT Lincoln Laboratory. In 1985, he moved to the University of New Mexico at Albuquerque, where he is currently a Professor in the Electrical and Computer Engineering Department and the Physics and Astronomy Deparments, and Director of the Center for High Technology Materials. His current research interests include nanoscale lithography, the physics of nanostructures, the nonlinear optics of poled glasses, and semiconductor laser physics.

0018-9197/02$17.00 © 2002 IEEE



Stephen D. Hersee (SM'96) was born in Hastings, U.K., in 1950. He received the B.Sc. degree (with first-class honors) in 1968 and the Ph.D. degree in 1975, both from Brighton Polytechnic, Brighton, U.K.

He was with The Plessey Co. Ltd. (U.K.) from 1975 to 1980, Thomson CSF (France) from 1980 to 1986, and General Electric (Syracuse, NY) from 1986 to 1991. He joined the University of New Mexico at Albuquerque in 1991, where he is currently a Professor in the Electrical and Computer Engineering Department. He teaches courses in semiconductor materials, advanced heterojunction devices, and semiconductor process technology. He is also a member of the Center for High Technology Materials. His research interests include novel semiconductor materials and heteroepitaxy for advanced materials and devices.



David Zubia was born in Delicias, Chihuahua, Mexico, in 1964. He received the Bachelor's and Master's degrees from the University of Texas at El Paso in 1989 and 1993, respectively, and the Ph.D. degree in electrical engineering from the University of New Mexico (UNM) at Albuquerque in 2000. His thesis work involved the heteroepitaxy of lattice-mismatched materials on nanopatterned substrates and led the development of nanoheteroepitaxy.

In 1993, he joined Golden Photon, Inc (GPI), Golden, CO, where his research involved investigation of the degradation mechanisms in CdS/CdTe solar cells. In 2000, he joined the University of New Mexico as Research Assistant Professor and Manager of the Crystal Growth Facility at the Center for High Technology Materials. His research included the utilization of nanopatterned substrates for integration of compound semiconductors with silicon, the development of radiation hard semiconductors, and GaInNAs vertical-cavity surface-emitting lasers. He taught classes in silicon VLSI technology while at UNM. In 2001, he joined the University of Texas at El Paso as an Assistant Professor, where his current research interests include studying the nanoscale aspects of materials and developing nanotechnology to create novel semiconductor materials and devices. This includes scientific studies of nucleation and coalescence of semiconductor nanocrystals and strain relief in nanoporous substrates. His technological interests include the development of small-, direct-bandgap semiconductor alloys, that are lattice and thermodynamically matched to silicon VLSI, the development of advanced materials utilized for energy cells, and the creation of novel nanodevices.

Dr. Zubia is a Guest Editor for the IEEE JOURNAL OF QUANTUM ELECTRONICS. He is a Ford Foundation Fellow, Chair of the Third International Conference on Alternative Substrate Technology, and a member of the American Association for Crystal Growth.

Nanoheteroepitaxy for the Integration of Highly Mismatched Semiconductor Materials

Stephen D. Hersee, *Senior Member, IEEE*, David Zubia, Xinyu Sun, R. Bommena, Mike Fairchild, S. Zhang, David Burckel, A. Frauenglass, and S. R. J. Brueck, *Fellow, IEEE*

Invited Paper

***Abstract*—We describe an ongoing study of nanoheteroepitaxy (NHE), the use of nanoscale growth-initiation areas for the integration of highly mismatched semiconductor materials. The concept and theory of NHE is briefly described and is followed by a discussion of the design and fabrication by interferometric lithography of practical sample structures that satisfy the requirements of NHE. Results of NHE growth of GaAs-on-Si and GaN-on-Si are described, following the NHE process from nucleation through to coalescence. Micro-Raman measurements indicate that the strain in partially coalesced NHE GaN-on-Si films is $<$0.1 GPa.**

***Index Terms*—Coalescence, GaN, MOCVD, nanoheteroepitaxy, Raman, selective growth.**

I. Introduction

EPITAXIAL crystal growth has developed from a paradigm of large-area film growth, where the formation and propagation of defects in the growing film has restricted the materials used to nearly lattice matched systems. There has been a recent increase in interest in selective-area and lateral overgrowth on patterned substrates, [1]–[4] because of the potential for fabricating low-defect density heterostructures [5], [6]. Of particular interest are materials such as GaN, for which lattice matched substrates are not available. The integration of such materials [7] has a huge potential payoff, for example, in increasing the functionality of Si-based integrated circuits through the integration of III–V based optical and electronic device technologies. Selective-area growth also offers new solutions to difficult processing challenges, such as contacting the base region of Al/GaN/GaN heterojunction bipolar transistors [8].

Although most research on patterned growth has concentrated on patterning on the $\gtrsim$1-μm scale, recent theoretical [9], [10] and experimental results [11] indicate that novel strain relief mechanisms are available when the substrate is patterned at the nanoscale. These mechanisms, active when the scale of the nanostructure approaches the critical thickness, offer new approaches for dealing with the stress associated with lattice and thermal mismatches [12]. Nanometer-scale structures are also of interest in their own right. Quantum-dot structures are being investigated particularly as gain media for semiconductor lasers [13], and nanoscale patterning [14] provides a potential route to an ordered, monosized distribution of dots that will have important device implications.

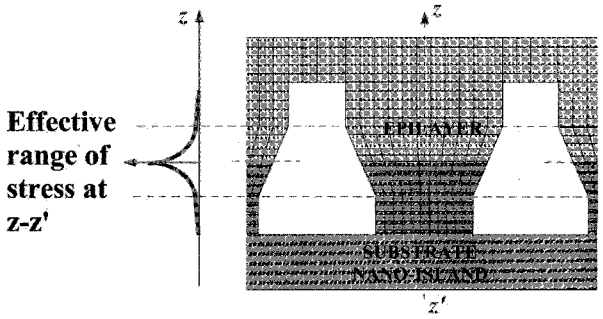


Fig. 1. Nanoheteroepitaxy concept. Epilayer nucleates selectively onto nanoscale substrate islands and coalescence occurs once strain free material is obtained.

This paper discusses an ongoing study of nanoheteroepitaxy (NHE) [10], which exploits the three dimensional strain relief mechanisms that are available to nanoscale nuclei, in a practical and scalable technology. Crystal growth on patterned substrates [15] is quite different from that observed on planar surfaces [16] and much remains to be understood. In patterned substrates, three-dimensional (3-D) nuclei are spatially ordered with a fixed density as opposed to the random distribution, in both nucleus size and nearest neighbor distance, that is characteristic of self assembled planar samples. Faceting of the nuclei is mainly determined by surface energy and not interfacial energy, and coarsening due to Ostwald ripening is less frequently observed in ordered arrays. Coalescence of the nuclei, which is necessary to obtain useful macroscopic areas of epilayer, is critically dependent on the initial faceting behavior. In NHE, coalescence occurs late in the growth stage, after the nuclei have grown sufficiently large to contact neighboring islands, rather than by the physical movement of small islands [17]. We show later that even when there is a large lattice mismatch an excellent coalescence can be achieved. However, if the structure is nonuniform competing facets can cause growth instabilities and lead to rough morphologies [1], [18].

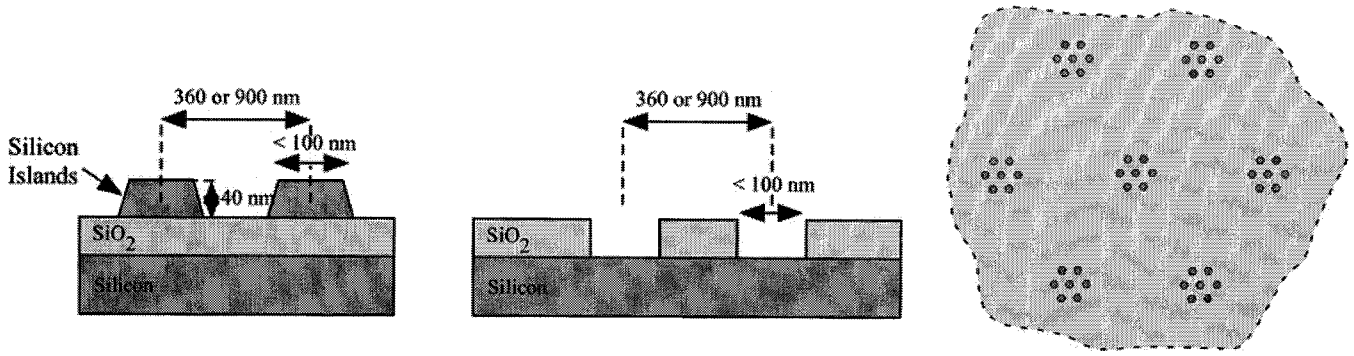
Manuscript received October 29, 2001; revised February 28, 2002. This work was partly supported by AFOSR, ARO/MURI, and DARPA.

S. D. Hersee, X. Sun, R. Bommena, M. Fairchild, S. Zhang, D. Burckel, A. Frauenglass, and S. R. J. Brueck are with the Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM 87106 USA.

D. Zubia is with the Electrical Engineering Department, University of Texas, El Paso, TX 79968 USA.

Publisher Item Identifier 10.1109/JQE.2002.800987.

0018-9197/02$17.00 © 2002 IEEE



(a) (b) (c)

Fig. 2. (a) SOI-based NHE sample. (b) Patterned oxide NHE substrate. (c) Plan view of grouped nanoscale pattern, where within the group each nucleus is at nanoscale and during growth there is coalescence between members of the same group. However, the large group-to-group spacing means that neighboring groups do not coalesce.

A facile, large-area nanopatterning technique is a requirement of NHE. Interferometric lithography [19] provides large areas of nanostructured material (scales as small as 10 nm and densities to 10^{11} cm^{-2} over areas of >10 cm^2) with a large flexibility in the pattern characteristics (size, density, array, and nanoelement symmetry, area, etc.). The current state-of-the-art in nanopatterning is discussed along with projections for developing smaller features and denser arrays for NHE substrates.

II. NHE

NHE is a novel approach [10] to the heteroepitaxy of highly mismatched materials. NHE uses selective epitaxy to nucleate nanosize epitaxial islands onto a substrate that is patterned with nanoscale features (Fig. 1). The growth is selective and the small size of the growth nucleus allows both the substrate and epilayer to deform in three dimensions in response to lattice mismatch stress. The stress and strain associated with lattice mismatch at the heterojunction decays exponentially away from the heterointerface [9] with a characteristic length on the order of the diameter of the nanoscale nucleus. The strain energy associated with the growing film saturates with film thickness as the film reaches its unstressed lattice parameters; if this saturation value can be kept below the energy required for a defect to form, strain-free and defect-free material can be grown. In addition, the local availability of a free surface at the edges of the nanosize islands modifies the defect formation and propagation kinetics and can, in some cases, confine the defects to a local region at the initial interface in the epitaxial structure [11].

III. NHE Substrate Preparation

Fig. 2(a) shows the sample structure used in our initial NHE experiments, based on an silicon-on-insulator (SOI) substrate where the top silicon layer is thinned to <100 nm and then patterned using interferometric lithography (described later). During MOCVD growth of GaAs or GaN films, the SiO_2 in the SOI provides excellent growth selectivity, allowing growth only on the Si islands. For this structure, the substrate islands are patterned at nanoscale and NHE theory predicts that a significant portion (approximately half) of the mismatch strain energy will be taken up by the substrate [10]. While this SOI-based NHE sample structure represents an ideal substrate in terms of the practical implementation of NHE theory, high-quality $\langle 111 \rangle$ SOI (as required for GaN growth) is much less readily available than bulk $\langle 111 \rangle$ Si.

Recently, we have evaluated an alternative NHE substrate structure shown in Fig. 2(b). Here, a thermal oxide is grown onto a bulk silicon substrate and then patterned using interferometric lithography. This structure offers the key properties required for NHE, Si versus SiO_2 growth selectivity and a nanoscale growth nucleus; however, the substrate compliance is reduced to essentially zero because it is bulk material. We anticipate that, in this case, the mismatch strain energy will be accommodated almost entirely by the epitaxial layer and this structure is equivalent to that first discussed by Luryi and Suhir [9]. While this structure will exhibit higher strain energy than that shown in Fig. 2(a), it has an important advantage in that it uses standard silicon processing and is straightforward to fabricate over large substrate areas. The modification of the defect formation due to the local availability of free surfaces remains active. Furthermore, this approach allows NHE to be performed on refractory substrates such as SiC and sapphire, which are difficult to pattern directly.

A coalesced NHE epilayer will also exhibit strain due to thermal mismatch between the epilayer and substrate. In many cases, e.g., GaN on Si, this is sufficient to crack the epilayer. We show later that the effects of thermal strain can be mitigated by restricting the size of the heteroepitaxial area. This anticipates the eventual application, where, for example, a group of III–V devices with a limited footprint, is fabricated within a larger Si CMOS circuit. To achieve this limited area, we are also evaluating NHE structures of the type shown in Fig. 2(c). Here, the nanoscale nuclei are grouped and the number of nuclei within each group may be as small as one. Each nucleus within the group is nanosize, providing the enhanced strain relief of the NHE approach, and within the group there is coalescence. However, the group-to-group spacing is made large such that at the end of growth the groups remain separate (they do not coalesce). Raman measurements of strain in such structures (presented later) indicate that there is significantly less thermal mismatch strain in this partially coalesced NHE structure.

A. Interferometric Lithography for Nanoscale Substrate Preparation

A clear requirement for nanoheteroepitaxy is the development of a facile, scalable and inexpensive nanoscale lithographic technique. Nanoscale features are required over the large areas associated with crystal growth, up to 300-mm diameter wafers in the case of Si and SOI. There is a broad parameter space to investigate: not only the dimensions of the nm-scale features, their density, the array pattern symmetry and the nanoelement symmetry; but also the details of the nano-element structure, the grouping of nanoelements, the underlying layer structure, alternatives for passivation at the edges of the nano-elements, etc. must be explored. Coupled with the vast parameter space inherent in epitaxial growth techniques, the need for large numbers and variety of samples is evident. This argues for a nanolithography technique that both covers large areas and is easy to adapt to new patterns and feature sizes. As nanoheteroepitaxy matures and impacts devices, the demands of manufacturing will impose new constraints on any nanolithography technique.

Interferometric lithography (IL) uniquely addresses all of these requirements [19]. In this section, a brief introduction to the accessible pattern space is presented along with a comparison with other possible nanolithography/nanopatterning approaches. IL is a maskless process that uses interference between coherent laser beams to define a pattern. An interference pattern between two beams provides a periodic one-dimensional (1-D) or grating pattern. Additional exposures with the wafer rotated relative to the grating direction provide the two-dimensional (2-D) patterns often required for NHE. Alternatively, multiple laser beams can be used to write the pattern in a single exposure [20].

In order to understand the fundamental limits of IL, it is simplest to begin by analyzing 1-D-grating patterns. The fundamental relation describing any interferometric process (including conventional optical imaging) is that the pattern period p is given by:

$$p = \lambda/2\sin(\theta) \tag{1}$$

where λ is the optical wavelength and 2θ is the angle between the two incident beams with the normal to the wafer parallel to the bisector between the incident optical wavevectors. Large incident angles of 80° to 85° are realizable in practice so the limiting period is very close to $\lambda/2$. Note that this corresponds to an optical system with an $NA(=\sin\theta)$ of 0.96 to 0.98, much closer to unity than is possible with achievable optics in an imaging system. Fig. 3 shows a 180-nm period 1-D grating pattern consisting of ~40-nm wide SiO_2 stripes on a GaAs substrate written using a 355-nm frequency-tripled YAG laser source. A reactive-ion etch process was used to transfer the pattern from the photoresist into the oxide layer.

For all of the experiments reported in this paper, we used wavelengths of either 364 nm (Ar-ion laser) or 355 nm (third harmonic of a YAG laser), which are compatible with commercial I-line photoresists. Shorter wavelength coherent sources are available, in particular at 244 nm (a doubled Ar laser matched to commercial 248-nm resists for KrF laser sources) and at 213 nm (fifth-harmonic of YAG matched to 193 nm resists being developed for ArF lithography laser sources). Use of these sources

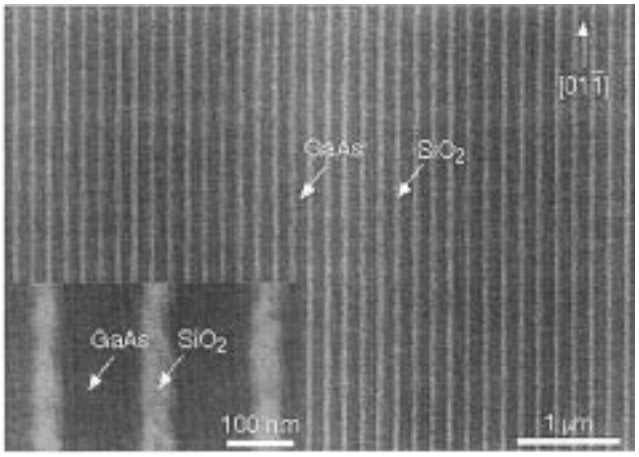


Fig. 3. Example of limits of IL: 180-nm period SiO_2 on GaAs grating using a 355 nm laser source corresponding to an NA of 0.986 (after [5]).

will allow a corresponding reduction in the period or increase in the pattern density. Further reductions can be realized by using immersion techniques, which reduce the achievable period by an extra factor of $1/n$, where $n \sim 1.5$ is the refractive index of the immersion fluid to [21, Eq. (1)]. Immersion has been demonstrated for wavelengths as short as 157 nm [22]. While there are practical issues in adapting a full optical system for immersion, for IL which only involves a limited range of angles of incidence, the optics is quite straightforward.

Another important result to take from Fig. 3 is that while the period is limited by the optical wavelength, there is no such limit on the smallest feature dimensions (often called the critical dimension or CD in lithography). In the case of Fig. 3, the CD is the oxide linewidth of 30–40 nm. These smaller features are possible as a result of the nonlinear response of the photoresist and of other processing steps which introduce higher harmonics of the fundamental frequency in the optical exposure. The resist response is inherently square law, responding to intensity not to electric field. In addition, modern commercial resists have been engineered to have highly nonlinear development character, very close to a thresholding behavior. The transition from unexposed to fully exposed occurs over a narrow dose range. This provides process latitude for the semiconductor industry, and allows us to achieve small features with almost vertical sidewalls. Additional nonlinearities are available in the processes such as reactive-ion etching used to transfer the photoresist patterns into hard layers such as the oxide layer shown in Fig. 3. The semiconductor industry routinely takes advantage of these nonlinearities in its manufacturing processes.

The pattern-array symmetry can be controlled by adjusting the IL parameters including the number of beams and the number of exposures. Fig. 4(a) and (b) show examples of both square and hexagonal pattern symmetries achieved by two IL exposures with the wafer rotated between exposures by 90° for the square and 60° for the hexagonal symmetry. These were achieved over areas of ~4 cm^2.

Controlling nano-element symmetry requires additional processing steps. Fig. 4(c) shows an example of ~100 × 100 nm square holes in photoresist that was printed with a double exposure, double-develop technique [23]. Other possibilities involve transferring partial patterns into sacrificial layers such as oxide

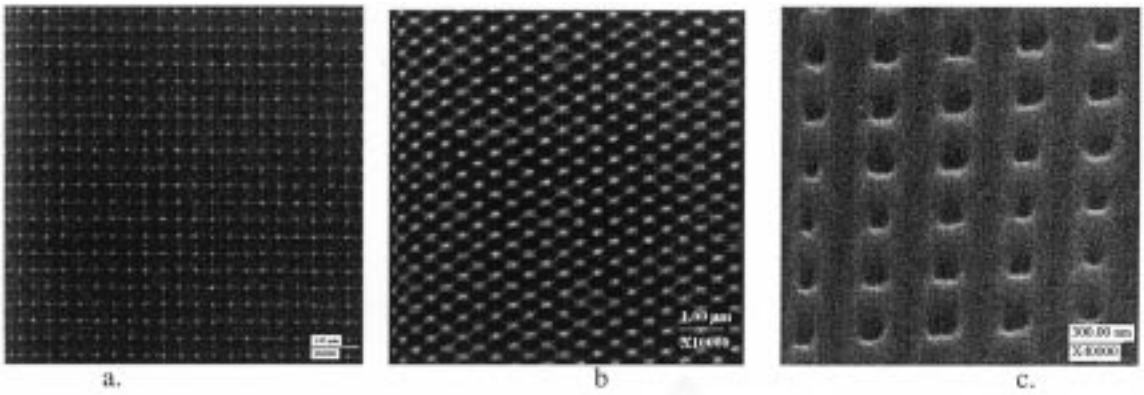


Fig. 4. Examples of 2-D pattern arrays at a period of 360 nm achieved by rotation of the wafer by: (a) 90° for the square pattern and (b) 60° for the hexagonal pattern. Feature sizes are much less than the optical wavelength due to nonlinearities the exposure and pattern transfer steps. (c) Square array of 100-nm CD square via holes printed by a double exposure/double development technique (after [24]).

TABLE I
ASPECTS OF NANOSTRUCTURE PATTERNING CONTROLLABLE BY INTERFEROMETRIC LITHOGRAPHY

Nanoisland Parameter	*NHE Impact*	*IL approach*	*Accessible Range*
Pattern periodicity	Film coalescence	Incident angles $\{p = \lambda/[2n\sin(\theta)]\}$, nonlinear IL & integration of optical and interferometric lithographies	> 70 nm
Size	Island growth, strain partitioning, active compliance	Vary exposure flux, use photoresist and other process nonlinearities.	to 20 nm
Pattern symmetry	Film coalescence	Number and orientation of beams	Rectangular, hexagonal, …
Nano-element symmetry	Island growth, defect formation	Multiple exposures, hard masks, additional nonlinearities.	Circular, square, hexagonal., …
Localized growth areas	Thermal expansion strain, integration with Si circuitry	Imaging IL, integration of conventional optical and IL, hard masks.	Fully arbitrary to ~ $\lambda/6$

or nitride films between IL steps. (These techniques were not necessary to fabricate the samples reported here.)

Isolated areas of nanostructures can be generated by combining IL with conventional optical lithography to define larger areas or by using Moiré techniques; a Moiré example will be illustrated in the GaN growth described later. More sophisticated structures are possible with imaging interferometric lithography (IIL) that allows definition of arbitrary patterns (as in imaging) with a resolution approaching that of interferometric lithography [24]. We have demonstrated dense features at a resolution of $\lambda/2$. The theoretical limit is $\sim\lambda/3$. Nonlinear techniques extend this range to $\lambda/6$. Table I summarizes the capabilities and limitation of IL for the definition of NHE substrates.

As NHE matures and impacts device capabilities, any nanolithographic technique will have to be adapted to the volume and cost requirements of manufacturing. Because of the large area capabilities of IL and the similarity to traditional optical lithography, IL is scalable to manufacturing [25]. Imprint lithography is also an attractive candidate for this role [26]. Other advanced lithographic tools such as 157-nm, F_2-laser-based steppers, extreme ultraviolet steppers, and e-beam lithography tools are likely to be too expensive for this application.

IV. NHE OF GaAs-ON-Si AND GaN-ON-Si

Fig. 5 shows the calculated total strain energy/unit area as a function of epilayer thickness for two materials systems. (a) GaAs on $\langle 001\rangle$ Si with an epitaxial alignment $[100]_{\mathrm{GaAs}}$, parallel to $[100]_{\mathrm{Si}}$. Here the lattice mismatch in the epitaxial plane is 4%. (b) GaN on silicon where the $\langle 111\rangle_{\mathrm{Si}}$ direction is parallel to the $[0001]_{\mathrm{GaN}}$ direction. In this case the lattice mismatch in the basal plane is 22%. Four heteroepitaxy approaches are plotted for each case. The superior strain relief offered by NHE is evident as NHE yields the lowest strain energy for any epilayer thickness in both material systems. The energy required to create a screw dislocation [27] is also plotted. In the case of GaAs/Si with a growth island diameter of 40 nm, the NHE strain energy remains below the dislocation formation energy for all epilayer thicknesses, suggesting that any thickness of GaAs can be grown without dislocations on islands of <40-nm diameter. The other growth approaches modeled in Fig. 5 are:

- *planar*: conventional growth on an unpatterned, thick Si wafer;
- *compliant*: growth on a 40-nm-thick Si wafer;
- *nanoscale*—patterned growth where only the epilayer (not the substrate) undergoes 3-D strain.

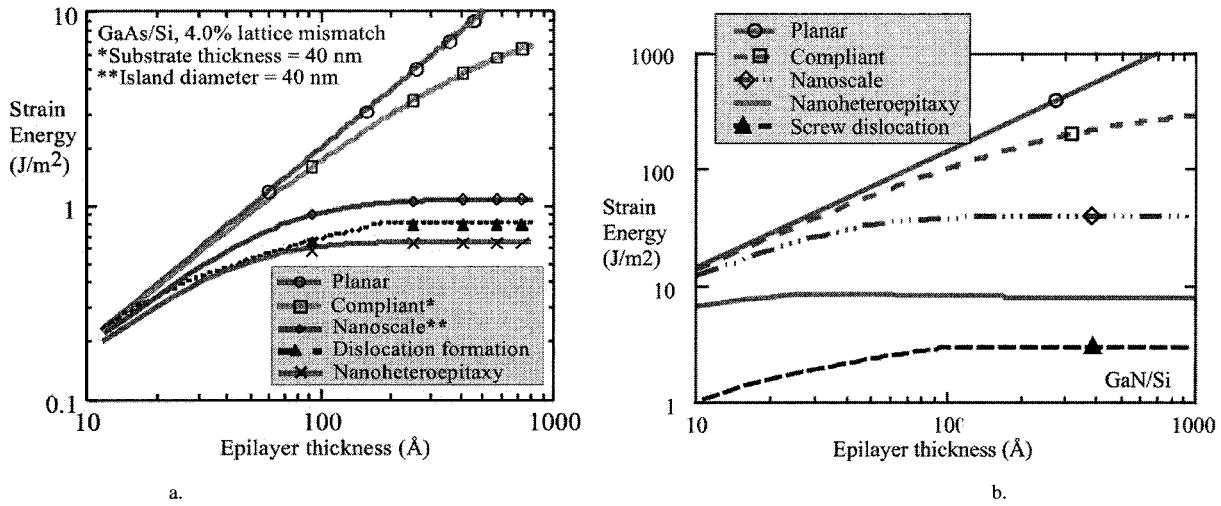


Fig. 5. Strain energy versus epilayer thickness for NHE and other growth approaches (after [10]) for two materials systems: (a) GaAs-on-Si and (b) GaN-on-Si.

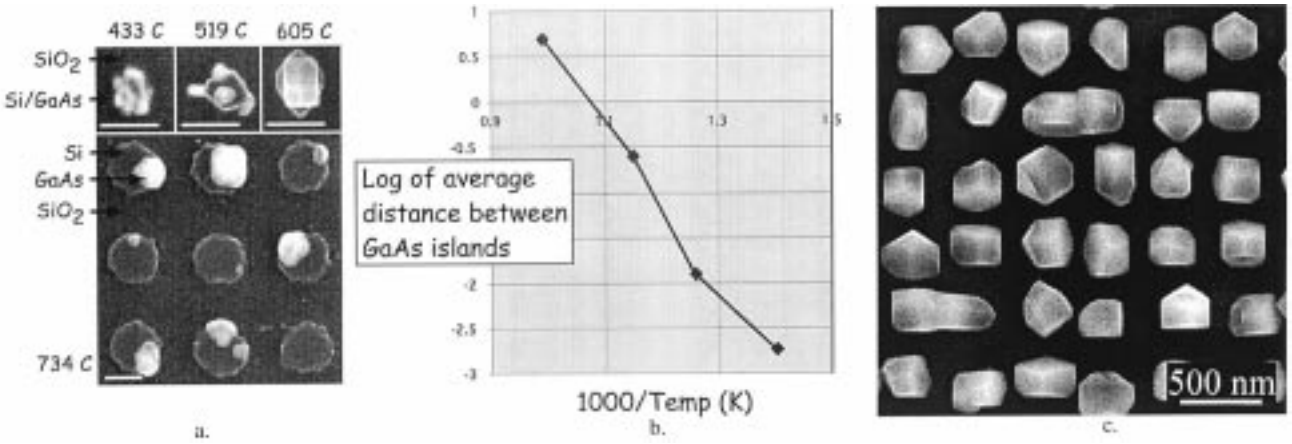


Fig. 6. (a) Plan view SEM micrograph of GaAs nucleation on Si islands at four different temperatures. (b) Arrenhius plot of average distance between GaAs islands versus inverse temperature. (c) SEM micrograph showing NHE GaAs nuclei grown at 605 °C.

All of these are above the screw dislocation energy, suggesting that dislocations are inevitable for this island diameter.

The large mismatch in the GaN/Si system means that the strain energy in the epilayer is always above the energy required to create a screw dislocation, and we do not expect to eliminate mismatch defects by NHE strain management. However, the unique substrate/epilayer interface of the NHE approach contains extensive free surface area which offers opportunities for defect termination. There is prior evidence that the density of mismatch defects in large mesa structures can be reduced when these defects are allowed to diffuse to the mesa edge [5], [6]. For GaN, where the commonly observed defects are c-oriented edge dislocations [28] which have a 1/3 [11$\bar{2}$0] Burger's vector in the basal plane, it is expected that these defects will glide to the edge of a nanoscale growth nucleus. In an NHE sample, where each nucleus is disconnected from its neighbor, a defect can terminate at the edge of the nucleus. However, we do not expect this type of defect termination to be possible during growth on a porous-Si substrate, where there is a crystalline connection through the substrate skeleton and where the Burger's vector must be conserved across the whole layer.

The NHE approach is quite generic and is not limited to these two materials systems. Amongst other heteroepitaxial materials combinations of interest are: $Si_{1-x}Ge_x$/Si, GaN/sapphire, GaN/SiC, $Ga_{1-x}In_xAs$/InP, and $B_{12}As_2$/SiC. NHE theory predicts that lattice mismatch defect elimination should be possible for mismatch strains up approximately 4%. However, as we show later in the case of GaN on Si, even with a mismatch strain much higher than this 4% value, the unique nanoscale features of the NHE interface presents additional mechanisms for defect reduction.

V. The Importance of Growth Selectivity in NHE

A critical step in NHE is the initial nucleation on the patterned substrate. We require that there is selective nucleation on each substrate island as shown schematically in Fig. 1, but not on the surrounding growth mask. Optimization of the MOCVD growth temperature allows this requirement to be satisfied for both GaAs/Si and GaN/Si.

Fig. 6 gives a synopsis of the observed GaAs/Si nucleation process. The upper part of Fig. 6(a) shows the typical nucleation

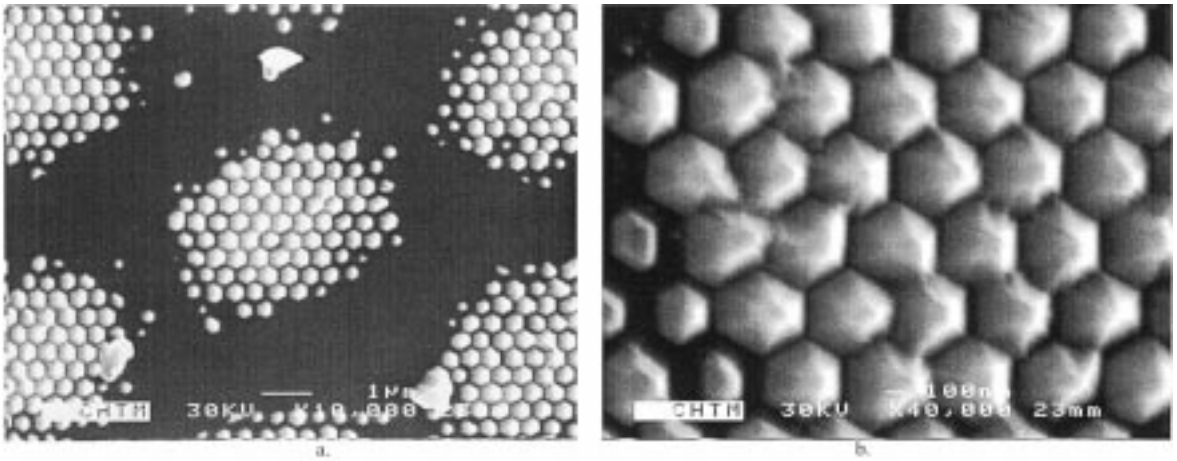


Fig. 7. Plan view SEM micrographs of GaN-on-Si using the "grouped" hexagonal array NHE sample structure that was shown in Fig. 2(c).

behavior at growth temperatures of 433 °C, 519 °C, and 605 °C. The lower part of this figure shows the behavior at 734 °C. The surface mobility of the growth species is adequate to provide selectivity between the Si and SiO_2, so that growth only occurs on the Si islands. This is expected from previous results [29], which indicate that, under the growth conditions used here, the migration length for GaAs on SiO_2 masked surfaces will be in the range 100–1000 μm. However, the surface migration length on the Si islands is much shorter, and in Fig. 6(b), we plot the average distance between GaAs growth nuclei on the Si islands as a function of temperature. At low temperature (433 °C), the GaAs nucleus separation is small and there are multiple nuclei on each Si island. This separation increases with growth temperature until at 605 °C we have the optimum number (one) of nuclei per Si island. If we increase the temperature further, then as the lower part of Fig. 6(a) shows, the migration length becomes too long and some of the Si islands are left without GaAs nuclei. This migration appears to be a well behaved, temperature activated process and Fig. 6(b) yields an activation energy of 0.71 eV for the migration of the GaAs growth species on the Si islands. These data illustrate the critical role of growth selectivity in the NHE process. If the growth temperature (and selectivity) are too low, we have multiple nuclei on each growth island which can lead to twinning and polycrystalline growth [15]. The nanopatterned substrate for this growth was produced using the Moiré techniques outlined earlier. Four two-beam exposures, each with identical 360-nm pitch were used. The wafer rotation was 0°, 60°, 62°, and 122° for the four exposures. This produced clustered arrays of nanoscale (~120-nm diameter) holes with a local spacing of 360 nm and a cluster spacing of ~10 μm. Fig. 7(a) shows GaN nucleation and that the selectivity offered by the simple oxide growth mask is adequate. This figure also reveals that the GaN nuclei exhibit well defined hexagonal facets, which have been determined previously [30] to be of the prismatic $\langle 01\underline{1}n\rangle$ family, which are more stable than the orthogonal $\langle 11\underline{2}n\rangle$ facets.

In both material systems discussed earlier, the nuclei exhibit 3-D growth behavior, reflecting either a Volmer–Weber or a Stranski–Krastanov growth mode, which is expected due to the large lattice mismatches involved. In the absence of a growth mask, the size and separation of these 3-D nuclei would be more random but would increase naturally with increasing growth temperature. The optimum growth temperature (typified by growth at 605 °C for the GaAs-on-Si system) therefore represents a resonance condition, where the natural spacing of the nuclei that is determined by migration length, is equal to the period of the pattern on the NHE substrate.

VI. Uncoalesced Nanoscale Nuclei

Previously reported TEM examination of GaAs growth on the Si islands [15] revealed a tendency to twinning at lower growth temperature, while at 605°C the GaAs was typically single crystal, but has stacking fault defects on the $\langle 111\rangle$ planes. While defects have not yet been eliminated by NHE growth, the defect structure in NHE samples of GaAs-on-Si was significantly different from that observed in regular GaAs on planar Si growth. Specifically, the incidence of threading dislocations has been largely replaced in NHE by stacking fault defects. This is also the case for NHE GaN-on-Si growth [31], where again the concentration of threading type defects was significantly reduced and the most frequently occurring defects were found to be stacking faults. In the case of GaAs-on-Si, the $\langle 111\rangle$ stacking faults are on inclined planes that propagate into the epilayer. In the case of GaN-on-Si, the stacking defects are in the basal plane, and while there is a high defect density at the heterointerface, these defects are confined to the interface and do not propagate into the epilayer. Cross-sectional transmission electron microscopy (XTEM) for a GaN nucleus measuring 0.32 μm across is shown in Fig. 8(a). The NHE sample in this case was of the type shown in Fig. 2(b) with an approximately 120-nm diameter hole size. At this hole size, NHE theory suggests that there will be little 3-D deformation or "nanoscale advantage." Interestingly, however, we observe the same defect behavior [see Fig. 8(b)] that we reported earlier [11] for much smaller (70-nm diameter) GaN-on-Si nuclei, i.e., disorder on the Si side of the interface, a high density of stacking faults in the first layers of GaN, defect-free GaN growth away from the interface, and a significantly reduced incidence of c-oriented threading dislocations.

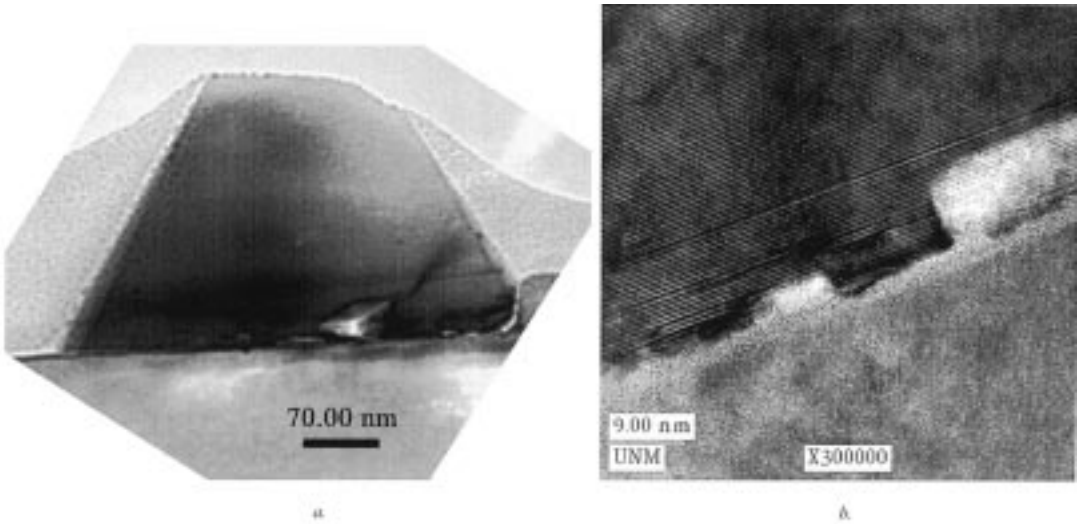


Fig. 8. XTEM micrographs showing (a) absence of threading dislocations in 0.32-μm nucleus. (b) High-resolution XTEM micrograph showing stacking faults at heterointerface.

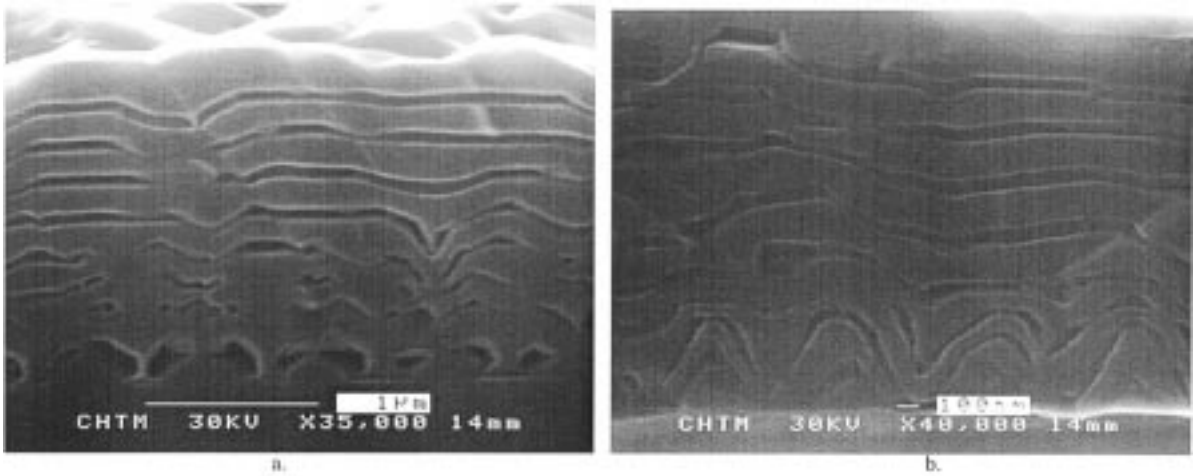


Fig. 9. Orthogonal (a) cleaved and (b) stained GaAs NHE samples.

VII. Coalescence of NHE Films

Coalescence behavior is very different in the GaAs-on-Si and GaN-on-Si materials systems and we attribute this difference in part to the different faceting behavior of the nuclei, that can be seen by a comparison of Figs. 6(c) and 7. GaAs nuclei exhibit nonuniform faceting that leads to local variations of growth rate and uneven coalescence. In contrast, the facets of the GaN nuclei are extremely uniform and this permits the simultaneous coalescence of all parts of the film.

A second and related cause of poor coalescence in NHE GaAs is revealed by Fig. 9, which shows SEM views of orthogonal cleaves from a coalesced GaAs-on-GaAs NHE film. (In this experiment, GaAs was grown directly on a nanopatterned NHE GaAs substrate to specifically study GaAs coalescence in the absence of strain.) The darker regions are marker layers of $Al_{0.4}Ga_{0.6}As$ that were grown approximately every 300 nm, to allow visualization of the evolution of the film morphology. The nanopatterning period is 0.5 μm and the initial pyramidal shape of the GaAs nuclei can be clearly seen at the bottom of each figure. For the cleave direction shown in Fig. 9(a), one can define three regions. In region 1 (up to the second marker layer), the GaAs film is planarizing. In region 2 (second marker layer to fifth marker layer), the film becomes nonplanar with a roughness that is unrelated to the pattern period. In region 3 (after the fifth marker layer), the film begins to planarize again. Fig. 9(b) shows the orthogonal cleave for the same growth. In this case, the growth is conformal up to the third marker layer, then the film begins to planarize. This difference in growth habit is, we believe, a second major cause of the poor morphology observed for the coalesced GaAs films. Even though the GaAs surface planarizes rapidly in one direction [Fig. 9(a)] this planarity becomes disturbed by the nonplanar, conformal growth that is simultaneously taking place in the orthogonal direction [Fig. 9(b)]. The conformal growth will develop a range of facets, and the faster growth of these facets will compete for growth nutrient causing a local reduction of growth rate in the orthogonal direction. Overall planarization

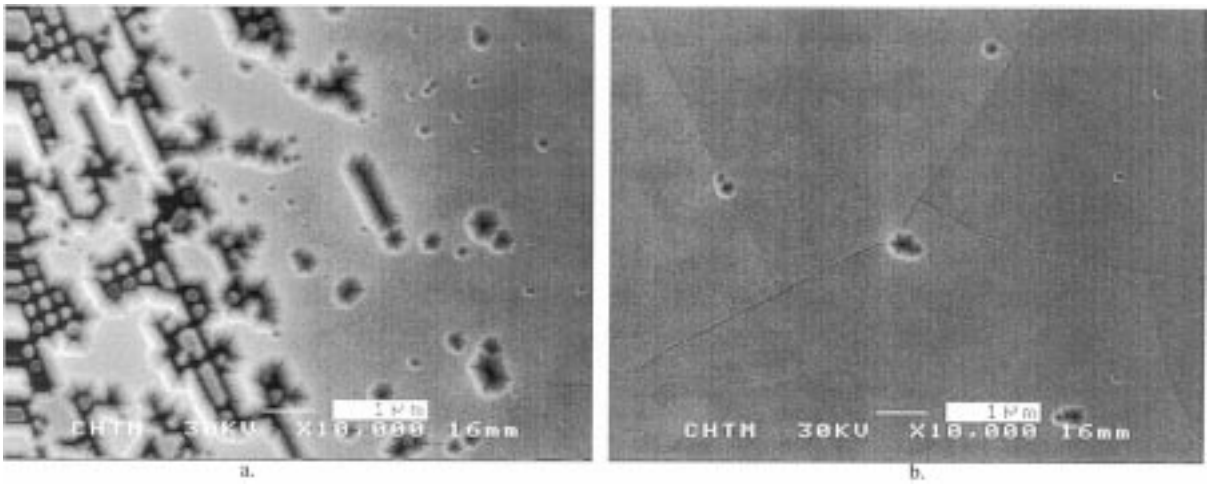


Fig. 10. NHE GaN-on-Si using a square 2-D sample structure. (a) Edge region where growth rate is reduced. (b) Center of sample where full coalescence has taken place.

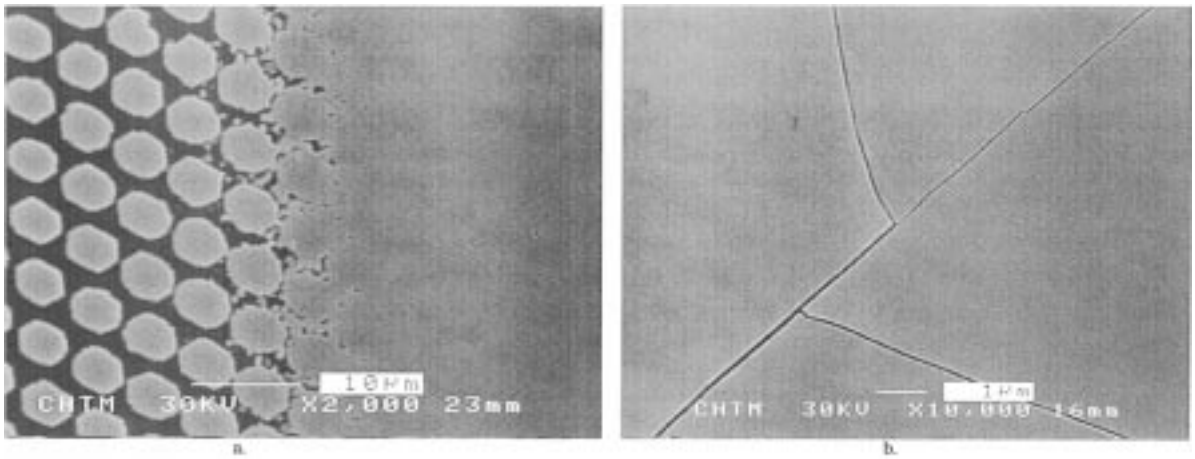


Fig. 11. NHE GaN-on-Si using a hexagonal "grouped" sample structure. (a) Edge region where growth rate is reduced. (b) Center of sample where full coalescence has taken place.

can, therefore, only take place when both directions planarize equivalently. As might be expected, macroscopic measurements (XRD, PL, etc.,) of fully coalesced GaAs films of this type indicate substandard quality by comparison with bulk GaAs and more optimization of NHE coalescence conditions is clearly necessary.

Despite the large mismatch between GaN and Si, the coalescence of GaN appears to be well behaved and large areas of smooth films are readily produced. This is a result of the highly uniform size and shape of the nuclei as shown in Fig. 7. Figs. 10 and 11 compare the coalescence of NHE GaN-on-Si for a uniform square array (Fig. 10) and a grouped hexagonal array (Fig. 11). Figs. 10(a) and 11(a) show the sample edge where the growth rate is lower and this region reveals the superior (more uniform) coalescence obtained with the hexagonal array. The sample shown in Fig. 11(a) used the "grouped" NHE structure shown in Fig. 2(c); thus, the elongated hexagons at the left of Fig. 11(a) are the already coalesced groups of nuclei [note the scale difference between Figs. 10(a) and 11(a)]. Figs. 10(b) and 11(b) show the corresponding fully coalesced regions of these samples and, as expected, cracking is evident as a result of the large thermal expansion mismatch.

VIII. Characterization of Coalesced NHE GaN-on-Si Films

Atomic force microscopy measurements on the coalesced NHE GaN films showed rms roughness values of between 5 and 6 Å for all samples, including the regions exhibiting cracks. Cracking is evident in the fully coalesced regions of all GaN-on-Si samples and is expected from the thermal mismatch between GaN and Si, which puts the GaN in tension. Bright field cross-sectional TEM (Fig. 12) on fully coalesced GaN-on-Si NHE films (2-D square array sample structure) yields a defect density of 3×10^9 cm^{-2}, assuming a specimen thickness of 750 Å.

Room temperature PL measurements reported earlier [32] indicated that coalesced NHE GaN-on-Si films have a significantly higher luminescence efficiency than either planar GaN-on-Si films or planar GaN-on-sapphire films.

Residual strain in these samples was measured by micro-Raman scattering measurements made in a standard backscattering geometry using a 488-nm, Ar-ion-laser source, a $40\times$ microscope objective (spot size ~ 1 μm), a 1-m double spectrometer and an optical multichannel analyzer (OMA).

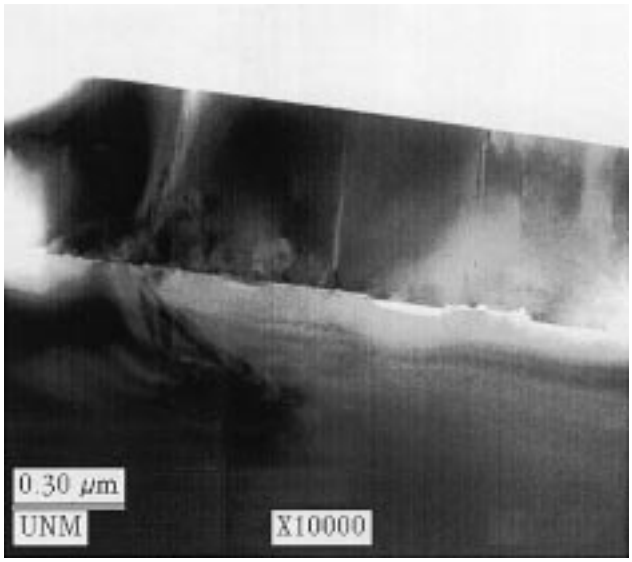


Fig. 12. Bright field cross sectional TEM micrograph on coalesced NHE GaN-on-Si on the 2-D square array sample structure.

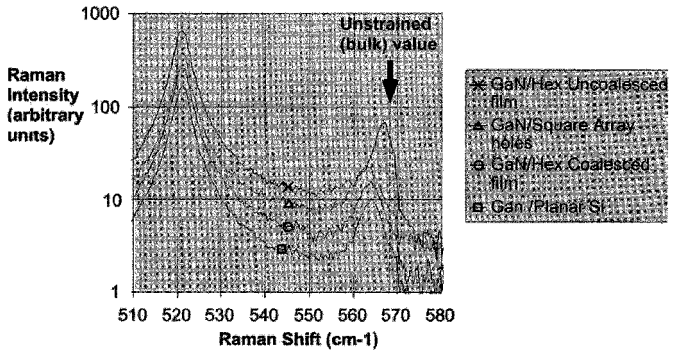


Fig. 13. Raman spectrum showing the Si substrate peak at 521 cm^{-1} and the GaN (E2) peak in the range 560 to 570 cm^{-1}. (Note: The traces have been displaced vertically for clarity.)

Spectral shifts were calibrated against emission lines from the Ar plasma with a repeatability of $\sim$ one diode in the OMA array or 0.12 cm^{-1}. Typical micro-Raman spectra for various sample structures, showing a Si peak at 521 cm^{-1} and the GaN E_2 peak in the vicinity of 570 cm^{-1}, are given in Fig. 13. Table II summarizes Raman data from these measurements for fully coalesced and partially coalesced GaN films. [The partially coalesced films are on the "grouped" NHE sample structure as shown at the left of Fig. 11(a), where there is coalescence within the group of nanonuclei but adjacent groups have not yet coalesced.] The shift of the Raman E_2 peak was translated to biaxial stress in the epitaxial GaN film using the approach described by Reiger *et al.* [33].

For all of the GaN growths on silicon, the E_2 Raman peak wavenumber is less than the bulk (unstrained) value of 568 cm^{-1}, indicating that the GaN is in tension. For the reference sample of GaN grown on sapphire, the E_2 Raman peak wavenumber is greater than the bulk value indicating that the GaN film is in compression. These results are consistent with previous data [33] and with the in-plane thermal expansion coefficients: $\alpha_{\mathrm{Si}} = 2.6 \times 10^{-6}/^{\circ}\mathrm{C} < \alpha_{\mathrm{GaN}} = 5.6 \times 10^{-6}/^{\circ}\mathrm{C} < \alpha_{\mathrm{sapphire}} = 7.5 \times 10^{-6}/^{\circ}\mathrm{C}$. The rms variation in wavenumber detailed in the σ column of Table II is much larger than the experimental repeatability and represents a real variation of local strain in the sample. This variation is at the measurement limit for the planar samples (0.12 cm^{-1}) and large ($\sim$1 cm^{-1}) for the coalesced NHE samples. Interestingly, the partially coalesced NHE hexagonal sample exhibits a small spread in wavenumbers and also has the smallest stress of only -0.09 GPa. This stress is less than 20% of the stress measured in the other NHE samples and is approximately half of the low stress achieved by growing GaN onto porous GaN sublayers [34]. This suggests that the "grouped" NHE sample structure [Fig. 2(c)] may be a viable route to higher quality GaN-on-Si. More detailed mapping [35] of these stress variations, in both the coalesced and partially coalesced cases, is underway and should yield additional useful information to guide future development of NHE.

The GaN bandgap peak energy measured by photoluminescence (PL) can also be used to determine residual stress in GaN films. At room temperature, the band edge peak shifts by approximately 24 meV per GPa of biaxial stress [33]. The right-hand column of Table II lists the GaN room temperature, photoluminescence peak energy measured for these GaN samples. We calculate that the GaN-on-planar-Si and the GaN-on-Si square array samples have a tensile stress that is approximately -1 GPa relative to the stress in the partially coalesced NHE GaN film (on the "grouped" hexagonal array). This is approximately twice the value of stress indicated by the Raman measurement but the agreement is not unreasonable given the large width of the room temperature PL peak.

IX. Discussion and Conclusion

There are both similarities and differences in the nucleation and coalescence phases of the NHE growth in the two material systems we have investigated. For both GaAs/Si and GaN/Si, stacking faults are found to be the dominant defect. For the GaAs/Si case, these defects nucleate on the $\{111\}$ sidewalls of the growth nuclei and are problematic because they propagate up into the epitaxial layer. This type of sidewall growth nucleation may be eliminated by using for example, the NHE structure shown in Fig. 2(b) (holes in an oxide overlayer atop a blanket Si film, similar to that demonstrated for the GaN growth). However, this type of NHE structure has an inherently lower compliance than the SOI structure of Fig. 2(a).

In GaN/Si NHE growth, the $\langle 111 \rangle$ Si facet is the lowest energy face and the problem of nucleation on the sidewalls does not occur with island growth [11] (furthermore, sidewall growth is impossible for the SiO_2-masked blanket Si growth investigated here). However, the lattice mismatch in this material system is too large to be accommodated by the nanoscale growth and defect formation is inevitable. The most significant impact of the NHE growth is in the character of the defects which are stacking faults confined to a thin region just at the initiation of the growth. These defects appear to largely replace the threading dislocations that propagate throughout the film and that are characteristic of planar (large area) GaN/Si growth.

TABLE II
MICRO-RAMAN E_2 MEASUREMENTS AND AVERAGE STRESS FOR GaN FILMS. (NOTE: A NEGATIVE STRESS INDICATES THAT GaN FILM IS IN TENSION AND A POSITIVE STRESS INDICATES COMPRESSION)

Sample	$<E_2>$ (cm^{-1})	σ_{E_2} (cm^{-1})	Average biaxial stress (GPa)	RT PL peak (eV)
GaN on planar Si	563.82	0.12	- 0.52	3.396
NHE GaN on square 2D array (coalesced)	563.53	1.21	- 0.56	3.396
NHE GaN on hexagonal 2D array (coalesced)	563.95	0.84	- 0.51	3.420
NHE GaN on hexagonal 2D array (partially coalesced)	567.3	0.25	- 0.09	
GaN on planar sapphire	569.18	0.12	+ 0.15	--

The kinetics of the coalescence are also dramatically different for the two material systems. Because of the differences in growth rate along different directions of the GaAs surface, any uneven growth at the initial separated island stage is not smoothed over in the coalescence but rather is amplified, leading to an uneven growth habit. Clearly, much further investigation is needed before useful device quality material is achieved.

Raman measurements on coalesced GaN/Si films show a level of strain consistent with the thermal expansion coefficients of the materials involved. This implies that the lattice mismatch is largely accommodated at growth temperature and that the residual strain is due mainly to differential thermal expansion arising during postgrowth cooling. In the absence of other strain relief mechanisms, Raman measurements should yield the same level of strain, a dimensionless parameter, for large and small areas. The observation of a significantly reduced strain in the partly coalesced NHE GaN films, therefore, suggests an additional mechanism is at work in these samples. We speculate that in these smaller area, partially coalesced samples, much of the thermal expansion mismatch is being accommodated by a relative motion of the film and the substrate during cooling as proposed by Lo *et al.* [12]. This relative motion is clearly an extensive parameter that must increase with the linear dimensions of the film and, thus, is not possible in the large areas of the fully coalesced samples. We are currently pursuing micro-Raman, high-resolution "mapping" of these samples to study the spatial variation of strain and correlate this with the underlying NHE sample structure. Clearly, more investigations of this strain relaxation as a function of the size of the NHE group and the extent of the overgrowth will be critical to understanding this phenomenon.

Another interesting observation is the apparent loss of selectivity just before coalescence in the GaN/Si system as seen in Fig. 11(a). On the right side of the figure, the film is fully coalesced. On the left side, individual hexagons of GaN at a scale of several nanometers are evident. Just at the transition region, there is a large number of small nuclei to the side of, and independent of, the Si nucleation sites that are buried under the large hexagons. Presumably, this would result in a polycrystalline material since these nuclei are likely randomly oriented. Yet the resulting film is single crystal. Do these nuclei reorient as the film comes together? Why does the selectivity, which is very good away from this interface, break down in this transition region? Clearly, much more needs to be understood about the coalescence regime.

For both material systems, the unique features of NHE lead to substantial changes in the defect structure over traditional blanket film growth. There is a large parameter space to investigate and systematic investigation is necessary to find the optimal combination of NHE substrate and growth conditions. Interferometric lithography offers a large degree of flexibility to aid in this investigation while, at the same time, is suited for scaling to higher levels of manufacturing as devices and circuits based on the integration that NHE offers between Si and other material systems evolves. While the potential of defect-free lattice mismatched materials and devices has yet to be realized, NHE continues to demonstrate promise toward that potential. At the same time NHE is providing a rich array of new epitaxial growth phenomena that will lead to significant improvements in the science of heterostructure growth.

REFERENCES

[1] S. Koshiba, Y. Nakamura, T. Noda, S. Watanabe, H. Akiyama, and H. Sakaki, "Transformation of GaAs (001)-(111)B facet structure by surface diffusion during molecular beam epitaxy on patterned substrates," *J. Cryst. Growth*, vol. 227–228, pp. 62–66, 2001.

[2] Y. T. Sun, E. Rodriguez Messmer, D. Soderstrom, D. Jahan, and S. Lourdudoss, "Temporally resolved selective area growth of InP in the openings off-oriented from [110] direction," *J. Cryst. Growth*, vol. 225, pp. 9–15, 2001.

[3] Y. Honda, Y. Kawaguchi, Y. Ohtake, S. Tanaka, M. Yamaguchi, and N. Sawaki, "Selective area growth of GaN microstructures on patterned (111) and (001) Si substrate," *J. Cryst. Growth*, vol. 230, pp. 346–350, 2001.

[4] K. Hiramatsu, "Epitaxial lateral overgrowth techniques used in group III nitride epitaxy," *J. Phys: Condens. Matter.*, vol. 13, pp. 6961–6975, 2001.

[5] S. Guha, A. Madhukar, and L. Chen, "Defect reduction in strained InxGal–xAs via growth on GaAs (100) substrates patterned to submicron dimensions," *Appl. Phys. Lett.*, vol. 56, pp. 2304–2306, 1990.

[6] E. A. Fitzgerald, G. P. Watson, R. E. Proano, D. G. Ast, P. D. Kirchner, G. D. Pettit, and J. M. Woodall, "Nucleation mechanisms and the elimination of misfit dislocations at mismatched interfaces by reduction in growth area," *J. Appl. Phys.*, vol. 65, pp. 2220–2237, 1989.

[7] K. T. Campbell, C. H. Joyner, C. W. Ebert, A. Robertson, D. P. Wilt, and J. L. Zilko, "Enabling terabit systems through MOVPE," *J. Cryst. Growth*, vol. 225, pp. 391–396, 2001.

[8] B. S. Shelton, D. J. H. Lambert, J. J. Huang, M. M. Wong, U. Chowdhury, T. G. Zhu, H. K. Kwon, Z. Leliental-Weber, M. Bernarama, M. Feng, and R. D. Dupuis, "Selective-area growth and characterization of AlGaN/GaN heterojunction bipolar transistors by metalorganic chemical vapor deposition," *IEEE Trans. Electron Dev.*, vol. 48, pp. 490–494, 2001.

[9] S. Luryi and E. Suhir, "New approach to the high quality epitaxial growth of lattice-mismatched materials," *Appl. Phys. Lett.*, vol. 49, pp. 140–142, 1986.
[10] D. Zubia and S. D. Hersee, "Nanoheteroepitaxy: The application of nanostructuring and substrate compliance to the heteroepitaxy of mismatched semiconductor materials," *J. Appl. Phys.*, vol. 85, pp. 6492–6496, 1999.
[11] D. Zubia, S. H. Zaidi, S. R. J. Brueck, and S. D. Hersee, "Nanoheteroepitaxial growth of GaN on Si by OMVPE," *Appl. Phys. Lett.*, vol. 76, pp. 858–860, 2000.
[12] Y. H. Lo and Z. H. Zhu, "Compliant substrates with an embedded twist boundary," in *Mat. Res. Soc. Symp. Proc.*, vol. 510, 1998, pp. 81–87.
[13] M. Sugawara, Ed., *Self-Assembled InGaAs/GaAs Quantum Dots*. San Diego, CA: Academic, 1999.
[14] S.-C. Lee, L. R. Dawson, K. J. Malloy, and S. R. J. Brueck, "Molecular beam epitaxial growth of one-dimensional rows of InAs quantum dots on nanoscale-patterned GaAs," *Appl. Phys. Lett.*, vol. 79, pp. 2630–2633, 2001.
[15] D. Zubia, S. Zhang, R. Bommena, X. Sun, S. R. J. Brueck, and S. D. Hersee, "Initial nanoheteroepitaxial growth of GaAs on Si(100) by OMVPE," *J. Electron. Mater.*, vol. 30, pp. 812–816, 2001.
[16] K. L. Chopra, *Thin Film Phenomena*. New York: McGraw-Hill, 1969, pp. 163–171.
[17] M. Ohring, *The Materials Science of Thin Films*. New York: Academic, 1992, pp. 213–219.
[18] H. Matsuchima, M. Yamaguchi, K. Hiramatsu, and N. Sawaki, "Submicron fine structure of GaN by metal–organic vapor phase epitaxy (MOVPE) selective area growth (SAG) and buried structure by epitaxial lateral overgrowth (ELO)," *J. Cryst. Growth*, vol. 189/190, pp. 78–82, 1998.
[19] S. R. J. Brueck, "Interferometric lithography—From periodic arrays to arbitrary structures," in *Micro and Nano Eng. '97*, M. Hazatkis and E. Gogolides, Eds. Amsterdam: Elsevier, 1998, pp. 145–148.
[20] X. Chen, S. H. Zaidi, S. R. J. Brueck, and D. J. Devine, "Interferometric lithography of sub-micrometer sparse hole arrays for field-emission display applications," *J. Vac. Sci. Technol.*, vol. B14, pp. 3339–3349, 1996.
[21] J. A. Hoffnagle, W. D. Hinsberg, M. Sanchez, and F. A. Houle, "Liquid immersion deep ultraviolet interferometric lithography," *J. Vac. Sci. Technol.*, vol. B17, pp. 3306–3309, 1999.
[22] M. Switkes and M. Rothschild, "Immersion lithography at 157 nm," *J. Vac. Sci. Technol.*, vol. B19, pp. 2353–2356, 2001.
[23] S. H. Zaidi and S. R. J. Brueck, "Nonlinear processes to extend interferometric lithography," *Proc. SPIE Microlithography Emerging Technologies III*, vol. 3676, pp. 371–378, 1999.
[24] X. Chen and S. R. J. Brueck, "Imaging interferometric lithography—Approaching the resolution limits of optics," *Opt. Lett.*, vol. 24, pp. 124–126, 1999.
[25] S. H. Zaidi, S. R. J. Brueck, F. M. Schellenberg, R. S. Mackay, K. Uekert, and J. J. Persoff, "Interferometric lithography tool for 180-nm structures," in *SPIE Microlithography '97, Emerging Lithographic Technologies*, vol. 3048, 1997, pp. 248–254.
[26] Y. Xia and G. M. Whitesides, "Soft lithography," *Angew. Chem. Int. Ed.*, vol. 37, pp. 550–575, 1998.
[27] R. People and J. C. Bean, "Calculation of critical layer thickness versus lattice mismatch for Ge_xSi_{1-x}/Si strained-layer heterostructures," *Appl. Phys. Lett.*, vol. 47, pp. 322–324, 1985.
[28] S. D. Hersee, J. C. Ramer, and K. J. Malloy, "The microstructure of MOCVD GaN on sapphire," *MRS Bullet.*, vol. 22, pp. 45–51, 1997.
[29] K. Kiruma, T. Haga, and M. Miyazaki, "Surface migration and reaction mechanism during selective growth of GaAs and AlAs by metalorganic chemical vapor deposition," *J. Cryst. Growth*, vol. 102, pp. 717–721, 1990.
[30] S. Nitta, T. Kashima, M. Kariya, Y. Yukawa, S. Yamaguchi, H. Amano, and I. Akasaki, "Mass transport, faceting and behavior of dislocations in GaN," *MRS Internet J. Nitride Semicond. Res.*, vol. 5S1, p. W2.8, 2000.
[31] D. Zubia, S. H. Zaidi, S. R. J. Brueck, and S. D. Hersee, "Nanoheteroepitaxial growth of GaN on Si by OMVPE," *Appl. Phys. Lett.*, vol. 76, pp. 858–860, 2000.
[32] D. Zubia, S. H. Zaidi, S. D. Hersee, and S. R. J. Brueck, "Nanoheteroepitaxy: Nanofabrication route to improved epitaxial growth," *J. Vac. Science Technol.*, vol. 18, pp. 3514–3520, 2000.
[33] W. Reiger, T. Metzger, H. Angerer, R. Dimitrov, O. Ambacher, and M. Stutzmann, "Influence of substrate-induced biaxial compressive stress on the optical properties of thin GaN films," *Appl. Phys. Lett.*, vol. 68, pp. 970–972, 1996.
[34] M. Mynbaeva, A. Titkov, A. Kryzhanovski, I. Kotousova, A. S. Zubrilov, V. V. Ratnikov, V. Yu. Davydov, N. I. Kuznetsov, K. Mynbaev, D. V. Tsvetkov, S. Stepanov, A. Cherenkov, and V. Dmitriev, "Strain relaxation in GaN layers grown on porous GaN sublayers," *MRS Internet J. Nitride Semicond. Res.*, vol. 4, p. 14, 1999.
[35] S. R. J. Brueck, B.-Y. Tsaur, J. C. C. Fan, D. V. Murphy, T. F. Deutsch, and D. J. Silversmith, "Raman measurements of stress in silicon-on-sapphire device structures," *Appl. Phys. Lett.*, vol. 40, pp. 895–897, 1982.



Stephen D. Hersee (SM'96) was born in Hastings, U.K., in 1950. He received the B.Sc. degree (with first-class honors) in 1968 and the Ph.D. degree in 1975, both from Brighton Polytechnic, Brighton, U.K.

He was with The Plessey Co. Ltd. (U.K.) from 1975 to 1980, Thomson CSF (France) from 1980 to 1986, and General Electric (Syracuse, NY) from 1986 to 1991. He joined the University of New Mexico at Albuquerque in 1991, where he is currently a Professor in the Electrical and Computer Engineering Department. He teaches courses in semiconductor materials, advanced heterojunction devices, and semiconductor process technology. He is also a member of the Center for High Technology Materials. His research interests include novel semiconductor materials and heteroepitaxy for advanced materials and devices.



David Zubia was born in Delicias, Chihuahua, Mexico, in 1964. He received the Bachelor's and Master's degrees from the University of Texas at El Paso in 1989 and 1993, respectively, and the Ph.D. degree in electrical engineering from the University of New Mexico (UNM) at Albuquerque in 2000. His thesis work involved the heteroepitaxy of lattice mismatched materials on nanopatterned substrates and led the development of nanoheteroepitaxy.

In 1993, he joined Golden Photon, Inc (GPI), Golden, CO, where his research involved investigation of the degradation mechanisms in CdS/CdTe solar cells. In 2000, he joined UNM as Research Assistant Professor and Manager of the Crystal Growth Facility at the Center for High Technology Materials. His research included the utilization of nanopatterned substrates for integration of compound semiconductors with silicon, the development of radiation hard semiconductors, and GaInNAs vertical-cavity surface-emitting lasers. He taught classes in silicon VLSI technology while at UNM. In 2001, he joined the University of Texas at El Paso as an Assistant Professor, where his current research interests include studying the nanoscale aspects of materials and developing nanotechnology to create novel semiconductor materials and devices, including scientific studies of nucleation and coalescence of semiconductor nanocrystals and strain relief in nanoporous substrates. His technological interests include the development of small- and direct-bandgap semiconductor alloys, lattice and thermodynamically matched to silicon VLSI. His other technological interests are the development of advanced materials utilized for energy cells and the creation of novel nanodevices.

Dr. Zubia is a Guest Editor for the IEEE JOURNAL OF QUANTUM ELECTRONICS. He is a Ford Foundation Fellow, Chair of the Third International Conference on Alternative Substrate Technology, and a member of the American Association for Crystal Growth.



Xinyu Sun received the Bachelor's degree in 1994 and the Master's degree in 1997 from Hebei University of Technology, Hebei, China. She is currently working toward the Ph.D. degree in microelectronics at the University of New Mexico at Albuquerque.

Her research interests include studies of III-V semiconductor, wide bandgap nitrides, MOCVD growth, and defect elimination in mismatched heteroepitaxy.



R. Bommena received the Bachelor's degree in electronics and instrumentation from Kakatiya Institute of Technology and Science, A.P., India, in 1999. He is currently working toward the Ph.D. degree at the University of New Mexico at Albuquerque (UNM).

He is a Research Assistant with the Center for High Technology Materials, UNM. His research has been mainly focused on nanoheteroepitaxy, a technique developed at UNM to reduce defects in mismatch epitaxy. His experience includes interferometric lithography and thin-film characterization techniques, including scanning electron microscopy, atomic force microscopy, photoluminescence, x-ray diffraction, and Raman spectroscopy studies for stress measurements in epitaxial films.



Mike Fairchild received the B.A. degree in philosophy, with minors in mathematics and physics, in 1989 from the University of New Mexico at Albuquerque, where he is currently enrolled in the Electrical Engineering program.

His principal research interests are epitaxial growth of III-V materials and metrology.

S. Zhang received the B.S. and M.S. degrees in physics from Jilin University, ChangChun, China, in 1993 and 1996, respectively, and the M.S. degree in physics from Northeastern University in 1999. He is currently working toward the Ph.D. degree in the electrical and computer engineering at the Center for High Technology Materials, University of New Mexico.



David Burckel received the B.S. and M.S. degrees in electrical engineering in 1993 and 1995, respectively, from the University of New Mexico at Albuquerque, where he is currently working toward the Ph.D. degree in electrical engineering, researching the role of nanostructures in microelectronic and optoelectronic devices.

A. Frauenglass received the diploma (*magna cum laude*) in laser electrooptics technologies from the Albuquerque Technical Vocational Institute, Albuquerque, NM, in 1984.

He was an Engineering Technician at Hughes Aircraft from 1984 to 1987. From 1987 to 1989, he was a Technical Assistant at the Institute for Accelerator and Plasma Beam Technology, University of New Mexico (UNM), Albuquerque. In 1989, he joined the Center for High Technology Materials at UNM, where he is currently a Research Engineer at Supervisor of the Laser Lithography Research Laboratory.



S. R. J. Brueck (S'63–M'71–SM'89–F'93) was born in New York City in 1944. He received the B.S. degree from Columbia University, NY, in 1965 and the M.S. and Ph.D. degrees from Massachusetts Institute of Technology (MIT), Cambridge, in 1967 and 1971, respectively, all in electrical engineering.

From 1971 to 1985, he was a Member of the Technical Staff at MIT Lincoln Laboratory. In 1985, he moved to the University of New Mexico at Albuquerque, where he is currently a Professor in the EECE and Physics and Astronomy Deparments and Director of the Center for High Technology Materials. His current research interests include nanoscale lithography, the physics of nanostructures, the nonlinear optics of poled glasses, and semiconductor laser physics.

ELSEVIER

Journal of Crystal Growth 240 (2002) 333–339

JOURNAL OF CRYSTAL GROWTH

www.elsevier.com/locate/jcrysgro

Dynamical faceting and nanoscale lateral growth of GaAs by molecular beam epitaxy

S.C. Lee[a,b,*], L.R. Dawson[a,b], S.R.J. Brueck[a,b,c]

[a] *Center for High Technology Materials, The University of New Mexico, 1313 Goddard SE, Albuquerque, NM 87106, USA*
[b] *Department of Electrical and Computer Engineering, The University of New Mexico, 1313 Goddard SE, Albuquerque, NM 87106, USA*
[c] *Department of Physics and Astronomy, The University of New Mexico, 1313 Goddard SE, Albuquerque, NM 87106, USA*

Received 12 November 2001; accepted 19 December 2001
Communicated by A.Y. Cho

Abstract

Dynamical faceting during homoepitaxial growth of GaAs on nanoscale-patterned surfaces by molecular beam epitaxy is examined. Selective deposition on open GaAs(1 0 0) surfaces with lateral dimensions ranging from 130 to 250 nm, separated by 15–80 nm-wide (25-nm-thick) SiO_2 stripes aligned along the [0 $\bar{1}$ 1] direction results in facet formation and lateral growth over the SiO_2 mask. At the early stage of growth, (3 1 1) facets appear on sidewalls near the boundary between an open GaAs surface and SiO_2 mask, these are replaced by (1 1 1) facets starting from SiO_2 boundaries as growth continues. After complete replacement, growth proceeds laterally in the direction perpendicular to [0 $\bar{1}$ 1] retaining the (1 1 1) facets until coalescence occurs between adjacent triangular cross-sectioned GaAs stripes. Nanoscale fabrication nonuniformity results in dynamical formation and retention of multiple (3 1 1) facets even for growth thicknesses much greater than the thickness of the SiO_2 mask stripes. This dynamical faceting is interpreted by minimization of total surface free energy based on equilibrium crystal shape, in qualitative agreement with our experimental results. © 2002 Elsevier Science B.V. All rights reserved.

PACS: 68.55.B

Keywords: A1. Faceting; A2. Lateral growth; A3. Molecular beam epitaxy; A3. Selective epitaxy

1. Introduction

Nanoscale-patterned (nanopatterned) growth is attracting considerable interest because of the possibility of a deterministic alignment of quantum dots (QDs) with improved size uniformity and of migration-assisted selective growth in molecular beam epitaxy (MBE) [1,2]. For 1-dimensional (1-D) patterning, as the lateral dimension of the open substrate surface, d, is decreased to a scale comparable to or smaller than the thickness, t, of an epilayer, faceting will determine the overall shape of a selectively grown epilayer. This faceting provides a powerful new approach to the formation of nanopatterned surfaces that avoids many

*Corresponding author. Center for High Technology Materials, The University of New Mexico, 1313 Goddard SE, Albuquerque, NM 87106, USA.
E-mail address: sclee@chtm.unm.edu (S.C. Lee).

0022-0248/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0022-0248(02)00867-9

of the damage and crystal disruption issues associated with subtractive (etching) preparation of nanopatterned substrates. In this paper, we illustrate this behavior in three regimes for homoepitaxial growth of GaAs by MBE on ~100- to 200-nm-wide GaAs(1 0 0) open faces bounded by thin ~50-nm-wide SiO_2 stripes oriented along a [0 $\bar{1}$ 1] direction: (1) growth within the original open GaAs stripes where the original facets adjust dynamically as the open area is filled for $t \lesssim d$; (2) lateral growth over the SiO_2 stripes for $t \sim d$; and (3) a return to the original faceting for films much thicker than the stripe spacing, $t > d$. According to the basic concept of equilibrium crystal shape (ECS) theory, faceting occurs to minimize the total surface free energy, E [3] (for review of Wulff's construction, see Ref. [4]). Faceting is very important for nanopatterned growth and associated nanoscale lateral growth (NLG) investigated in this work; evolution of the shape and of the faceting in nanoscale growth leads to lateral growth which has not been observed in μm-scale-patterned MBE growth.

We investigate homoepitaxial growth of GaAs with a 1-D SiO_2 stripe pattern atop a GaAs (1 0 0) substrate, which demonstrates the basic idea of ECS without any complications from heteroepitaxial strain. Surface migration of incident atoms is one of major physical origins of faceting in MBE. Boundary faceting occurs near the edge between a SiO_2 mask and an open substrate surface [5]. Microscopically, boundary faceting has been explained by variation of surface-bonding configuration, leading to different surface reconstructions and different surface free energies. As a consequence of the restricted growth area, faceting must evolve dynamically to maintain a minimum surface free energy. Thus, generation and disappearance of various facets, referred as dynamical faceting, occurs during growth. As the growth thickness increases beyond the thickness of the SiO_2 masking film, dynamical faceting will determine lateral or vertical growth mode and the continued evolution of the facet structure.

In this work, we present what is to our knowledge the first experimental evidence of dynamical faceting observed on a 1-D SiO_2 stripe-patterned GaAs substrate with homoepitaxial MBE. We examine NLG on SiO_2-patterned substrates showing the evolution of surface faceting and propose a qualitative interpretation using minimization of total free energy based on ECS.

2. Nanoscale patterning and MBE growth

A 1-D 25-nm-thick SiO_2 stripe pattern oriented along the [0 $\bar{1}$ 1] direction was generated on a GaAs(1 0 0) substrate by large-area interferometric lithography and dry etching. Fig. 1(a) a shows a 45°-tilted scanning electron microscope (SEM) image of a 180-nm period SiO_2-patterned substrate for which d is about 130 nm and w, the SiO_2 stripe width, is about 50 nm. The SiO_2 stripes have an edge roughness of about 5–10 nm. The edge definition in the SEM is due partly to this roughness and partly to the SEM resolution limit. Interferometric lithography and the detailed patterning procedure have been presented elsewhere [2].

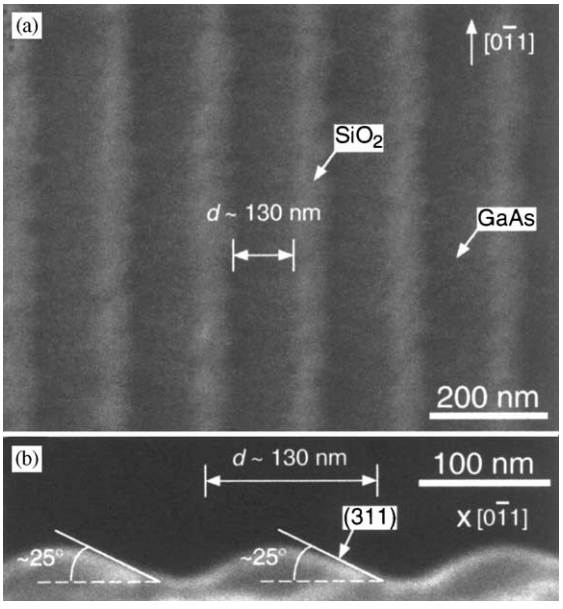


Fig. 1. (a) A 45°-tilted SEM image of a 180-nm period SiO_2 pattern with $d \sim 130$ nm, generated on a GaAs (1 0 0) substrate by IL and dry etching. (b) A cross-sectional SEM image of a 180-nm period ($d \sim 130$ nm) SiO_2-patterned substrate after GaAs growth of $t_d = 15$ nm.

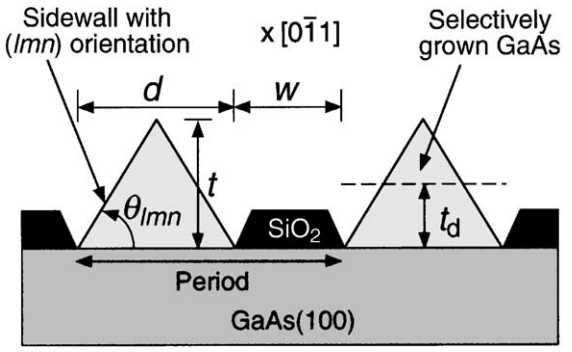


Fig. 2. A schematic illustration of a cross-sectional view of an as-grown sample with the parameters used in this work. SiO_2 stripe masks in every sample were removed by HF treatment after growth.

MBE was performed at 610 ± 5°C with growth rate of about 0.1 monolayers. Under these conditions, the sticking coefficient of a Ga atom on the SiO_2 surface is very low and selective growth is realized [1]. Fig. 2 is a schematic illustration of a cross-sectional view of selectively grown GaAs stripes of which each sidewall consists of a single facet. Deposition thickness, t_d, on a SiO_2-patterned substrate was determined from growth time multiplied by growth rate and means the thickness of GaAs deposited on an unpatterned substrate under the same growth condition, as indicated by the dashed line in Fig. 2. Fig. 1(b) shows a SEM cross-section of a different sample with $d \sim 130$ nm after selective growth of GaAs of about $t_d = 15$ nm. For this and all subsequent SEMs, the SiO_2 has been removed with an HF etch after the growth. For this work, d was varied from 130 to 250 nm by changing the period and the etch parameters, and t_d was varied from 15 to 400 nm.

3. Results and discussion

Fig. 3 shows cross-sectional SEM images taken from a single growth run on a sample with regions of different periods and d's between SiO_2 stripes. In Fig. 3, $t_d = 46$ nm. As expected, several different facet structures were generated depending on d. All of the cross sections show a (1 1 1) facet. For $d = 130$ nm [Fig. 3(a)], the (1 1 1) facets fill the entire space. For the SEMs in Figs. 3(b) and (c),

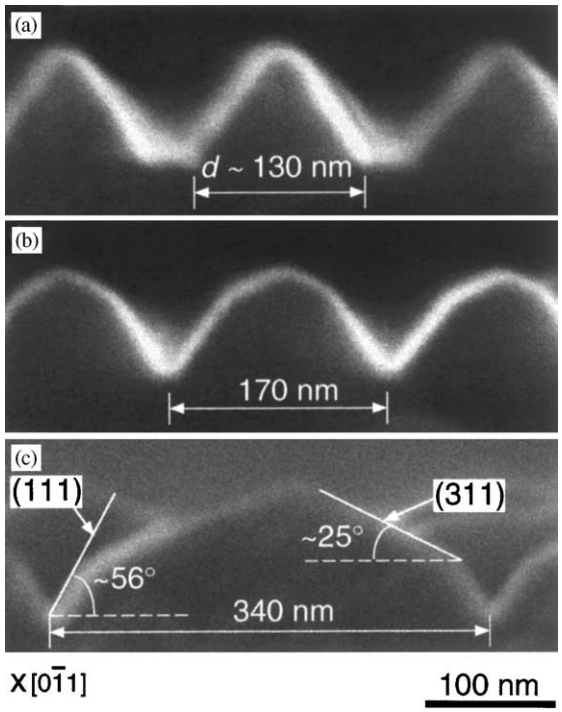


Fig. 3. Cross-sectional SEM images for $d \sim 130$ nm (a), 170 nm (b) and 340 nm (c) SiO_2-patterned substrate after GaAs growth of $t_d = 46$ nm.

both (1 1 1) and (3 1 1) facets are seen, the lengths of the (3 1 1) facets increase with d.

Phenomenologically, the faceting observed in Fig. 3 can be understood as follows. At the initial stage of growth, a (3 1 1) facet begins to appear on an open GaAs surface near the boundary between SiO_2 mask and open substrate surface as seen in Fig. 1(b). Study on the very early stage of facet formation is beyond the scope of this work. For $t_d = 15$ nm, a facet generated on sidewalls near the boundary is (3 1 1)-type. The maximum deposition thickness, t_{d311}, for which this facet is allowed is given by $t_{d,311} = t/2 = d \tan \theta_{lmn}/4$ where $\theta_{lmn} = \theta_{311} \sim 25°$, as indicated in Fig. 2. For $d = 130$ nm, $t_{d,311}$ is about 15 nm as is clear from Fig. 1(b). If d is increased, the allowable thickness to retain a (3 1 1) facet also increases as is evident from Figs. 3(b) and (c). For $d = 340$ nm of Fig. 3(c), $t_{d,311}$ is ~ 40 nm, the pictured case of $t_d \sim 46$ nm, a (1 1 1) facet has just begun to form at the edges of the open GaAs surface. For $d = 170$ nm, Fig. 3(b),

the (1 1 1) facet boundary is closer to the apex of the growth, and finally for $d = 130$ nm, the (3 1 1) facets is gone entirely and the (1 1 1) facets completely fill the open area ($t_d = 46\,\text{nm} \sim t_{d,111} = d \tan\theta_{111}/4$ for $d \sim 130$ nm where $\theta_{111} \sim 56^\circ$). Coexistence of (3 1 1) and (1 1 1) facets on a sub-μm-scale-patterned GaAs(1 0 0) substrate has been reported by Dilger et al. [6].

Initial faceting of Fig. 1(b) is consistent with minimization of E if (3 1 1) facets are more stable or have lower surface free energy than (1 1 1) facets. Although there has been no absolute comparison of surface free energies between different orientations of GaAs, Scheffler and his coworkers have evaluated the surface free energy of several GaAs surfaces with density-functional theory, and suggest that a (3 1 1) surface has a surface free energy comparable with other lower index surfaces [7,8]. According to their calculation, a (3 1 1) facet has surface free energy of 47 meV/Å^2 compared to 51–54 meV/Å^2 for a (1 1 1) facet in an As-rich environment, consistent with our experimental findings. Irrespective of its surface free energy, the initial (1 0 0) facet must be replaced by other facets as growth proceeds in a confined geometry.

We explored the possibility of lateral growth through continued growth over the faceted surface, e.g., for $t_d > t_{d,111} = d \tan\theta_{111}/4$ for a given d. For this purpose, we employed 265-nm period, 25-nm-thick SiO_2 patterns providing better resolution. Fig. 4 presents 45°-tilted [(a) and (b)] and cross-sectional [(c)] SEM images after deposition of $t_d = 200$ nm of GaAs which is more than twice $t_{d,111} \sim 75$ nm for $d \sim 180-220$ nm. Adjacent triangular cross-sectioned GaAs stripes undergo coalescence as the growth proceeds. The coalescence is nonuniform as a result of variations in d and w associated with process limitations at this sub-100-nm scale. If the SiO_2 pattern were uniform, simultaneous overall coalescence might have occurred although the edge definition is intrinsically limited at this nm-scale by the discrete atomic nature of the interfaces. Coalescence over the SiO_2 mask stripe is retarded for a larger w since the sticking coefficient of Ga on the SiO_2 surface is negligible under our growth conditions. Additionally, the extrapolated surface diffusion length of a

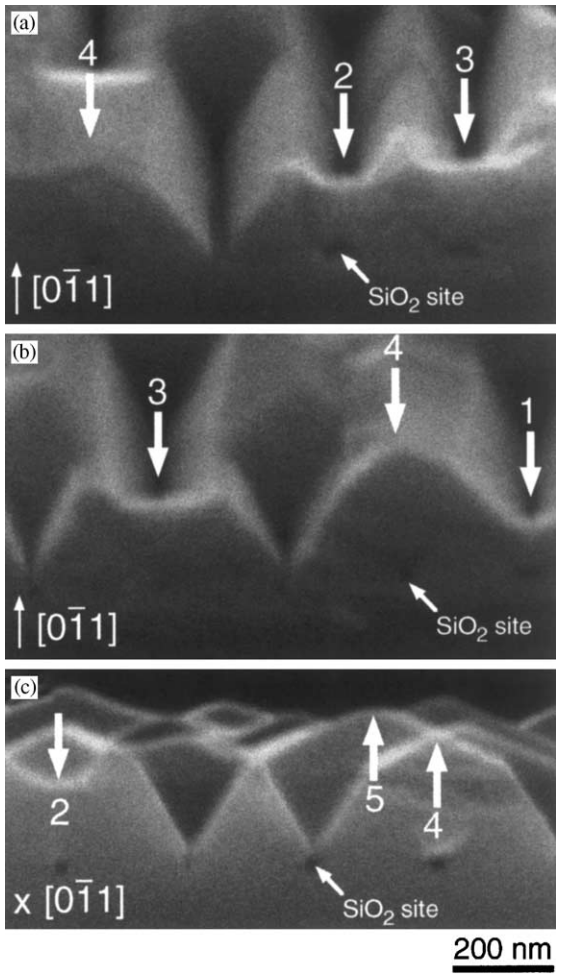


Fig. 4. 45°-tilted [(a) and (b)] and cross-sectional [(c)] SEM images of a 265-nm period SiO_2-patterned substrate after GaAs growth of $t_d = 200$ nm.

Ga atom on GaAs (1 0 0) surface along the [0 $\bar{1}$ 1] direction at 615°C is several μm [9]. Also, it has been reported that the surface diffusion length on (1 1 1) surface is greater than that on (1 0 0) surface [10]. This implies that once initial coalescence occurs at some point over SiO_2 mask which is generally around a small-w region, growth partly proceeds along the [0 $\bar{1}$ 1] direction, onto the plane of the cross sections, as well as filling in the V-groove between the (1 1 1) faces (see below). This and other nonlinear growth dynamics appear to amplify the effects of the differences in w and might account for the strong nonuniformities in

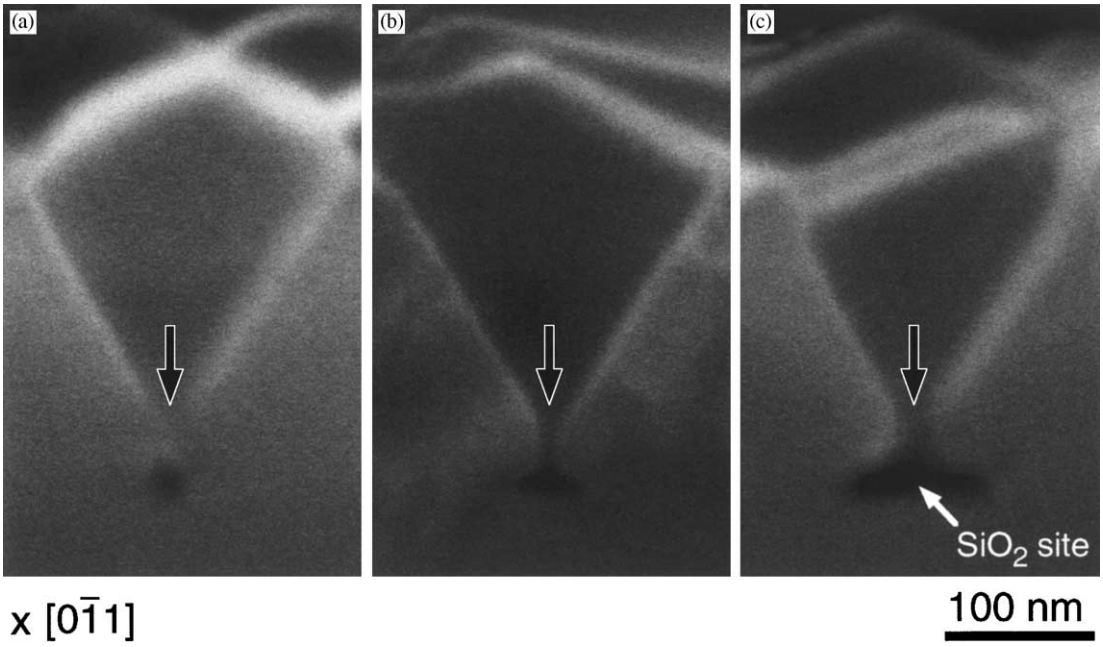


Fig. 5. Cross-sectional SEM images of a 265-nm period SiO_2-patterned substrate having different w ($\sim$15–80 nm) after GaAs growth of $t_d = 200$ nm.

the original SiO_2 mask. The coalescence shown in Fig. 4 results in growth of single crystalline GaAs over the SiO_2 mask. The clear faceting of the surface along with the absence of any ring patterns in in situ reflection high-energy electron diffraction shows that no polycrystalline phase occurs over the entire growth plane.

Magnified views of NLG and coalescence regions are shown in Fig. 5 for different w values ranging from $\sim$15 to 80 nm. These images correspond to different regions of the sample prepared with differing etch times to vary w, but with a single growth to ensure identical growth conditions. Coalescence depends both on w and on the oxide thickness. In Fig. 5(a), coalescence is evident for $w \sim 15$ nm, whereas a thicker deposition is required for $w \sim 85$ nm in Fig. 5(c), which clearly exhibits NLG. In this figure, growth proceeds in lateral direction, over the SiO_2 mask, and retaining (1 1 1) sidewall facets. It is not clear whether some particular facets were generated along the sidewalls of SiO_2 mask in Fig. 5. One of most probable facets is (0 1 1)-type perpendicular to (1 0 0) if SiO_2 edge has vertical sidewalls. The filling process of Fig. 4 is then continued until the height of the (0 1 1)-type facets reaches the thickness of the SiO_2 mask. To minimize E, growth is continued with extension of (1 1 1) facets over SiO_2 mask leading to lateral growth.

More investigation of the details of the growth up to the thickness of SiO_2 mask is clearly necessary. In particular, the mask used in this work is only 20–30 nm thick and does not have vertical sidewalls. Regardless of its thickness, it is evident that the presence of SiO_2 mask results in discontinuity of bonding sites available for epitaxial growth and causes facet formation on adjacent open GaAs surface. However, these slanted SiO_2 sidewalls may allow extension of the (1 1 1) facets over SiO_2 mask. In selective growth, initiation of over-growth also requires a certain activation energy which may depend on the slanted degree of SiO_2 sidewalls. Growth without lateral expansion after filling out open GaAs (1 0 0) surface has been reported in circular-patterned selective growth that does not provide any particular facet

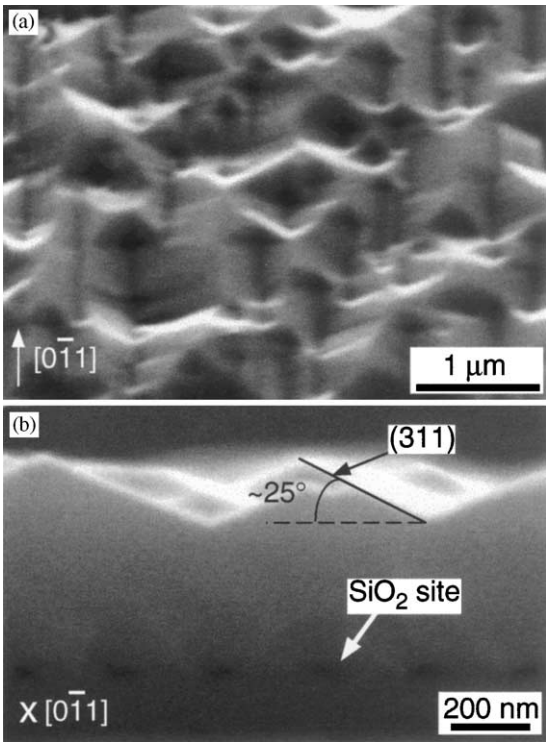


Fig. 6. 45°-tilted (a) and cross-section (b) SEM images of a 265-nm period SiO_2-patterned substrate after GaAs growth of $t_d = 400$ nm.

in sidewall [2]. If extension of existing facets over SiO_2 mask is energetically more favorable than creation of the next-highest-surface-free-energy facet, epitaxial growth would keep its minimum E by facet extension.[1] This interpretation can be applied to the replacement of (3 1 1) by (1 1 1) of Fig. 2. Lateral growth is then accompanied by over-growth filling out the space bounded by open GaAs (1 0 0) surface and slanted sidewalls of SiO_2 mask. Experimentally, we observe lateral growth of GaAs over the SiO_2 mask through extension of (1 1 1)-type facets leading to coalescence of adjacent growth stripes. The field-emission SEM used in this work does not have sufficient resolution to allow investigation of the fine details of the region along the SiO_2 sidewall.

What happens as growth is continued further after coalescence? For the SEM images of Fig. 6(b) taken after a deposition of $t_d = 400$ nm on $d = 180-220$ nm SiO_2-patterned GaAs (1 0 0) substrate, all of the SiO_2 stripes are buried in deposited GaAs. The top surface shown in Fig. 6 is not flat, but mainly consists of ~(3 1 1) facets just as in the initial stage of facet formation of Fig. 1(b). Fig. 4 gives some clues about the reasons for this (3 1 1) terminated surface after complete coalescence. In Fig. 4, the numbered arrows (1–5) indicate the sequential stages of evolution after coalescence. Owing to the fluctuation in w, several different evolution stages can be observed from a single sample. The evolution order is from stage 1 to 5 with the coalescence point shown in Fig. 4(b) (1) as the starting point. Once coalescence occurs between adjacent (1 1 1) type facets, the triangular dip in the cross section starts to fill in. We speculate that the steep junction of the two (1 1 1) faces at the coalescence point will form a rounded concave profile with a radius of high curvature at the very initial stage. A rounded profile with high curvature may be energetically unstable compared with a flat one since such a surface does not have a well-defined surface reconstruction. Then, some portion of subsequently arriving Ga atoms which have a surface migration length much greater than the lateral dimension of the (1 1 1) facets move to this area to reduce surface free energy and total area, and as a result gradually fill out this triangular dip. The rounded surface of stage 2 of Figs. 4(a) and (c) are examples of this filling process. A well-known example similar to this filling is growth of a quantum wire in a V-groove. As seen in stage 4 of Figs. 4(a) and (b), however, this filling process eventually results in new (3 1 1)-type facets on the top surface. This also could be due to the fluctuation of w in SiO_2 pattern, although as mentioned above, nonlinearities in the growth process seem to accentuate the roughness compared with the original pattern. Since the degree of filling between adjacent triangular dips is different because of pattern nonuniformity, dips will fill [e.g. form flat (1 0 0) surfaces] at different growth times.

[1] Facet-extension accompanies generation of other facets along the interface between deposited GaAs and SiO_2 mask. Besides, kink energies due to facet intersections must be also considered for rigorous approach. More detailed analysis is presently underway and will be reported elsewhere.

As the growth is continued further, these area-limited (1 0 0) surfaces will transform into (3 1 1)-type surfaces just as observed in Fig. 1(b). Unlike Fig. 1(b), however, the transformed (3 1 1)-type surfaces of Fig. 4 do not have nearby SiO_2 mask stripes anymore. Depending on the status of neighboring ridges, these (3 1 1)-type surfaces can grow larger than the original d as indicated by stage 5 in Fig. 4(c). Thus, (3 1 1) facets appear first at the earliest growth stages, are replaced by (1 1 1) facets as the growth proceeds until coalescence, and finally reappear on the top surface as the growth is continued.

4. Summary and conclusion

We have examined dynamical faceting in nanoscale homoepitaxial lateral growth on GaAs (1 0 0) by molecular beam epitaxy. SiO_2 patterns prepared by interferometric lithography and dry etching were used to define open GaAs stripe surfaces with a lateral dimension ranging from 130 to 250 nm along [0 $\bar{1}$ 1] directions. Selective growth resulted in facet formation and lateral growth over 15–80 nm-wide and 25-nm-thick SiO_2 mask depending on the amount of deposition. At the early stage of growth, (3 1 1) facets appear and are gradually replaced by (1 1 1) facets as growth is continued in these restricted areas. When (1 1 1) facets are fully developed as sidewalls of a GaAs stripe, the growth proceeds in lateral direction perpendicular to [0 $\bar{1}$ 1] until coalescence between adjacent triangular cross-sectioned GaAs stripes occurs. Dynamical faceting was interpreted with minimization of total surface free energy based on equilibrium crystal shape theory which qualitatively agreed with our experimental results.

Acknowledgements

The authors wish to acknowledge the support of DARPA and of ARO/MURI.

References

[1] S.C. Lee, L.R. Dawson, K.J. Malloy, S.R.J. Brueck, Appl. Phys. Lett. 79 (2001) 2630.
[2] S.C. Lee, K.J. Malloy, S.R.J. Brueck, J. Appl. Phys. 90 (2001) 4163.
[3] G. Wulff, Z. Kristallogr. Mineral. 34 (1901) 449.
[4] A. Pimpinelli, J. Villain, Physics of Crystal Growth, Cambridge, 1998.
[5] W.T. Tsang, A.Y. Cho, Appl. Phys. Lett. 30 (1977) 293.
[6] M. Dilger, M. Hohenstein, F. Phillipp, K. Eberl, A. Kurtenbach, P. Grambow, A. Lehmman, D. Heitmann, K. von Kilizing, Semicond. Sci. Technol. 9 (1994) 2258.
[7] N. Moll, A. Kley, E. Pehlke, M. Scheffler, Phys. Rev. B 54 (1996) 8844.
[8] J. Platen, A. Kley, C. Setzer, K. Jacobi, P. Ruggerone, M. Scheffler, J. Appl. Phys. 85 (1999) 3597.
[9] M. Hata, T. Isu, A. Watanabe, Y. Kytayama, J. Vac. Sci. Technol. B 8 (1990) 692.
[10] T. Takebe, M. Fujii, T. Yamamoto, K. Fujita, T. Watanabe, J. Appl. Phys. 81 (1997) 7273.

Nanoscale limited area growth of InAs islands on GaAs(001) by molecular beam epitaxy

Seung-Chang Lee[a], Andreas Stintz[b], and Steven R. J. Brueck[a,b]
Center for High Technology Materials, University of New Mexico, 1313 Goddard, SE, Albuquerque, NM 87106

Growth of InAs islands on a GaAs(001) substrate patterned with ~ 50- to 200-nm diameter holes in an SiO_2 mask overlayer providing selective GaAs nucleation areas is reported. The nanoscale pattern was generated in the SiO_2 film by large-area interferometric lithography and dry etching. Two-dimensional, 285-nm period, arrays of InAs islands having heights of 10- to 15-nm with three different bottom diameters of 50- to 100-nm, ~150 nm, and ~200 nm, were selectively grown on SiO_2 patterned substrates by molecular beam epitaxy. Growth conditions were chosen to provide a very low sticking coefficient of In atoms on the SiO_2 surface excluding any volume contribution from migration of In atoms incident on SiO_2 mask region to nearby open GaAs surface areas. Formation of spherical-section InAs dots with diameters of about 50 nm entirely relying on nanoscale-limited area growth is demonstrated. As the diameter of hole increases beyond 150 nm, InAs islands deviate from a spherical-section and self-assembled QDs confined within the open GaAs surface appear. A relation between activation of self-assembly and the nanoscale growth area is proposed, with an apparent transition from layer-by layer growth to self-assembly occuring at hole diameters of ~ 100- to 150-nm.

I. Introduction

Self-assembled QDs have significant potential for many device applications. In many cases, additional control of the size and positional distribution of the dots over that possible so far with self-assembled growth would be desirable. Recently, there have been several attempts to control the nucleation of self-assembled quantum dots (QDs) with patterned substrates[1]. Most of these have concentrated on growth of InAs or $In_xGa_{1-x}As$ self-assembled QDs on a GaAs or a InP substrate using optical or electron-beam lithography for patterning either an overlayer mask or the substrate[2-6] or have exploited facet-dependent migration.[2, 7-13]

Recent progress in lithographic technologies has expanded the regime of accessible mask/substrate patterning into the nanoscale. A pattern for which the scale approaches the lateral dimension of typical self-assembled QDs strongly suggests the capability to grow QDs on a limited area of a substrate with better controllability in both size and spatial distribution. This so-called nanoscale patterned growth basically requires a highly uniform pattern and a selective growth mode.

We have previously reported[14] on nanoscale, homoepitaxial molecular beam epitaxy (MBE) selective growth of GaAs involving migration-enhancement by surface diffusion of incident Ga atoms from the masked areas onto the growth areas. Recently, formation of InAs islands on a Ga_2O_3- or SiO_2 -patterned substrates[3,6] using similar In-atom migration processes have been reported. However, in these experiments the lateral size of the open substrate surface is far greater than the lateral dimension of typical self-assembled InAs QDs.

InAs island formation on a patterned substrate is basically different from self-assembled QD formation on a unpatterned substrate. First, mask-patterned growth does not provide an extensive wetting layer for initiation of the Stranski-Krastanov (S-K) growth mode. Self-assembly is a relaxation process for reduction of strain energy. This suggests that the migration of an In atom on a GaAs surface becomes a directional movement for strain reduction once nucleation of QDs begins. However, migration of an In atom on a SiO_2 surface is irrelevant to strain relaxation. This random movement provides a chance for In atoms to be accidentally captured by nearby InAs island being grown on a open substrate surface, or ends up with desorption from the SiO_2 surface during migration. As a pattern period decreases, this accidental capturing due to migration becomes more significant. In spite of deposition of very thin InAs layer of about ~ 2 monolayers (MLs), accidental capturing makes it possible to form islands or QDs in reported patterned growths.[3,6] We refer to this mode of InAs island (or QD) formation on a patterned substrate as migration-assisted QD formation.

Recently, the role of a wetting layer in the activation of S-K growth mode,[16,17] the control of shape and size[18] and the carrier transport[19,20] between QDs on unpatterned substrates have been investigated. However, the impact of the lateral area of the wetting layer on SK mode growth has not been clearly established. This is another subject which must be investigated in nanoscale patterned growth. In this work, we examine the growth mode for formation of InAs QDs on

a. Also Department of Electrical and Computer Engineering.
b. Also Department of Physics and Astronomy.

a nanoscale SiO_2-patterned substrate in a new regime. We choose growth conditions where the sticking coefficient of an In atom on a SiO_2 surface is essentially zero as evidenced by the absence of any detectable growth on an unpatterned SiO_2 film. Then, formation of InAs islands is only due to the incoming flux to a open surface of a GaAs substrate bounded by SiO_2 mask. This implies that accidental capture due to In migration along the SiO_2 surface is minimized since essentially all of the incident In atoms are rapidly desorbed from the SiO_2 surface. Unselective growth, for which the surface diffusion length along the mask is less than the QD size, using lift-off processes also results in a limited area deposition and does not allow accidental capture of itinerant In atoms. In a broad sense, we refer to these QD formation mechanisms without a large-area wetting layer and without growth contributions from atoms incident on the masking layer in the nanoscale regime as nanoscale limited area growth (NLAG). The focus of this work is to examine QD formation by MBE in nanoscale limited areas, the ultimate goal of NLAG will be to customize the QD shape, size, spacing, density and area coverage.

We examine the formation of InAs islands on a nanoscale limited area of GaAs(001) substrate. A 285-nm period, 2-D array of circular holes with diameters ranging from $\lesssim$ 100 nm to 200 nm was generated through a SiO_2 film on GaAs. This SiO_2-patterned substrate was prepared by large-area interferometric lithography and dry etching. InAs islands were grown by MBE. Different growth modes, indicating the nanoscale regime for which self-assembly (S-K growth) becomes important, were found across this range of hole diameters in the SiO_2. We propose a relationship between self-assembly and the area of any wetting layer in the S-K growth mode.

II. Nanoscale patterning

Reproducible nanoscale patterning is essential to realizing NLAG. For pattern formation, we used large-area IL which is very promising as a low-cost, reliable, and scalable technology for deep sub-100-nm scales. The structure of the sample for processing consists of a negative-tone photoresist (PR) film, an antireflection coating (ARC), a SiO_2 film, and a GaAs(001) substrate. A 120-nm thick ARC film, with strong absorption at the IL wavelength, was embedded between the PR and the SiO_2 film to suppress vertical standing wave effects. A 80-nm thick SiO_2 film deposited by electron-beam evaporation was used as the mask material. A single longitudinal and transverse mode, frequency tripled YAG laser at a wavelength of 355 nm was used as an ultraviolet light source. Using IL, a 285-nm period interference pattern was generated on the PR film. Figure 1(a) shows a 45°-tilted scanning electron microscope (SEM) image of a developed PR pattern (all SEM images presented herein are 45°-tilted). As seen in the figure, a 2-D array of circular holes of diameter ~100 nm is generated on the PR film. The thickness of the PR film after development was 140-160 nm. The detailed process for 2-D array of hole patterns by IL has been summarized elsewhere.[15] Figure 1(b) is a SEM image revealing the sample film stack.

Transfer of the PR pattern into the mask material without degradation is a very important step for NLAG. Dry etching with CF_4 plasma was used for pattern transfer. Figure 2 shows SEM images of a 2-D array of holes transferred onto the SiO_2 film through the ARC material. Both SiO_2, ARC, and PR films can be etched by CF_4 plasma. The ARC has an etch rate that is about 50% that of the SiO_2 film while it has almost same etch rate as the PR film. These three materials do not have good selectivities in CF_4 plasma etching, but the film thicknesses are nonetheless sufficient to allow good pattern transfer. As seen in Fig. 2 (a), the transferred pattern is quite uniform. Also, note that the hole diameter, d, in the SiO_2 film of Fig. 2 (b) is less than the ~100-nm diameter of the starting PR film of Fig. 1. This is due to the formation of slanted sidewalls for the holes in the SiO_2 film during the etch process. The range of hole diameters in Fig. 2 (b) is about 50- to 80-nm. To our knowledge, this represents the smallest area substrate opening ever explored for 2-D patterned growth. These sub-100-nm diameter holes provide incorporation routes for In atoms to the substrate surface for epitaxial growth. By varying the etching time, different d's ranging up to 200 nm could be produced. For this work, three different hole diameters were prepared: 50- to 80-nm, ~150 nm, and ~200 nm.

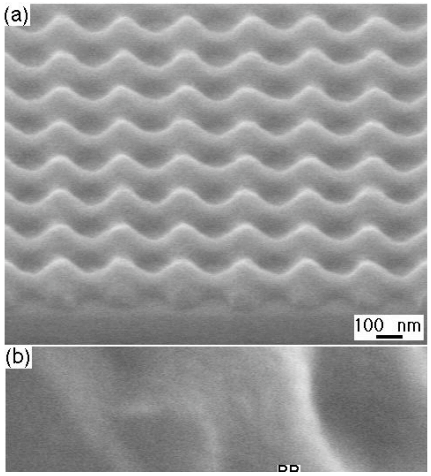


Figure 1 A 45°-tilted SEM image of (a) a 2-D array of holes generated in a negative PR film and (b) the sample film structure.

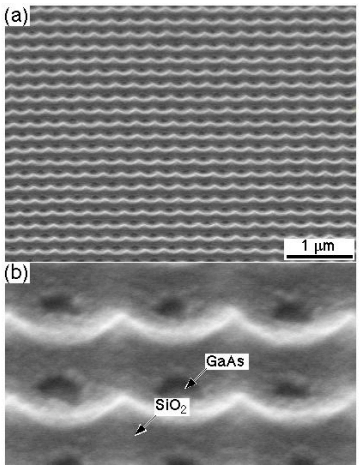


Figure 2 (a) A 45°-tilted SEM image of a 2-D array of holes transferred through the SiO_2 film by dry etching and (b) a higher magnification SEM image.

All of the InAs QDs and islands presented in this work including the InAs layer grown on an unpatterned substrate were obtained from a single growth run on a single substrate which was treated with different dry etch times using a mechanical mask to obtain the various hole diameters. Thus, all of the results were obtained under exactly the same growth conditions with the same In flux.

III. MBE growth

To remove any surface contamination and assure good crystal quality at the growth surface, about 5-nm of GaAs was removed from the open surface of the substrate by wet etching using a highly diluted etchant (H_2SO_4 : H_2O_2 = 1:4) having etch rate of 2.0 - 2.5 nm/s for 2 seconds immediately before loading the sample into the MBE chamber. A small part of the SiO_2 pattern on the substrate was completely removed so that surface oxide removal from the substrate before InAs deposition could be confirmed through the RHEED pattern. The surface oxide removal step was performed before the start of growth at 620°C for more than 10 minutes.

Approximately 10 nm of InAs was then deposited by MBE on the SiO_2-patterned substrate shown in Fig.2 with the flux ratio of As to In set to 10 in beam equivalent pressure. This thickness is about twice the typical height of self-assembled InAs QDs on an unpatterned GaAs substrate. The growth temperature was maintained at 535°C, where desorption of In atoms from GaAs(001) already begins.[21] The reported incorporation fraction of an In atom at 547°C is about 0.94 for an $In_{0.15}Ga_{0.85}As$ surface which is negligible,[22] but could be higher than this value in case of InAs surface.[23] High quality InAs QDs have been generated under these growth conditions.[1] The growth temperature was monitored by optical pyrometry. The flux of In atoms was controlled to achieve a growth rate of 0.1 ML/s which was separately calibrated by reflection high energy electron diffraction (RHEED) oscillations. By examining growth on unpatterned SiO_2 films, we found that the sticking coefficient of an In atom on an unpatterned SiO_2 film is significantly reduced around 510°C and becomes unobservable at 535°C under these flux conditions. Therefore, these growth conditions enable growth of InAs QDs without any contributions from migration-assisted QD formation.

IV. Results and discussion

Figure 3 shows SEM images of the as-grown sample (a) and of a section of the same sample after an HF-etch to remove the SiO_2 mask (b) for the smallest diameter hole array. Comparing Fig. 2 and Fig. 3(a), no growth is observed on the SiO_2 film. During growth, no ring patterns appeared in the *in situ* RHEED, implying that polycrystalline InAs was not formed on the SiO_2 surface. These results demonstrate that selective growth occurred in this heteroepitaxy. In this selective growth regime, dipping the sample in diluted-HF completely removes the SiO_2 mask as seen in Fig. 3(b).

The bottom diameter of the InAs island shown in Fig. 3 is in the range of 50- to 100-nm close to those of the holes

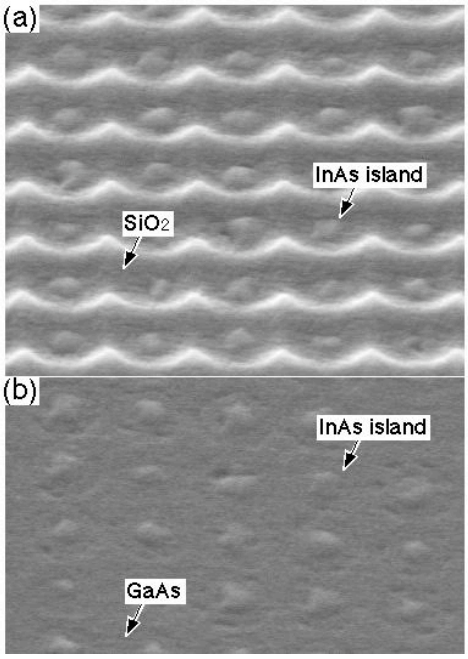


Figure 3 45°-tilted SEM images of a 2-D array of InAs islands deposited on the SiO_2-patterned substrate (a) as-grown state and (b) after HF treatment to strip the SiO_2.

of Fig. 2(b) originally patterned on the SiO_2 mask. Although the SiO_2 pattern in Fig. 2 looks uniform, the InAs islands of Fig. 3 have a noticeable fluctuation in size and shape. Generally, small-diameter holes, nearer the limits of our fabrication techniques, have larger size fluctuations. As shown in Fig. 2(b), this fluctuation is partly due to irregularities in the hole pattern. The general shape of holes in Fig. 2(b) is close to a circle but with a rough periphery. Since deposited InAs was only about 10 nm in thickness, no faceting was observed on the surface of InAs islands. As a result, the overall shape of the InAs islands seems to depend on the shape of holes allowing epitaxial growth nucleation. In Fig. 3, InAs islands having $d < 80$ nm have a shape similar to a spherical section. The detailed island shape will be discussed below.

Figure 4 shows a SEM image of a single as-grown InAs dot for which the lateral dimension is ~60 nm. It is known that the average lateral dimension of InAs QDs increases with growth temperature and becomes greater than 30 nm at 535°C.[1] Although there remain issues as to the precise shape and size of self-assembled QDs,[24-26] the dot in Fig. 4 is within about a factor of two in linear dimension compared with typical self-assembled QDs but, it should be emphasized, any wetting layer does not extend beyond the lateral dimensions of the dot. The InAs dot of Fig. 4 is thus equivalent to a QD grown in an artificial Volmer-Weber (V-W) growth mode. Without S-K mode, therefore, growth of InAs QDs having diameter of 60 nm only relying on NLAG is demonstrated.

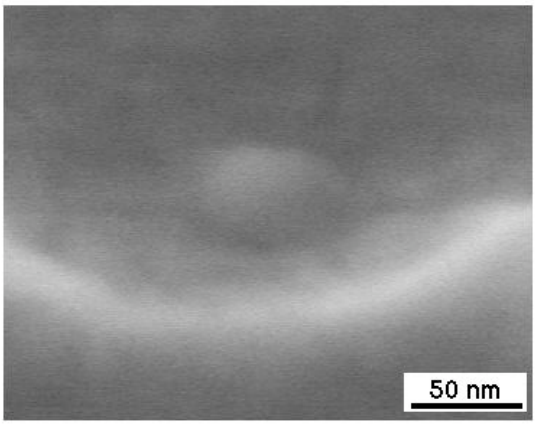


Figure 4 A 45°-tilted SEM image of an as-grown InAs dot of $d \sim 50$ nm.

Recently, there have been two reports of InAs and InGaAs QD formation without a wetting layer. Mano *et al*. reported InGaAs QDs on a GaAs substrate using alternative deposition of In and Ga droplets by low temperature MBE.[27] Tsatsul'nikov *et al*. reported V-W mode growth of InAs-(Al,Ga)As quantum dots with an ultrathin AlAs capping layer.[28] In both cases, unpatterned substrates were used but no wetting layers were formed[27] or remained[28] after growth of a capping layer. One of the common results is a change of QD shape. The QD height was in the range of about 8 - 12 nm which is greater than the 4- to 6-nm typical of self-assembled QDs on an unpatterned substrate[1] while the lateral dimensions are in different ranges (30 nm in Ref. 27 and 7- to 9-nm in Ref. 28). They confirmed the absence of dislocations on their QDs with photoluminescence and transmission electron microscopy. These results imply that the absence of wetting layers allows a different shape of QDs, taller than self-assembled QDs, on an unpatterned substrate. Although the physical origins of this QD shaping have not been explicitly clarified, these reports suggest that the presence of wetting layers is very important in QD formation by S-K mode, as expected.

In consideration of volume comparison between the InAs dot of Fig. 4 and typical self-assembled QDs with and without a wetting layer, the InAs dot might be strain relaxed because its lateral size is about twice that of self-assembled QDs. It has been reported that self-assembled InAs QDs relaxed by an excess supply of In atoms still keep a similar shape to unrelaxed QDs except for their larger size. Even if it is relaxed, minimization of surface free energy would lead to formation of a spherical shape for an InAs island as observed in Fig. 3 and 4.[29,30]

As previously mentioned, different etching times for pattern transfer resulted in a variation of hole diameter. Figure 5 shows SEM images obtained from InAs islands having $d < 100$ nm [(a) and (b)], ~ 150 nm [(c)], ~ 200 nm [(d), (e) and (f)], and an InAs layer grown on an unpatterned substrate region [(g)]. Figures 6 and 7 are atomic force microscope (AFM) images of InAs islands and their corresponding profiles which have similar diameters to those of Fig. 5, respectively. Figure 7 represents the profiles obtained along the direction of the line (A to B) indicated in Fig. 7(f). Dotted lines in Fig. 7 are the reference height corresponding to the GaAs substrate surface.

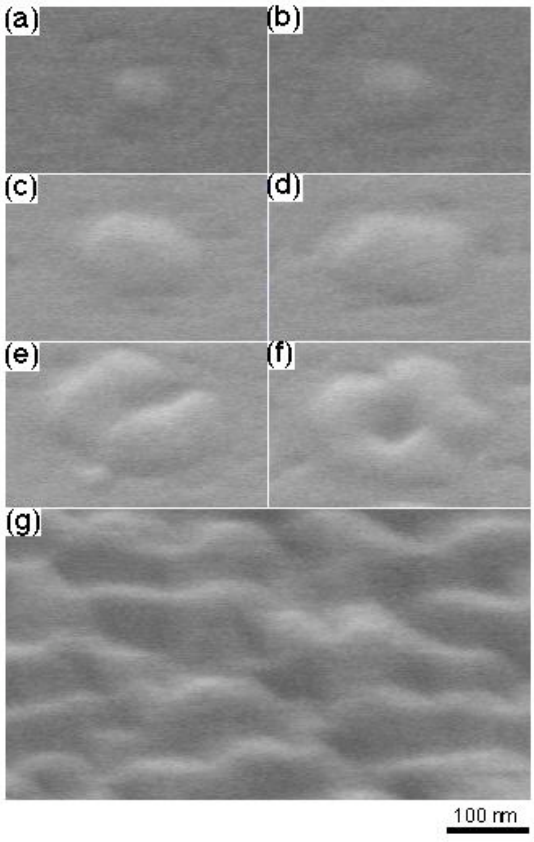


Figure 5 45°-tilted SEM images of InAs islands deposited on holes with a diameter of 50- to 100-nm [(a) and (b)], ~ 150 nm [(c)], ~ 200 nm [(d), (e) and (f)], and an InAs layer grown on an unpatterned substrate region [(g)].

As seen in Figs. 5 and 6, the InAs islands have a circular shape in 45°-tilted and top view when d is less than or comparable to 150 nm. Fig 7 (a-c) shows the shape and height of these islands which have a spherical-section shape with a height, h, of about 10- to 15-nm. This height is close to that reported in Ref. 27 and 28. The spherical-section shape can be understood since In atoms incident near the boundary between GaAs and SiO_2 surfaces move to an inner side of a hole to find a more stable adsorption site.

Although the physical dimension of these dot-shaped InAs islands or 'big dots' is not highly uniform, their heights are 2- to 3-times larger than those of self-assembled InAs QDs on GaAs(001). Also, the height of an InAs big dot is at most 50% greater than the thickness expected from the In flux and the deposition time. Assuming an ellipsoidal section shape, however, the extra volume increment is almost zero compared with the volume of a cylindrical-shaped island with the same bottom area and the nominal growth height. These results imply that there is no additional volume contribution due to In atoms migrating from nearby SiO_2 regions to the open GaAs surface, counter to the situation in migration-assisted QD formation. The size-independent height shown in Fig. 7 is further evidence of

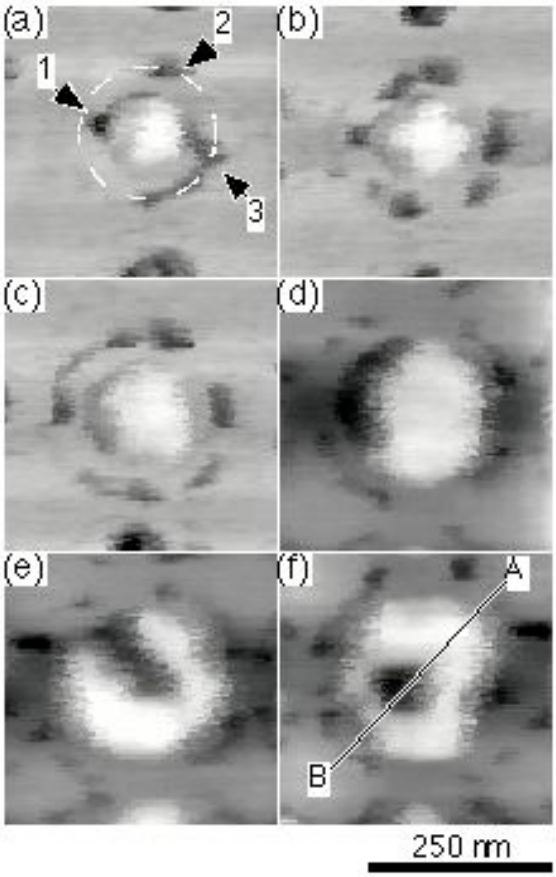


Figure 6 AFM images of InAs islands deposited on holes with a diameter of 50- to 100-nm [(a) and (b)], ~ 150 nm [(c)], and ~ 200 nm [(d), (e) and (f)], similar to Fig. 5.

the absence of contributions to the growth from In-atoms migrating from the surrounding SiO_2 material. The very low sticking coefficient of an In atom on a SiO_2 surface at the given growth conditions, implying extremely short surface lifetime, therefore, excludes migration-assisted QD formation, as predicted.

Around each InAs island, a circular step along with several small dips indicated by a white dashed circle and black arrows in Fig. 6(a), are observed in all AFM images of Fig. 6. Although such steps and dips were already observed in the SEM results (Fig. 3), they appear much more clearly in the AFM images. The depth of these dips is in the range of 5- to 9-nm. While the circular steps resulted from the wet etching of the substrate which was performed before loading into the MBE chamber, their depth was not precisely measured but is less than that of the small dips around them. The dips [the numbered black spots in Fig. 6(a)] were generated either near the periphery of an InAs island (1) or about 40 - 60 nm away from it (2 and 3) and are independent of the size of the InAs islands. Once a dip is generated on the periphery, no additional dips form in the same direction from the center of an island. They are likely due to nonuniform wet etching. A GaAs surface interfaced with a SiO_2 film can have an etch rate faster than the bulk as a result of stress at the interface and capillary effect, and can form a circular step with a diameter greater than the hole size in the SiO_2. Moreover, some wet etchant trapped near the boundary between the SiO_2 film and the GaAs surface might not be removed rapidly in the deionized water rinse used to quench the etch. Such etchant trapping could produce the dips shown in Fig. 6. Thus, reliable surface treatment for nanoscale open substrate surfaces is another important issue to be dealt with in NLAG.

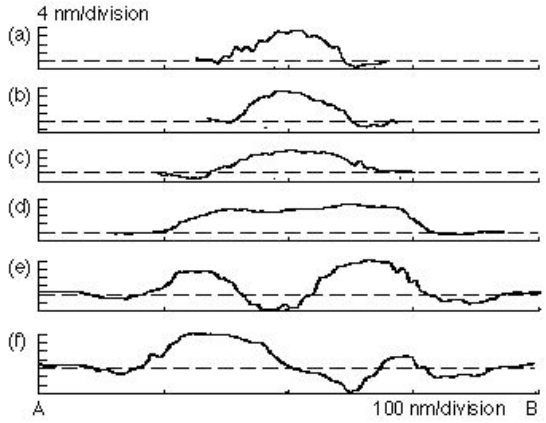


Figure 7 AFM profiles of InAs islands of Fig. 6 obtained along the direction of the line (AB) indicated in Fig. 7(f). Dotted lines are the reference height corresponding to the GaAs substrate surface.

In Fig. 5 and 6, the shape of the InAs islands deforms from a spherical-section when d exceeds ~150 nm. For example, the InAs island of Fig. 6(d) of which the top view is close to a circle has a deformed spherical-section profile, as seen in Fig. 7(d). If d is increased further, as shown in Fig.5(e) and (f) and Fig. 6(e) and (f), the deposited InAs does not retain a circular shape but rather forms a random-shaped island within each hole. Fig. 7(e) and (f) reveal strongly distorted profiles. However, their shape is still different from that of an InAs layer deposited on an unpatterned region which exhibits surface undulations for the reduction of strain energy as shown in Fig 5(g). As d increases beyond 150 nm, a self-assembling mechanism seems to be activated. The ring-type InAs island of Fig. 5(f) is similar to a coalesced form of several self-assembled QDs. Since the deposited InAs was 10-nm thick, formation of self-assembled QDs and their coalescence is very feasible in wide holes. From Fig. 5, 6, and 7, it can be deduced that the area allowed for nucleation of InAs on GaAs(100) surface strongly effects the growth mode, especially the island shape and morphology. This variation is mainly due to lattice-mismatch-induced stress effects which increase with deposition area.

As indicated by arrows in Fig. 8 (a), very small-sized InAs dots were occasionally formed alongside the larger InAs islands. Approximately 15% to 20% of the InAs islands with $d \gtrsim 150$ nm exhibit such small dots. These small dots were not observed when $d \lesssim 100$ nm. The diameter of these small dots is about 30 - 40 nm which is comparable to the lateral dimension of typical[1] self-assembled InAs QDs grown at 535°C. In Fig. 8(b), the height of the small dot is about 4 nm which is also very close to that of typical self-assembled InAs QDs while other large-sized dots in Fig. 8(b) are more than 8-nm high. The large-sized dots in Fig. 8(b) are consistent with the 10-nm deposition. Therefore, we argue that the small dots are evidence of S-K growth mode self-assembly for open hole areas with diameters $d \gtrsim 150$ nm on a GaAs(001) surface bounded by an SiO_2 mask.

As mentioned above, the shape of the InAs islands for $d \sim 200$ nm in Fig 5(f) could be due to coalescence between self-assembled QDs during growth. More than 20% of the InAs islands grown in holes with $d \gtrsim 150$ nm then might have self-assembled QDs at least in coalesced form. Formation of self-assembled QDs on those holes can be understood by large open areas which, while still limited, are much greater than the size of self-assembled QDs. Strain energy depends on area as well as thickness.[31,32] At the early stages of growth, a few monolayer-thick strained InAs film deposited on such a relatively large area could have sufficient strain energy to activate the S-K growth mode. In Fig. 8(b), more than one InAs dot appears within a single growth area. If multiple self-assembled InAs QDs can be generated on a limited area for $d \gtrsim 150$ nm, this also indirectly supports the possibility of the coalescence between self-assembled QDs for larger growth areas. Again, it must be emphasized that the self-assembled InAs QDs shown in Fig. 8 were not observed for $d \lesssim 100$ nm (Fig. 3). Coalescence between dots on such small holes is less probable because of the spherical-section shape and the comparable height of InAs islands to those of QDs without a wetting layer.[27,28] It is evident that only a single dot can be generated within a single hole as the hole diameter approaches the lateral dimension of a self-assembled QD. Because of insufficient strain energy to activate self-assembling due to a small, finite-area wetting layer, a single dot-shaped layer-by-layer growth may proceed on the holes of $d \lesssim 100$ nm for a 10-nm-thick InAs deposition.

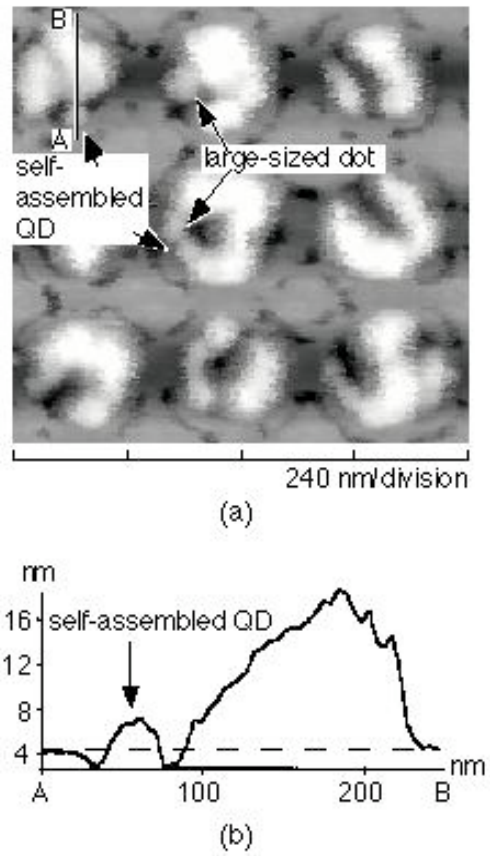


Figure 8 (a) AFM image revealing formation of self-assembled QDs and large-sized dots, and (b) an AFM profile of a self-assembled QD.

These results imply a criterion on the critical diameter of holes, d_C, for generation of self-assembled QDs. In this work, 100 nm < d_C < 150 nm at 535°C at a growth rate of 0.1 ML/s. This d_C which is about 5- to 10-times the lateral dimension of self-assembled InAs QDs (20- to 30-nm**???**), provides a boundary between layer-by-layer growth and S-K mode growth on a large-area wetting layer. In other words, d_C is the minimum diameter for activation of S-K growth mode self-assembly. The reported density of self-assembled InAs QD grown on an unpatterned GaAs substrate at 535°C is about 1.5×10^{10} cm^{-2}.[1,30] Thus, the average area allocated to a single QD is about 80×80 nm^2, comparable to d_C. Further study and theoretical modeling will be necessary to understand the role of the wetting layer and to understand the meaning of d_C in S-K growth mode QD formation. Experimentally, these results suggest an important relation between activation of self-assembly and area of the wetting layer in S-K mode.

The hole density for a 285-nm period 2-D array pattern is 1.23×10^9 cm^{-2}, which is lower than the dot density of typical InAs QDs randomly generated on unpatterned GaAs(001) by about an order of magnitude. However, sub-100-nm IL is currently being developed,[33] and nanoscale lithography technologies ultimately will be able to provide deep sub-100-nm patterns with hole densities higher than the dot density of an unpatterned substrate.

Our smallest d is 50 nm and most of InAs islands in this size range have a spherical-section shape. There are several possible ways to control the shape and size of QDs in NLAG. For example, QD shape different from pyramidal one such as a cylindrical QD could be realized if a hole diameter less than the lateral dimension of self-assembled QDs is achieved by nanoscale patterning. Since height of such QDs can be exactly controlled by deposition time, customizing QD shape and size in three-dimension will be possible with NLAG.

Relying on S-K mode for QD formation, there might be an upper limit in dot density desirable for device application. This is because the S-K mode results from a strain relaxation process, and dot density, correlated with the dot size and shape, is mainly governed by lattice-mismatch and growth conditions. If we use migration-assisted QD formation, the dot density can be increased by increasing the hole density. But, the dot size is influenced by accidental capturing of itinerant In atoms moving along the SiO_2 surface. Generally, this accidental capturing is proportional to the area of unit period of the mask, and as a result is critically affected by pattern uniformity. Therefore, there is a tradeoff between dot density and size uniformity especially in height depending on the uniformity of the lithographic and pattern transfer processes.

Dot density and size uniformity are directly related with gain of QD lasers. Enhancement of dot density while retaining quantum-size effects and narrowing the size distribution is one of the major goals in QD growth. Deep sub-100 nm NLAG which enables independent control of dot density and size in wide range could be a potential approach towards this goal.

IV. Summary

Nanoscale limited area growth (NLAG) of InAs islands on a SiO_2-patterned GaAs(001) substrate is reported. Two-dimensional arrays of 285-nm-period circular holes with three different bottom diameters of 50- to 100-nm, ~150 nm, and ~200 nm were generated on a SiO_2 film by large-area IL and dry etching. At a growth temperature of 535°C and a growth rate of 0.1 ML/s, a 10-nm thickness of InAs was selectively deposited on the GaAs surface of the SiO_2-masked substrate by MBE. Owing to the selectivity resulting from the very low zero sticking coefficient of In atoms on the SiO_2 surface, two different regimes of InAs QDs were generated on open GaAs surfaces bounded by the SiO_2 mask without migration-assisted QD formation. One regime gives spherical-section InAs dots with diameters of about 50- to 100-nm and a height of 10- to 15-nm entirely relying on NLAG, and the second regime is self-assembled QDs similar to those on an unpatterned substrate formed on holes of diameter greater than ~150 nm. The former corresponds to QDs without a wetting layer, whereas a nanoscale finite-area wetting layer plays a significant role in the latter regime. This leads to InAs islands severely impacted by coalescence between self-assembled QDs that was not observed for holes of diameter less than ~100 nm. The lateral size of the wetting layer strongly affects QD formation in S-K mode growth. From the hole-diameter dependent QD formation, a criterion on the size of holes for activation of self-assembling in S-K mode was proposed and is 100- to 150-nm which is 5- to 10-times greater than the lateral size of typical self-assembled QDs on an unpatterned GaAs substrate under similar growth conditions.

Acknowledgment. The authors wish to acknowledge the support of DARPA and of ARO/MURI.

[1] For a review, see "*Self-Assembled InGaAs/GaAs Quantum Dots,*" ed. M. Sugawara, Academic Press (1999).
[2] S. . Jeppesen, M. S. Miller, D. Hessman, B. Kowalski, I. Maximov, and L. Samuelson, Appl. Phys. Lett. **68**, 2228 (1996).
[3] C. H. Hahn, Y. J. Park, E. K. Kim, D. K. Min, S. K. Jung, and J. H. Park, Appl. Phys. Lett. **73**, 2479 (1998).
[4] T. Ishikawa, T, Nishimura, S. Kohmoto, and K. Asakawa, Appl. Phys. Lett. **76**, 167 (2000).
[5] 5. H. Lee, J. A. Johnson, M. Y. He, J. S. Speck, and P.M. Petroff, Appl. Phys. Lett. **78**, 105 (2001).
[6] 6. B. H. Choi, C. M. Park, S. H. Song, N. H. Son, S. W. Hwang, D. Ahn, and E. K. Kim, Appl. Phys. Lett. **78**, 1403 (2001).
[7] D. S. Mui, D. Leonard, L. A. Coldren, and P. M. Petroff, Appl. Phys. Lett. **66**, 1620 (1995).
[8] M. Kitamura, M. Nishioka, J. Oshinowo, and Y. Arakawa, Appl. Phys. Lett. **66**, 3663 (1995).
[9] Y. Sugiyama, Y. Sakuma, S. Muto, and N. Yokoyama, Appl. Phys. Lett. **67**, 256 (1995).
[10] W. Seifert, N. Carlsson, A. Petersson, L.-E. Wernersson, and L. Samuelson, Appl. Phys. Lett. **68**, 1684 (1996).
[11] M. Araki, Y. Hanada, H. Fuhikura, and L. Hasegawa, Jpn. J. Appl. Phys. **36**, 1763 (1997).
[12] R. Tsui, R. Zhang, K. Shiralagi, and H. Gorokin, Appl. Phys. Lett. **71**, 3254 (1997).
[13] A. Konkar, A. Madhukar, and P. Chen, Appl. Phys. Lett. **72**, 220 (1998).
[14] S. C. Lee, K. J. Malloy, and S. R. J. Brueck, J. Appl. Phys. (to be published).
[15] S. C. Lee, K. J. Malloy, and S. R. J. Brueck, J. Appl. Phys. (to be published).
[16] T. R. Ramachandran, A. Madhukar, I. Mukhametzhanov, R. Heitz, A. Kalburge, Q. Xie, and P. Chen, J. Vac. Sci. Technol. **B16**, 1330 (1998).
[17] H. R. Eisenberg and D. Kandel, Phys. Rev. Lett. **85**, 1286 (2000).
[18] H. J. Kim and Y. H. Xie, Appl. Phys. Lett. **79**, 263 (2001).
[19] C. J. Huang, Y. Yang, D. Z. Li, B. W. Cheng, L. P. Luo, J. Z. Yu, and Q. M. Wang, Appl. Phys. Lett. **78**, 2006 (2001).
[20] S. Hinooda, S. Loualiche, B. Lambert, N. Bertru, M. Paillard, X. Marie, and T. Amend, Appl. Phys. Lett. **78**, 3052 (2001).
[21] A. Stintz (unpublished).
[22] K. R. Evans, C. E. Stutz, E. N. Taylor, and J. E. Ehret, J. Vac. Sci. Technol. **B9**, 2427 (1991).
[23] G. D. Lian, J. Yuan, L. M. Brown, G. H. Kim, and D. A. Ritchie, Appl. Phys. Lett. **73**, 49 (1998).
[24] N. Moll, M. Scheffler, and E. Pehlke, Phys. Rev. **B58**, 4566 (1998).
[25] D. J. Bottomley, J. Vac. Sci. Technol. **B17**, 259 (1999).
[26] J. Marques, L. Geelhaar, and K. Jacobi, Appl. Phys. Lett. **78**, 2309 (2001).
[27] T. Mano, K. Watanabe, S. Tsukamoto, H. Fujioka, M. Oshima, and N. Koguchi, Jpn, J. Appl. Phys., Part 2 **38**, L1009 (1999).
[28] A. F. Tsatsul'nikov, A. R. Kovsh, A. E. Zhukov, Yu. M. Shernyakov, Yu. G. Musikhin, V. M. Ustinov, N. A. Bert, P. S. Kop'ev, Zh. I. Alferov, A. M. Minairov, J. L. Merz, N. N. Lendentsov, and D. Bimberg, J. Appl. Phys. **88**, 6272 (2000).
[29] Petroff et al.
[30] A. Stintz, G. T. Liu, A. L. Gray, R. Spillers, S. M. Delgado, and K. J. Malloy, J. Vac. Sci. Technol. **B18**, 1496 (2000).
[31] E. A. Fizgerald, G. P. Watson, R. E. Proano, D. G. Ast, P. D. Kirchner, G. D. Pettit, and J. M. Woodall, J. Appl. Phys. **65**, 2220 (1989).
[32] A. Fischer and H. Richter, Appl. Phys. Lett. **61**, 2656 (1992).

[33] S. R. J. Brueck, S. H. Zaidi, X. Chen, and Z. Zhang, Microelec. Engineering, **41/42**, 145 (1998).

Selective growth of Ge on Si(100) through vias of SiO_2 nanotemplate using solid source molecular beam epitaxy

Qiming Li, Sang M. Han,[a] Steven R. J. Brueck, Stephen Hersee, Ying-Bing Jiang, and Huifang Xu
University of New Mexico, Albuquerque, New Mexico 87131

(Received 9 July 2003; accepted 14 October 2003)

We demonstrate that Ge can be selectively grown on Si(100) through openings in a SiO_2 nanotemplate by solid source molecular beam epitaxy. The selectivity relies on the thermal instability of GeO and SiO near 650 °C. Ge islands grow in the template windows and coalesce on top of the template, forming an epitaxial lateral overgrowth (ELO) layer. Cross-sectional transmission electron microscopy images show that the Ge seeds and the ELO layer are free of threading dislocations. Only stacking faults are generated but terminate within 70 nm of the Ge–Si interface, while twins along {111} planes are observed in the ELO layer. The threading-dislocation-free Ge seeds and ELO layer are attributed to epitaxial necking as well as Ge–Si intermixing at the interface. © *2003 American Institute of Physics.*
[DOI: 10.1063/1.1632037]

Germanium quantum structures may give rise to advanced applications in quantum bit (i.e., qubit) computing[1,2] and optoelectronics.[3–6] Despite the promising applications, integrating Ge nanostructures for device manufacturing faces numerous challenges such as (1) defect-free epitaxial growth, (2) selective growth at precise locations, (3) uniform size and spatial distribution, (4) vertical alignment of dots in multiple stacks, and (5) controlled doping. Among these challenges, we intend to focus on the selective growth of defect-free Ge. Many methods exist to direct the self-assembly of Ge nanostructures[7–11] and to grow high quality epitaxial films.[12] We take advantage of the nanotemplate approach where Ge quantum structures grow selectively in a one-dimensional or two-dimensional (2D) array[7–11,13] through the template windows. The Ge nanostructures coalesce upon reaching the top of the template and form a high quality Ge epitaxial lateral overgrowth (ELO) layer.[12] The relaxed, defect-free epitaxial Ge layer on Si can also serve as a buffer layer on which III–V devices can be fabricated.[14]

Chemical vapor deposition and gas source molecular beam epitaxy (MBE) with Cl or H chemistry are successfully used to achieve selective growth of Ge on patterned Si substrates.[12,13,15,16] However, these growth techniques require extraordinary precaution to contain commonly used gas precursors such as GeH_4. Germane is pyrophoric and toxic to humans, the exposure threshold limit value of time-weighted average is only 0.2 parts per million.[17] In this work, we demonstrate that a 2D array of Ge nanostructures can be selectively grown on Si by solid source MBE, leading to a threading-dislocation-free Ge ELO layer.

To create the template, a 300-nm-thick layer of thermal SiO_2 is initially grown on an undoped Si(100) substrate by dry oxidation. Interferometric lithography[18–24] by the 355 nm Ar laser line and plasma etching[25–28] are used to pattern a 2D array of vias on the SiO_2 film, exposing the underlying Si substrate. The average diameter and depth of the SiO_2 vias are 200 and 300 nm, respectively. Figure 1(a) depicts the patterned vias along with the Ge seeds grown in a later step. The patterned samples are treated in Piranha solution for 20 min to remove carbon contamination and to chemically oxidize the exposed Si. The Piranha solution consists of four parts of H_2SO_4 (2 M) per one part of H_2O_2 (30 wt %) by volume. No HF treatment is performed on the patterned SiO_2 in order to preserve the template against chemical etching. After a deionized H_2O rinse and N_2 blow-dry, the patterned sample is loaded into a MBE chamber through a high-vacuum transfer loadlock. The base pressure of the ultrahigh vacuum (UHV) chamber is 4×10^{-10} Torr. Prior to Ge exposure, the loaded sample is heated to 900 °C in UHV for 15 min to remove surface contaminants introduced during sample transfer. The thermal treatment is also meant to partially remove the chemical oxide formed by Piranha treatment.[29,30] The sample is then cooled down to and maintained at 650 °C for Ge exposure. A Knudsen effusion cell is used to expose the sample to Ge. The effusion cell is operated at 1120 °C, and the Ge growth rate on Si is 0.7 ML/min. During growth, the chamber pressure remains below 1×10^{-9} Torr. As a control experiment, Ge is also deposited on a blank Si(100) sample under identical conditions. The blank sample is prepared by dissolving the thermally grown SiO_2 in HF (11 wt % in H_2O) for 10 min. The blank Si sample is then treated in Piranha and loaded into the UHV MBE chamber. The control sample subsequently undergoes

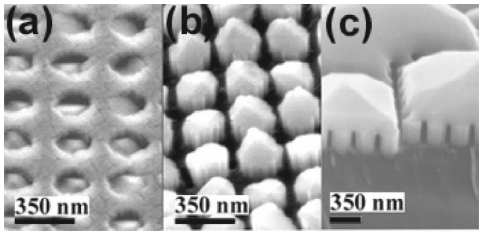


FIG. 1. SEM images: showing (a) selectively seeded Ge islands on Si in a 2D array through the vias of SiO_2 nanotemplate, (b) Ge seeds when the SiO_2 template is removed in HF, and (c) Ge ELO layer partially coalesced on top of the SiO_2 template.

[a]Electronic mail: meister@unm.edu

 © 2003 American Institute of Physics
Downloaded 30 Dec 2003 to 129.24.216.199. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

preparation steps identical to those of the templated sample. After Ge exposure, the samples are characterized by scanning electron microscopy (SEM), cross sectional transmission electron microscopy (XTEM), and energy dispersive spectroscopy (EDS). The samples for XTEM are prepared by mechanical polishing and Ar ion milling.

Figure 1 shows the SEM images that capture the surface morphology of epitaxially grown Ge islands and the ELO layer. In Fig. 1(a), Ge islands grow selectively on the exposed Si through the vias of the SiO_2 nanotemplate. No visible Ge nucleation occurs on the SiO_2 surface. The edges of the vias after plasma etching do not remain perfectly circular, and this roughness is transferred to the edges of the Ge structures [Fig. 1(b)]. Faceted Ge islands are revealed in Fig. 1(b) after the SiO_2 template is etched in HF (11 wt % in H_2O). Continued growth leads to the coalescence of Ge islands, forming an ELO layer shown in Fig. 1(c). The selective growth of Ge on Si over SiO_2 stems from the thermal instability GeO and SiO at the growth temperature, despite the use of Ge solid source and the absence of selectivity-controlling agents such as Cl or H.[13] Germanium is known to scavenge O from SiO_2 and to form GeO, which sublimates at 625 °C.[31] Surnev[32] also reports that GeO and SiO, produced during the reaction between Ge and SiO_2, start to sublimate at 410 and 670 °C, respectively. Wado *et al.* also suggest that the desorption of Ge and GeO from the SiO_2 surface at elevated temperatures enhances the selectivity.[33] In future publications, we will directly address the selectivity issue by probing Ge adspecies on partially oxidized Si, using scanning tunneling microscopy.

The interfacial defect morphology of the Ge layer grown on the blank Si substrate and the Ge ELO layer on the patterned Si substrate are compared by XTEM. Figure 2(a) shows a TEM image of a network of dislocation segments[34] that exist mostly within 300 nm from the Ge–Si interface when Ge is grown on a blank Si substrate. Despite the identical preparation steps used for the blank Si sample, except for the SiO_2 removal in HF, the Ge–Si interface of the blank Si sample roughens during Ge growth. Note that the images in Fig. 2 capture only the Ge–Si interface since the thickness of the Ge layer is 2 μm. The Ge layer far exceeds the field of view in Fig. 2. In comparison, the Ge ELO layer is free of dislocation segments, but twins are generated from the SiO_2–Ge interface mostly on top of the SiO_2 template. Figure 2(b) shows the twins propagating along the {111} planes to the surface. These twins may originate by two mechanisms. The roughness of the SiO_2–Ge interface causes lattice misplacement where the stacking sequence along the [111] directions is disrupted to follow the inverse order.[35] The merging fronts of coalescing Ge islands may also lead to the formation of these twins.[12,35]

Figure 3(a) is a magnified view of Region outlined by a square in Fig. 2(b). No threading dislocations are observed in the ELO layer and within the windows. The thin layers of contrast ($\sim$20 nm) near the template walls are due to the circular vias whose front and back are partially ion-milled for imaging, leaving two arced segments. These segments along with the depth averaging in TEM give rise to the contrast. The *z*-contrast image in Fig. 3(b) corroborates this interpretation by showing similar layers. The *z*-contrast image

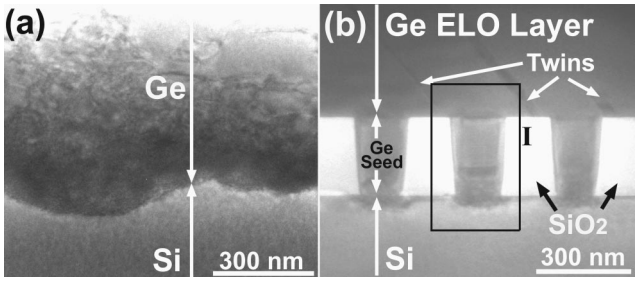


FIG. 2. XTEM images of (a) Ge layer grown on blank Si(100) with a network of dislocations at the Ge–Si interface and (b) Ge layer grown on templated Si(100) with no visible threading dislocations.

also provides an elemental map of Si, Ge, and O. Oxygen is the lightest element appearing dark whereas Ge is the heaviest element appearing light. The light contrast is also observed at the bottom of each Ge seed, indicating that Ge diffuses into Si during Ge growth. The diffusion layer is approximately 25 nm thick.

The bottom of the Ge seed is further magnified to interrogate the Ge–Si interface within the template windows and to determine whether the complex contrast near the bottom of Ge seed in Fig. 3(a) is due to tangled dislocation networks. The high-resolution TEM image in Fig. 4 shows a 3-nm-thick amorphous layer at the Ge–Si interface, and EDS detects a highly elevated level of oxygen near the interface. No carbon signal is detected near the interface. These observations suggest that the amorphous region is an oxidation layer incompletely removed during the thermal degassing step at 900 °C. Despite the oxidation layer, Ge grows epitaxially on the Si substrate. Similar epitaxial growth is observed by other authors.[36,37] Germanium is believed to scavenge O from SiO_2. The byproducts sublime as GeO and SiO,[32] creating areas of bare Si.[37] The size of the oxide-free openings is approximately 10 nm, and such region is labeled as II in Fig. 4 for clarity. Germanium crystal inclusions grow in these bare Si areas[37] and then grow laterally to form an epitaxial layer. No dislocation segments are seen in epitaxial Ge above the oxide-free region, whereas stacking faults emanate from the oxide region and propagate along {111} planes. Similar stacking faults are commonly observed in Si homoepitaxy at the SiO_2–Si interface.[38–40] During the oxidation process, the volume increase in the oxidation layer causes compressive stress, which in turn leads to the supersaturation of Si in the interstices of the epitaxial layer and

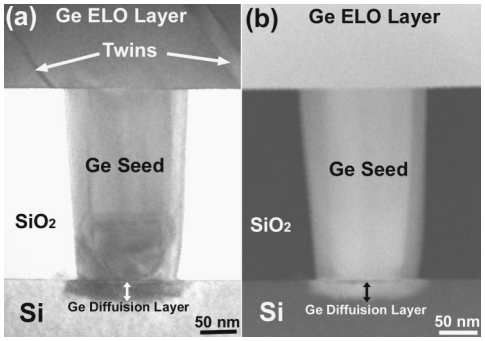


FIG. 3. Magnified view of Region I in Fig. 2(b): (a) XTEM image of Ge seeds and ELO layer showing twin defects and (b) *z*-contrast image showing 25-nm-thick Ge diffusion layer.

Downloaded 30 Dec 2003 to 129.24.216.199. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

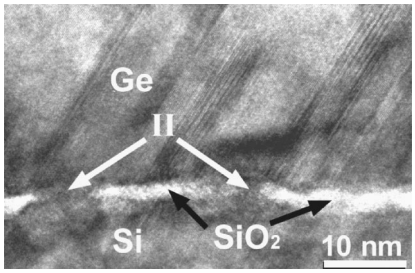


FIG. 4. High-resolution XTEM image to show the details of the Ge–Si interface where stacking faults emanate from the partially oxidized interface.

eventually to the nucleation of stacking faults.[39] In the case of Ge heteroepitaxy, the stress caused by Ge oxidation at the interface may similarly cause the generation of stacking faults. These stacking faults in Ge terminate within 70 nm from the interface unless they run into the SiO_2 template sidewall and terminate even closer to the interface. The termination of stacking faults transitions into a defect-free Ge region. The termination of defects by the SiO_2 sidewall, known as necking, may be also responsible for the absence of threading dislocations in the Ge seeds.[12]

The absence of threading dislocations is also attributed to Ge–Si intermixing below the Ge–Si interface. The lattice constant in the intermixing region shown in Fig. 3 is 5.57 Å based on its electron diffraction pattern. This lattice constant is greater than 5.43 Å of Si below the diffusion layer but less than 5.66 Å of Ge above the diffusion layer. That is, the intermixing of Ge and Si creates a graded Ge diffusion layer. The gradation helps to form a low-defect-density Ge epitaxial layer because the lattice mismatch is less than 4.2% within the graded area. Thus, no atomically abrupt interface exists between Ge and Si.

We speculate that the ELO layer is also fully relaxed. The speculation is based on the theory by Luryi and Suhir[41] that the strain energy density decays exponentially from the heterojunction and that the characteristic length of decay is comparable to the dimension of the heterojunction. For the 200-nm-wide Ge–Si interface, the strain density would decrease to its minimum within 200 nm of the interface, and the strain energy density may never exceed the threshold to create dislocations.

We have selectively grown Ge nanostructures in a 2D array, using solid source MBE. The selectivity relies on observed thermal instability of GeO and SiO that desorb from the SiO_2 template surface. The germanium structures coalesce to form an ELO layer over an extended period of Ge exposure. The germanium seeds as well as the ELO layer are free of threading dislocations. The only defects observed are stacking faults at the Ge–Si interface and twins in the ELO layer.

The authors thank the University of New Mexico SEED and the National Science Foundation (DMR-0094145) for their generous financial support. The TEM analysis was conducted in the Transmission Electron Microscopy Laboratory of the Department of Earth and Planetary Sciences at UNM. The National Science Foundation (CTS 98-71292) and the State of New Mexico support the TEM facility.

[1] Paul A. Cain, Haroon Ahmed, and David A. Williams, J. Appl. Phys. **92**, 346 (2002).
[2] Rutger Vrijen, Eli Yablonovitch, Kang Wang, Hong Wen Jiang, Alex Balandin, Vwani Roychowdhury, Tal Mor, and David DiVincenzo, Phys. Rev. A **62**, 012306 (2000).
[3] L. Vescan, T. Stoica, O. Chretien, M. Goryll, E. Mateeva, and A. Mück, J. Appl. Phys. **87**, 7275 (2000).
[4] M. Stoffel, U. Denker, and O. G. Schmidt, Appl. Phys. Lett. **82**, 3236 (2003).
[5] T. Stoica and L. Vescan, J. Appl. Phys. **93**, 4461 (2003).
[6] P. Schittenhelm, C. Engel, F. Findeis, G. Abstreiter, A. A. Darhuber, G. Bauer, A. O. Kosogov, and P. Werner, J. Vac. Sci. Technol. B **16**, 1575 (1998).
[7] G. Jin, J. L. Liu, and K. L. Wang, Appl. Phys. Lett. **76**, 3591 (2000).
[8] G. Jin, J. L. Liu, S. G. Thomas, Y. H. Luo, K. L. Wang, and B.-Y. Nguyen, Appl. Phys. Lett. **75**, 2752 (1999).
[9] T. I. Kamins and R. Stanley Williams, Appl. Phys. Lett. **71**, 1201 (1997).
[10] E. S. Kim, N. Usami, and Y. Shiraki, Appl. Phys. Lett. **72**, 1617 (1998).
[11] Jennifer L. Gray, Robert Hull, and Jerrold A. Floro, Appl. Phys. Lett. **81**, 2445 (2002).
[12] T. A. Langdo, C. W. Leitz, M. T. Currie, E. A. Fitzgerald, A. Lochtefeld, and D. A. Antoniadis, Appl. Phys. Lett. **76**, 3700 (2000).
[13] L. Vescan, Mater. Sci. Eng., B **28**, 1 (1994).
[14] R. Venkatasubramanian, M. L. Timmons, J. B. Posthill, B. M. Keyes, and R. K. Ahrenkiel, J. Cryst. Growth **107**, 489 (1991).
[15] D. J. Dumin, J. Cryst. Growth **8**, 33 (1971).
[16] Yoshiki Nitta, Motoshi Shibata, Ken Fujita, and Masakazu Ichikawa, Surf. Sci. **462**, L587 (2000).
[17] BOC Gases, Materials Safety Data Sheet, MSDSonline, http://www.msdsonline.com/ (1995).
[18] S. R. J. Brueck, S. H. Zaidi, X. Chen, and Z. Zhang, Microelectron. Eng. **42**, 145 (1998).
[19] Xiaolan Chen and S. R. J. Brueck, J. Vac. Sci. Technol. B **17**, 921 (1999).
[20] X. L. Chen and S. R. J. Brueck, Opt. Lett. **24**, 124 (1999).
[21] S. H. Zaidi and S. R. J. Brueck, J. Vac. Sci. Technol. B **11**, 658 (1993).
[22] S. Zaidi and S. R. J. Brueck, Proc. SPIE **3740**, 340 (1999).
[23] L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll, Appl. Opt. **17**, 1165 (1978).
[24] E. H. Anderson, C. M. Horwitz, and H. I. Smith, Appl. Phys. Lett. **43**, 874 (1983).
[25] Riccardo d'Agostino, J. Appl. Phys. **52**, 162 (1981).
[26] V. M. Donnelly, D. L. Flamm, W. C. Dautremont-Smith, and D. J. Werder, J. Appl. Phys. **55**, 242 (1984).
[27] N. Schaepkens, G. S. Oehrlein, and J. M. Cook, J. Vac. Sci. Technol. B **18**, 856 (2000).
[28] J. A. Mucha, D. W. Hess, and E. S. Aydil, in *Introduction to Microlithography*, 2nd ed., edited by L. F. Thompson, C. G. Willson, and M. J. Bowden (American Chemical Society, Washington, DC, 1994), Chap. 5, p. 377.
[29] Akitoshi Ishizaka and Yasuhiro Shiraki, J. Electrochem. Soc. **133**, 666 (1986).
[30] N. Miyata, M. Shigeno, Y. Arimoto, and T. Ito, J. Appl. Phys. **74**, 5275 (1993).
[31] J. F. Morar, B. S. Myerson, U. O. Karisson, F. J. Himpsel, F. R. McFeely, D. Rieger, A. Taleb-Ibrahimi, and J. A. Yarmoff, Appl. Phys. Lett. **50**, 463 (1987).
[32] S. Surnev, Surf. Sci. **282**, 10 (1993).
[33] Hiroyuki Wado, Tadami Shimizu, Seiji Ogura, Makoto Ishida, and Tetsuro Nakamura, J. Cryst. Growth **150**, 969 (1995).
[34] S. Oktyabrsky, H. Wu, R. D. Vispute, and J. Narayan, Philos. Mag. A **71**, 537 (1995).
[35] Y. C. Shih, J. B. Liu, W. G. Oldham, and R. Gronsky, Appl. Phys. Lett. **65**, 1142 (1994).
[36] Alexander A. Shklyaev, Motoshi Shibata, and Masakazu Ichikawa, Phys. Rev. B **62**, 1540 (2000).
[37] A. Barski, M. Derivaz, J. L. Rouviére, and D. Buttard, Appl. Phys. Lett. **77**, 3541 (2000).
[38] S. P. Murarka and G. Quintana, J. Appl. Phys. **48**, 46 (1977).
[39] T. Y. Tan and U. Gösele, Appl. Phys. Lett. **39**, 86 (1981).
[40] T. Y. Tan and U. Gösele, Chem. Phys. Lett. **40**, 616 (1982).
[41] Serge Luryi and Ephraim Suhir, Appl. Phys. Lett. **49**, 140 (1986).

Downloaded 30 Dec 2003 to 129.24.216.199. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/aplo/aplcr.jsp

HTML ABSTRACT + LINKS

Defect reduction mechanisms in the nanoheteroepitaxy of GaN on SiC

X. Y. Sun, R. Bommena, D. Burckel, A. Frauenglass, M. N. Fairchild, and S. R. J. Brueck
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

G. A. Garrett and M. Wraback
Army Research Lab, Adelphi, Maryland 20783

S. D. Hersee[a)]
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

(Received 25 August 2003; accepted 18 November 2003)

This article describes defect reduction mechanisms that are active during the growth of GaN by nanoheteroepitaxy on (0001) 6H SiC. Nanoheteroepitaxial (NHE) and planar GaN epitaxial films were grown and compared using transmission electron microscopy, photoluminescence, x-ray diffraction, and time resolved photoluminescence. It was found that in addition to the previously reported defect reduction mechanism that results from the high compliance of nanoscale nuclei, other independent defect reduction mechanisms are also active during NHE including: (i) filtering of substrate defects, (ii) improved coalescence at the nanoscale, and (iii) defect termination at local free surfaces. Also, it was found that the biaxial strain in the GaN film could be significantly reduced by using a ''grouped'' NHE pattern geometry. Time resolved photoluminescence measurements on NHE GaN samples with this geometry showed a more than tenfold increase in carrier lifetime compared to GaN grown on planar SiC. © *2004 American Institute of Physics.*
[DOI: 10.1063/1.1639952]

INTRODUCTION

GaN and the related III–N materials are typically grown on lattice mismatched substrates such as silicon, sapphire, and silicon carbide. A consequence of this mismatch is that III–N epitaxial layers contain a high density of threading dislocations that have been linked to the failure of lasers[1] and the breakdown of *pn* junctions.[2] Furthermore, it is anticipated that threading dislocations will aggravate compositional stability problems associated with In segregation from InGaN alloys.[3] To reduce this defect concentration several approaches exploiting lateral overgrowth on a patterned substrate have been demonstrated. These approaches include lateral epitaxial overgrowth,[4,5] pendeo epitaxy,[6] and cantilever epitaxy,[7] and significantly reduced dislocation densities in the range $10^6-10^7\ \mathrm{cm}^{-2}$ have been achieved. Theory[8,9] and experiment[10] indicate that a further reduction in defect density is possible if the lateral overgrowth approach is extended to the nanoscale. This paradigm is embodied in the nanoheteroepitaxy approach[9] where strain energy is reduced by leveraging the high, three dimensional (3D) compliance of nanoscale features.[11] Conservative calculations of the strain energy[9] suggest that the 3.2% lattice mismatch between GaN and SiC can be accommodated without defect generation when the nanoheteroepitaxial (NHE) pattern features are reduced to ~40 nm in diameter.

This article describes additional defect reduction mechanisms that are operative during the NHE process even when the NHE pattern size is larger than the 40 nm target value that theory[9] suggests is required for the elimination of strain-related defects: (1) filtering of substrate defects as a result of the small initial growth area; (2) improved nanoscale coalescence; and (3) the termination of defects at local free surfaces. It is found that these additional mechanisms contribute significantly to quality improvement in NHE GaN films. The thermal mismatch stress between the substrate and the epitaxial layers can also be controlled by employing a multiple-scale nanopatterning geometry leading to localized growth areas. In-plane biaxial stress values as low as 0.18 GPa were achieved within arrays of localized areas of ~10 μm linear dimensions, large enough for many device applications. The defect and stress reduction mechanisms reported here are also expected to occur during the NHE of other mismatched semiconductor materials, such as GaN on Si, which is of critical importance for solid-state lighting applications.

EXPERIMENTAL SETUP

The design of the NHE sample structure has been discussed previously.[9,10] The experiments described here used a simple structure, consisting of holes etched into a 35 nm thick, low-pressure chemical vapor deposition Si_3N_4 layer that was deposited onto a SiC substrate. Metalorganic chemical vapor deposition (MOCVD) growth of GaN on this structure is selective (growth only on the exposed SiC) and creates an array of compliant, nanoscale islands. Two nanoscale hole geometries: a continuous, large-area array pattern geometry [Fig. 1(a)], and a local-area, grouped array pattern [Fig. 1(b)] were fabricated by interferometric lithography.[12] In both cases, the pattern period was 500 nm and the hole diameter was approximately 100 nm. In the grouped array pattern [Fig. 1(b)] the center-to-center spacing of the groups

[a)]Author to whom correspondence should be addressed; electronic mail: shersee@chtm.unm.edu

 © 2004 American Institute of Physics

Downloaded 01 Mar 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

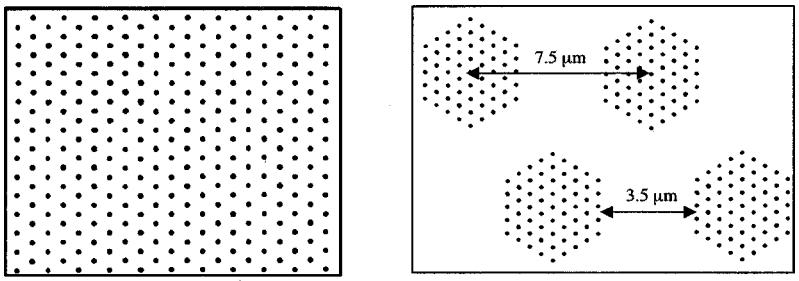


FIG. 1. Nanoscale pattern geometries used for NHE growths: (a) patterned array geometry and (b) "grouped" patterned array geometry.

was 7.5 μm and the span of each group was 3.5 μm. These dimensions were selected to allow coalescence within the group but not in the space between the groups. These moiré patterns were created by multiple exposures with slightly different periods and the holes near the edges of each group are approximately 30% smaller than the holes in the center of the pattern. While a hole diameter of ~100 nm is too large to offer sufficient reduction in strain energy to eliminate stress-induced defects via the 3D compliance NHE defect reduction mechanism,[9] a significant improvement in GaN material quality is nevertheless achieved as detailed below.

The GaN epitaxial layers were grown by MOCVD in a radio-frequency heated, horizontal reactor at 100 Torr using trimethygallium and ammonia sources. Growth was initiated using a 15 Å GaN buffer grown at 650 °C. The growth temperature was then increased to 1080 °C to grow the bulk of the GaN epilayer. During buffer layer growth the average growth rate (measured on the planar SiC reference substrate) was kept low at 1.4 nm/min to ensure selectivity, which is critical to the NHE process. This growth rate was increased during the higher temperature growth stage to 10 nm/min. In each experiment we grew GaN on the two arrays shown in Fig. 1 and on a planar (unpatterned) SiC reference wafer.

RESULTS

Figure 2 shows low resolution plan-view scanning electron microscope (SEM) images of 1.8 μm thick GaN films grown side-by-side on a planar SiC substrate [Fig. 2(a)] and on a continuously patterned NHE array [Fig. 2(b), substrate patterned as in Fig. 1(a)]. At this thickness the GaN film on the patterned array is fully coalesced. The GaN film grown on the planar substrate exhibits a "hole" defect density of $9.1\times10^6\,\mathrm{cm}^{-2}$. These "hole" defects are believed to be threading dislocations or nanopipe/micropipe defects that originate in the SiC substrate and thread into the GaN epitaxial layer. In the NHE GaN sample the density of "hole" defects is reduced to $\sim6\times10^4\,\mathrm{cm}^{-2}$ (this defect density was estimated from measurement of larger areas of the sample). This reduction in defect density in the NHE sample illustrates one of the additional defect reduction mechanisms that are active in NHE. Because much of the SiC substrate is covered by the Si_3N_4 growth mask, most substrate threading defects (including threading dislocations, nanopipes, and micropipes) will terminate at the Si_3N_4 mask and will not propagate into the GaN epitaxial layer. Both GaN layers also show cracks that originate from the thermal expansion mismatch on cooling from the growth temperature. As shown below, the grouped nanoscale pattern can eliminate these cracks.

Figure 3 shows the results of NHE GaN growth on the "grouped" array pattern geometry sample shown in Fig. 1(b). Each group of nanoscale nuclei has coalesced to form a perfectly faceted pyramid of GaN [Fig. 3(a)] that is approximately 6.5 μm across at the base (measured at the pyramid base between the $\{01\underline{1}n\}$ facets). The underlying "grouped" nucleation pattern can be seen in Fig. 3(a), where at the edge of the sample one of the pyramids became detached. It should be noted that while similar GaN facetted pyramids have been reported previously,[13,14] the pyramids reported here were nucleated from approximately 80 individual nuclei, which then coalesced to form the single crystal GaN pyramid. This excellent coalescence illustrates one of the key differences between NHE and the conventional lateral epitaxial overgrowth (LEO) approach. When coalescence occurs in the NHE process the epilayer is thin and compliant and high quality coalescence is obtained.[11] In the case of LEO,

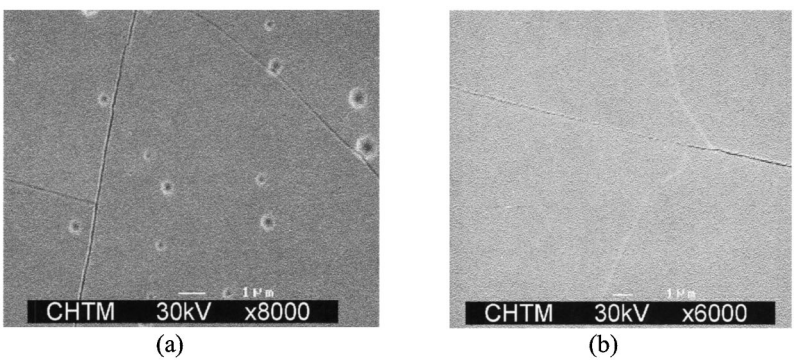


FIG. 2. (a) SEM image of GaN on planar SiC showing "hole" defects and cracks and (b) SEM image of fully coalesced NHE GaN on patterned SiC showing reduced density of "hole" defects.

Downloaded 01 Mar 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

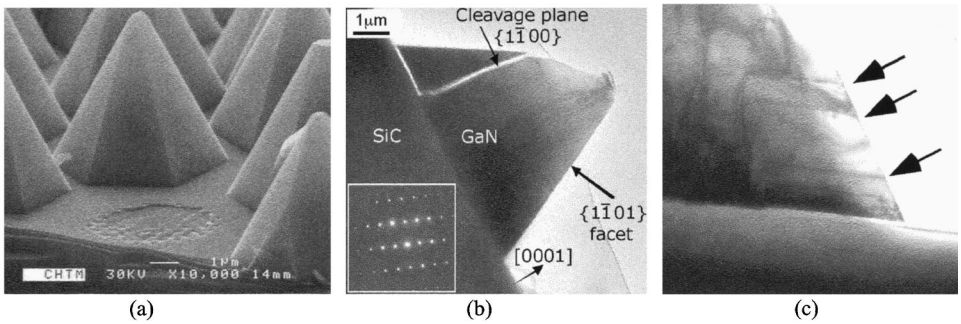


FIG. 3. (a) GaN pyramids on group-patterned SiC. (b) Bright-field XTEM image of pyramid. (Inset) electron diffraction pattern showing pyramid is single crystal. (c) XTEM image of pyramid side showing dislocation bending.

by the time coalescence occurs the epilayer is typically ''tilted'' and thick (between 1 and 10 μm) and defects are generally observed at the coalescence boundary. Note that neither the ''hole'' defects [shown in Fig. 2(a)] nor cracking due to thermal mismatch strain [shown in Figs. 2(a) and 2(b)] were observed in these GaN pyramids.

Transmission electron microscopy (TEM) analysis of a GaN pyramid feature is shown in Figs. 3(b) and 3(c). The crystallographic orientations and Miller indices of the facets in Fig. 3(b) are assigned based on measurement of the facet angle with respect to the [0001] direction ($61°\pm1°$) and are consistent with previous work[15] that has shown such facets to be of the prismatic $\{01\underline{1}n\}$ family. During TEM sample preparation an example of cleavage on a $\{1\underline{1}00\}$ facet was also observed. The electron diffraction pattern from the upper part of this GaN pyramid [inset in Fig. 3(b)] confirms that it is single crystal. Figure 3(c) is a TEM image of the pyramid adjacent to one of the $\{1\underline{1}01\}$ facets. This image confirms that the bending behavior of threading dislocations previously reported for GaN on Si[14] is also active during the growth of GaN on SiC. It is assumed that threading dislocations can minimize their energy by bending and terminating at a $\{1\underline{1}01\}$ surface. Three such dislocations are indicated by arrows in Fig. 3(c). In NHE, where growth begins as a dense array of nanoscale nuclei, there are many such sidewall surfaces available for this bending/termination process to occur. This bending mechanism is also responsible for the result that in NHE GaN nuclei threading dislocations appear to be replaced almost completely by in-plane type defects. This change in defect kinetics is another major aspect of nanoheteroepitaxial growth that has also been observed in the NHE growth of GaN on Si[10,14] as well as in the growth of GaN on SiC.

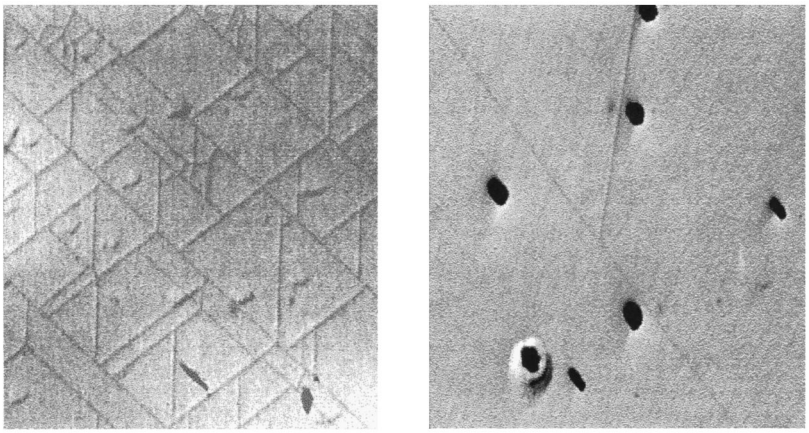
The difference in thermal expansion coefficients (the thermal mismatch) between the component materials in a heterostructure can be a source of significant biaxial stress that in thick samples causes cracking as seen in Fig. 2. In these GaN/SiC structures micro-Raman spectroscopy was used to measure the residual biaxial stress in the GaN films. The Raman scattering measurements were made in a standard backscattering $Z(XX)\underline{Z}$ geometry. For the grouped (pyramidal) samples, the nonplanar geometry effectively eliminates any orientational or polarization selectivity. The Raman measurement used a 488 nm Ar ion laser source, a $40\times$ microscope objective (spot size ~1 μm), a 1 m double spectrometer and an optical multichannel analyzer (OMA). Spectral shifts were calibrated against emission lines from the Ar plasma with a repeatability of approximately one diode in the OMA array or 0.12 cm^{-1}. In general, for an elastic material where the elastic constants are known, converting the phonon frequency shift to a biaxial stress value is straightforward. The phonon frequency shift is first converted to strain in the ''*c*'' lattice parameter then Hookes' law is used to calculate the biaxial stress in the basal plane.[16] In this Raman scattering geometry the GaN $E2$ peak is a convenient indicator of strain, however, previous work[16–22] has attributed a wide range of conversion factors relating the biaxial strain in the GaN epitaxial layer to the $E2$ Raman peak shift. [The conversion factor k ($cm^{-1}\,GPa^{-1}$) is in the range: $2.7<k<7.6$, where $\Delta\omega=k\sigma$, $\Delta\omega$ is the phonon frequency shift (cm^{-1}) and σ is the biaxial stress.] This wide range in k value is due largely to use of different values for the GaN elastic constants (Young's modulus and Poisson's ratio); moreover, some systematic inaccuracy is unavoidable as Hookes' law is being applied to a partially relaxed (and therefore nonelastic) GaN film. Despite these reservations, most authors typically measure a stress induced $E2$ phonon frequency reduction (relative to the unstrained $E2$ peak frequency of 568 cm^{-1}) of approximately 4 cm^{-1} in GaN/SiC structures.

The biaxial stresses shown in Table I were calculated using a conversion factor of $k=5.1\,cm^{-1}\,GPa^{-1}$. Each stress value is an average value calculated from at least five mea-

TABLE I. Biaxial stress measured in the GaN/SiC structures by Raman spectroscopy.

Sample structure	$\langle E_2\rangle$ (cm^{-1})	Shift from bulk (cm^{-1})	Biaxial stress (GPa)
GaN/SiC: planar	564.91	−3.09	−0.61
GaN/SiC: coalesced uniform hexagonal array	564.25	−3.74	−0.73
GaN/SiC: coalesced grouped hexagonal array	567.1	−0.9	−0.18

Downloaded 01 Mar 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp



a *b*

FIG. 4. Optical micrographs at 20× magnification showing reduction of crack density in planar, 2 μm thick, GaN films on SiC, with reduced post-growth cooling rate: (a) postgrowth cooling rate at 100 °C/min and (b) postgrowth cooling rate at 20 °C/min.

surements of the Raman spectrum on multiple samples. For all of the GaN growths on SiC, the $E2$ Raman peak wave number is less than the bulk (unstrained) value of 568 cm^{-1}, indicating that the GaN is in tension as expected from the thermal expansion coefficients. For the GaN/SiC planar growth sample and the coalesced NHE uniform array sample we measure biaxial stresses of approximately -0.61 and -0.73 GPa, respectively. The NHE ''grouped'' geometry sample consistently had a much smaller biaxial stress of only 0.18 GPa.

A simple model can be used to explain these stress values. We assume that at the GaN growth temperature of ~1040 °C, defect creation will reduce the mismatch strain to approximately zero. When a planar GaN/SiC sample is cooled from the GaN growth temperature to room temperature we assume that this defect structure remains largely unchanged (the defect structure is ''frozen in'') and the stress measured at room temperature will then be due to the thermal mismatch. In this case the strain at room temperature is given by $\epsilon = \Delta T(\alpha_{\text{sub}} - \alpha_{\text{epi}})$, where ΔT is the temperature difference and α is the thermal expansion coefficient. The biaxial stress value is obtained using the approach of Reiger[16] from Eq. (1)

$$\sigma = \frac{\epsilon E}{\nu}, \tag{1}$$

where σ is the biaxial stress (GPa), E is Young's modulus (GPa), and ν is Poisson's ratio. Using a growth temperature of ~1000 °C, thermal expansion coefficients of: $\alpha_{\text{SiC}} = 4.2 \times 10^{-6}/°\text{C}$ and $\alpha_{\text{GaN}} = 5.6 \times 10^{-6}/°\text{C}$, a Young's modulus of 200 GPa, and a Poisson's ratio of 0.38, this calculation yields a biaxial stress value of -0.74 GPa. This value is comparable to the biaxial stress values measured for the planar growth and coalesced NHE films in Table I.

The reduced biaxial stress measured in the GaN pyramid structures (Fig. 3) has been previously observed,[23,24] especially near the top of the pyramid, and it is of interest to understand why this stress reduction occurs. The theory of nanoheteroepitaxy[9] shows that the stress and strain fall off exponentially in a 3D semiconductor object, with a decay length that is comparable to the diameter of that object. It is important to understand that nanoscale size is not specifically required to achieve this exponential stress/strain decay, rather it is an approximately unity aspect ratio of the object (height to width) that must be achieved. The GaN pyramid, where the height is comparable to the width of the object, satisfies this aspect ratio requirement and a significant reduction stress and strain is thus expected. While this mechanism could account for most of the observed reduction in residual biaxial stress in the GaN pyramids, an additional stress reduction mechanism is also believed to be active.

After growth, as the GaN/SiC sample is cooled from high temperature, the thermal mismatch stress will change and there will be some rearrangement of the defect structure in response to this change. This defect rearrangement will be facilitated by the presence of a surface which can act as a source of vacancies and can allow defect termination or generation. In the grouped sample (Fig. 3) defects can also interact with the sidewall surfaces of the pyramid as shown in Fig. 3(c). We suggest that as the grouped sample is cooled from growth temperature, the defect structure in this sample will therefore be better able to respond to the changing thermal mismatch stress in the GaN/SiC structure. Support for this hypothesis was recently obtained by comparing the stress in fully-coalesced, 2 μm thick, planar GaN epilayers grown on SiC, which were cooled at different rates following growth. Figure 4 shows optical micrographs of these planar GaN films at 20× magnification. The crack density is clearly reduced when the postgrowth cooling rate is reduced from its normal value of 100 °C/min [Fig. 4(a)] to 20 °C/min [Fig. 4(b)]. This reduction in crack density indicates a reduced residual stress in the slower cooled film and indicates that there has been some change of defect structure (and stress) in the slower cooled film, in support of the above hypothesis. An interdependence of defect density and sample lateral size has also been observed in other materials such as InGaAs/GaAs.[25]

Time resolved photoluminescence measurements with subpicosecond resolution were also performed on these samples using a luminescence downconversion technique. PL excitation pulses of ~100 fs duration tunable from 375 nm to 225 nm were obtained by frequency doubling of the visible signal pulses from a 250 kHz Ti:sapphire regenerative-amplifier-pumped optical parametric amplifier.

Downloaded 01 Mar 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

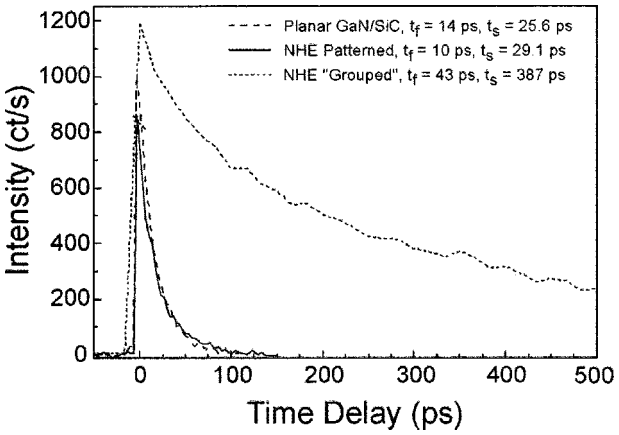


FIG. 5. Time resolved photoluminescence shows a significantly longer carrier lifetime for the NHE ''grouped'' GaN sample.

Figure 5 shows the decay of PL intensity for the three sample types after a 345 nm, 530 μW pump pulse. The decay is fitted by two exponentials to yield a fast (t_f) and a slow (t_s) component of carrier lifetime. The superior carrier lifetime of the NHE ''grouped'' pattern is clearly demonstrated with an $\sim 4\times$ increase in the ''fast'' carrier lifetime component and more than a $10\times$ increase in the ''slow'' carrier lifetime component for this sample. Similar improvements in the fast and slow components of carrier lifetime were also obtained in the grouped NHE sample for 345 nm, 1.56 mW pump pulses. The room temperature photoluminescence intensity of the ''grouped'' sample was approximately 100 times more intense than that of the planar GaN on SiC sample. Part of this efficiency improvement is undoubtedly due to the improved coupling of radiation both into and out of the pyramidal samples as compared with a planar geometry, where the high refractive index of the GaN prevents escape of much of the photoluminescence radiation; however, some of the increase is likely due to reduced nonradiative processes as reflected in the lifetime measurements.

CONCLUSIONS

We have analyzed NHE GaN grown on SiC by MOCVD and observed that the quality of NHE GaN is significantly improved compared to planar growth even when the nanoscale nuclei in these samples were larger than the target size of 40 nm and were not expected to show defect elimination via the 3D NHE compliance strain-relief mechanism.[9] It is found that the NHE growth mask is effective at terminating substrate defects. Also the close proximity of the growth surface to unbound sidewalls in the NHE sample structure leverages the bending of threading dislocations and their termination at sidewalls. The ''grouped'' NHE structures showed a dramatic improvement in both carrier lifetime and room temperature photoluminescence intensity, and a significantly reduced biaxial stress of -0.18 GPa.

While the GaN material quality in this grouped array structure is significantly improved and this geometry may be acceptable for some emitter array applications, the challenge remains to combine a low defect density with reduced stress (no cracking) in a fully coalesced, planar film.

ACKNOWLEDGMENTS

The authors wish to acknowledge support for this work from the Office of Naval Research, the Army Research Office under the MURI program, and the Army Research Laboratory. Part of the research reported in this document was performed in connection with Contract No. DAAD 17-02-C-0111 with the U.S. Army Research Laboratory. The views and conclusions contained in this document are those of the authors and should not be interpreted as presenting the official policies or position, either expressed or implied, of the U.S. Army Research Laboratory or the U.S. Government unless so designated by other authorized documents. Citation of manufacturer's or trade names does not constitute and official endorsement or approval of the use thereof. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

[1] S. Nakamura *et al.*, Appl. Phys. Lett. **72**, 211 (1998).
[2] P. Kozodoy, J. P. Ibbetson, H. Marchand, P. T. Fini, S. Keller, J. S. Speck, S. P. Denbaars, and U. K. Mishra, Appl. Phys. Lett. **73**, 975 (1998).
[3] J. Ramer, A. Hecht, and S. D. Hersee, J. Electron. Mater. **26**, 1109 (1997).
[4] D. Kapolnek, S. Keller, R. Vetury, R. D. Underwood, P. Kozodoy, S. P. Denbaars, and U. K. Mishra, Appl. Phys. Lett. **71**, 1204 (1997).
[5] T. S. Zheleva, O.-H. Nam, M. D. Bremser, and R. F. Davis, Appl. Phys. Lett. **71**, 2472 (1997).
[6] A. M. Roskowski, E. A. Preble, S. Einfeldt, P. M. Miraglia, and R. F. Davis, IEEE J. Quantum Electron. **38**, 1006 (2002).
[7] C. I. H. Ashby, C. C. Mitchell, J. Han, N. A. Missert, P. P. Provencio, D. M. Follstaedt, G. M. Peake, and L. Griego, Appl. Phys. Lett. **77**, 3233 (2000).
[8] S. Luryi and E. Suhir, Appl. Phys. Lett. **49**, 140 (1986).
[9] D. Zubia and S. D. Hersee, J. Appl. Phys. **85**, 6492 (1999).
[10] D. Zubia, S. H. Zaidi, S. R. J. Brueck, and S. D. Hersee, Appl. Phys. Lett. **76**, 858 (2000).
[11] S. D. Hersee *et al.*, IEEE J. Quantum Electron. **38**, 1017 (2002).
[12] S. H. Zaidi, A. S. Chu, and S. R. J. Brueck, Mater. Res. Soc. Symp. Proc. **358**, 957 (1995).
[13] S. Bidnyk, B. D. Little, Y. H. Cho, J. Krasinski, J. J. Song, W. Yang, and S. A. McPherson, MRS Internet J. Nitride Semicond. Res. **4S1**, G6.48 (1999).
[14] S. Tanaka, Y. Kawaguchi, N. Sawaki, M. Hibino, and K. Hiramatsu, Appl. Phys. Lett. **76**, 2701 (2000).
[15] S. Nitta, T. Kashima, M. Kariya, Y. Yukawa, S. Yamaguchi, H. Amano, and I. Akasaki, MRS Internet J. Nitride Semicond. Res. **5S1**, W2.8 (2000).
[16] W. Reiger *et al.*, Appl. Phys. Lett. **68**, 970 (1996).
[17] T. Kozawa, T. Kachi, H. Kano, H. Nagase, N. Koide, and K. Manabe, J. Appl. Phys. **77**, 4389 (1995).
[18] J. Kruger, C. Kisielowski, T. Suski, S. Ruvimov, Z. Liliental-Weber, J. W. Ager III, M. Rubin, and E. R. Weber, Proc. IEEE **9**, 89 (1996).
[19] V. Y. Davydov *et al.*, J. Appl. Phys. **82**, 5097 (1997).
[20] M. Mynbaeva *et al.*, MRS Internet J. Nitride Semicond. Res. **4**, 14 (1999).
[21] Q. K. K. Liu, A. Hoffmann, H. Siegle, A. Kaschner, C. Thomsen, J. Christen, and F. Bertram, Appl. Phys. Lett. **74**, 3122 (1999).
[22] S. Tripathy, S. J. Chua, P. Chen, and Z. L. Miao, J. Appl. Phys. **92**, 3503 (2002).
[23] A. Hoffmann *et al.*, J. Cryst. Growth **189/190**, 630 (1998).
[24] K. C. Zheng, J. Y. Lin, and H. X. Jiang, Appl. Phys. Lett. **74**, 1227 (1999).
[25] E. A. Fitzgerald, G. P. Watson, R. E. Proano, D. G. Ast, P. D. Kirchner, G. D. Petit, and J. M. Woodall, J. Appl. Phys. **65**, 2220 (1989).

Downloaded 01 Mar 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

Growth of GaN on a nanoscale periodic faceted Si substrate by metal organic vapor phase epitaxy

S C Lee,[1] X Y Sun,[1] S D Hersee,[1] J Lee,[1] Y-B Ziang, [2] H Xu[2] and S R J Brueck[1]

[1]Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard, SE, Albuquerque, NM 87106
[2]Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, NM 87131

Abstract
Growth of GaN on a nanoscale periodic faceted Si substrate by metal organic vapor phase epitaxy is reported. The surface of the Si(001) substrate is patterned with a one-dimensional, 355-nm period array of grooves. Each groove consisted of two facing Si{111} stripe facets separated by upper and lower Si(001) surfaces. GaN deposition on this groove array shows a strong selectivity with epitaxial growth of the hexagonal phase on the Si{111} facets being dominant. The initial stage of GaN growth appears to be amorphous. This is followed by GaN containing a high-density of stacking faults to accommodate the large lattice mismatch strain but the defect density decreases as growth proceeds. Lateral overgrowth from the Si{111} facets toward neighboring Si(001) facets leads to coalescence between GaN from opposing Si{111} facets. These GaN regions have misaligned *c*-axes and the crystal structure becomes unstable as they merge, resulting in a phase transition from hexagonal to cubic at the coalescence region. Experimental results of GaN grown on a nanoscale faceted Si surface including: nucleation, crystal structure, and lateral growth depending on Si orientation are presented.

1. Introduction

Gallium nitride (GaN) and related materials have attracted much attention since the achievement of blue laser diodes on a sapphire substrate (Nakamura 1996). Because of their large bandgap, there is also substantial interest for advanced communication systems that require high power and high speed. While sapphire has crystal symmetry identical to

that of (0001) GaN, its large lattice mismatch and low thermal conductivity are less favorable. Silicon (Si) has a strong potential as a substrate for GaN as a result of the availability of low-cost, large-area substrates with extremely high quality. It also has a large lattice mismatch but has a thermal conductivity comparable with GaN that is considerably larger than that of sapphire (Powell 2002). However, GaN on Si(001) is known to have phase instability due to polytypism which leads to a mixture of cubic and hexagonal phases and degrades the crystallinity and surface morphology. For device applications, phase control or phase separation into wide-area single phase GaN on Si(001) is required.

In this work, we investigate the use of nanoscale patterning of a Si substrate to control the phase instability. It is known that GaN exhibits a hexagonal phase on Si(111) and can have cubic phase on Si(001) (Lei 1991, Godlewski 1996, Kawaguchi 1998, Kung 1999, Zamir 2002). The basic idea is to prepare both nanoscale-wide (111) and (001) facets on a single Si(001) substrate. On this multi-faceted surface, both GaN phases are available at the same time but in different areas of the Si substrate. Then, the nucleation mechanisms and crystal structure of GaN which depend on Si facet orientation can be directly compared. We report two important experimental results: one is nucleation selectivity of GaN depending on Si orientation and associated lateral growth, and the other is phase transition of GaN from hexagonal to cubic phase on a nanoscale faceted (nanofaceted) Si surface. Nanoscale faceting (nanofaceting) on a Si substrate and metal-organic vapor-phase epitaxy (MOVPE) of GaN will be described. Experimental results obtained from GaN grown on this nanofaceted Si surface including orientation-dependent nucleation, crystal structure, lateral growth, and formation of cubic phase GaN, will be presented.

2. Experiment

With large-area interferometric lithography and anisotropic wet etching, a nanoscale one dimensional (1D) periodic grooves which were bounded by (100) and {111} facets was fabricated on Si(100) substrate (Brueck 1998, Lee and Brueck 2003). Figure 1(a) shows 45°-tilted (top) and cross sectional (bottom) scanning electron microscopy (SEM) images of a 355-nm period nanofaceted Si substrate. As seen in Fig. 1(a), each period has both top and bottom (100) facets and {111} sidewalls; the overall shape is a flat-bottom groove. The widths of the top and the bottom (100) facets are approximately 100 and 120 nm, respectively.

On this nanofaceted Si surface, GaN was deposited by MOVPE using $(CH_3)_3Ga$ and NH_3. First, a ~1.5-nm thick GaN buffer was grown at 650°C. On top of this buffer, an ~600-nm-thick GaN epilayer was grown at 1050°C. Figure 1(b) shows a cross sectional SEM image of the as-grown sample. This sample was examined with transmission electron microscopy (TEM), atomic force microscopy (AFM), and photoluminescence (PL).

3. Results and discussion

Figure 2(a) is a cross sectional TEM (XTEM) image of the GaN grown on the nanofaceted Si surface. Figures 1(d) and 1(e) are an AFM and a top view SEM image revealing surface morphology of the as-grown GaN, respectively. Some surface roughening due to an irregular corrugation running across the groove direction was observed in Figs. 1(d) and 1(e), and also in Fig. 2(a). However, this is significantly flat and smooth compared with the GaN grown on an unfaceted large-area Si(001) substrate under the same growth conditions [Fig. 1(f)] and with the reported results (Lei 1991, Godlewski 1999, Zhang 1999). This improvement of surface morphology is a major difference between GaN epilayers grown on nanofaceted and unfaceted Si(001).

In Fig. 1(c), which is the magnified image of the region marked by the white box in Fig. 1(b), a void is observed on the bottom Si(001) facet. The shape and location of the voids are more clearly revealed in the XTEM images shown in Fig. 2. Voids are formed on both the bottom and the top Si(001) facets. In Fig. 2(b), the overall shapes of these voids are approximately triangular with GaN$\{1\bar{1}01\}$ sidewalls on the bottom Si(001) and a GaN{0001} on the top Si(001) facet.

Figure 3(a), corresponding to the area in box a in Fig. 2(b), is a high resolution XTEM image taken near the corner of the groove where the bottom Si(001) and Si(111) intersect. The amorphous-like GaN which could be due to the stress from the large lattice mismatch appears at the beginning of deposition but evolves to a crystalline structure with

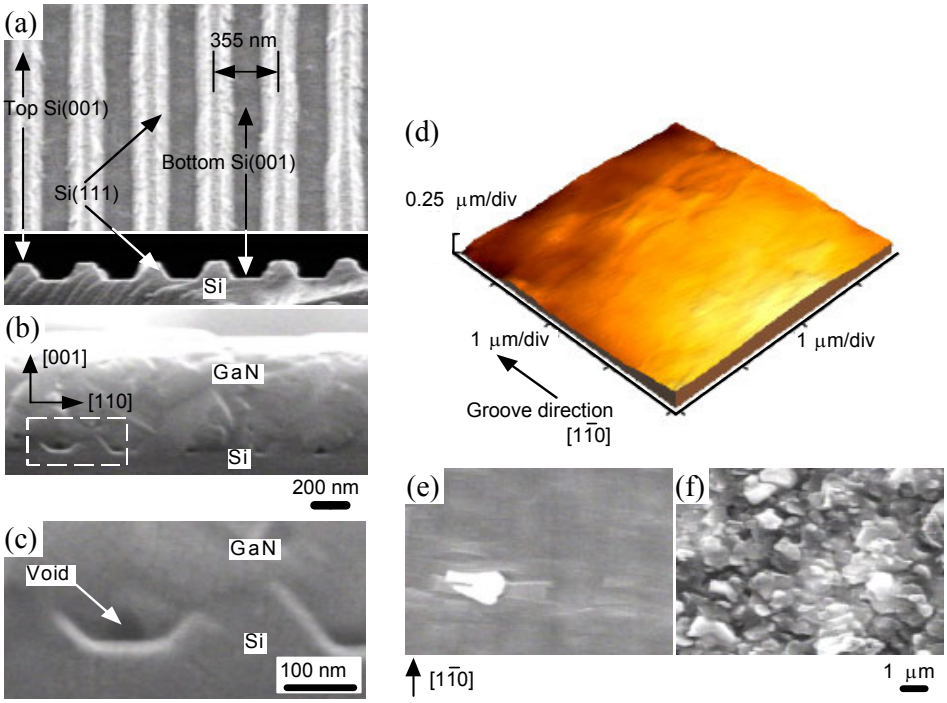


Figure 1. (a) A 45°-tilted (top) and a cross sectional (bottom) SEM image of a 355-nm period 1D array of flat-bottom grooves fabricated on a Si(001) substrate. (b) A cross sectional SEM image of a 600-nm thick GaN film grown on a nanofaceted Si surface shown in (a). (c) A magnified SEM image of the white dashed box in (b). (d) An AFM image of the as-grown sample. (e) A top view SEM image of GaN shown in (b). (f) A top view SEM image of GaN grown on a wide-area unfaceted Si(001) under the same growth conditions as those of (b).

a large number of stacking faults after a few nm of growth. Stacking faults are generated partially to release the excess strain energy (Zubia 2000). Figure 3(b), corresponds to the area in box b of Fig. 2(b), and shows a defect-free hexagonal phase GaN region formed above the stacking faults. The *c*-axes in the hexagonal GaN phases are perpendicular to the Si{111} facets as indicated by the arrows in Fig. 2(b). Figure 3(a) shows that after GaN nucleation on the Si{111} facets, there is lateral growth towards the adjacent Si (001) facets. Figure 3(c) corresponds to the area in box c of Fig. 2(b) and shows a high resolution XTEM image taken near the void at the top Si(001) stripe facet. As seen in this figure, the two *c*-axes of the coalesced GaN are misaligned since they are originated from opposing Si(111) faces and make an angle of ~70°. Figure 3(d) is the high-resolution image of the area in box d of Fig. 2(a). The cubic phase GaN in Fig. 3(d) was consistently formed above the coalescence boundary of the highly defected hexagonal phase regions.

Figure 4 is a room-temperature PL spectrum obtained from the as-grown sample. The PL exhibits a single peak at 3.389 eV which is on a background of scattered laser light. This is close to the reported PL peak energy of the transition of a donor-bound exciton in hexagonal phase GaN on Si(111) (Zamir 2002, Yodo 2002). This also implies that the major crystal structure of the as-grown GaN on a nanofaceted Si surface is hexagonal and supports the TEM results.

Based on the experimental results, the growth evolution outlined schematically in Fig. 5 is proposed. As suggested in Fig. 5(a), nucleation selectively occurs on Si{111} with a hexagonal GaN phase forming at this initial growth stage. From the shape of the voids, it can be noticed that the GaN selectively grown on Si{111} has a trapezoidal cross section with $\{1\bar{1}01\}$ as the sidewall orientation at the beginning of growth. This orientation selectivity results in negligible deposition on the Si(001) facets, but lateral growth from Si{111} toward these facets begins to fill in these facets. The arrowheads in Fig. 5(b) indicate the directions of lateral growth conjectured from the TEM image of Fig. 2. From

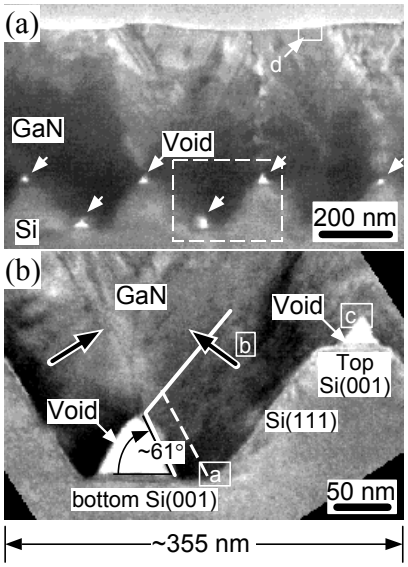


Figure 2. (a) An XTEM image of the as-grown sample. (b) A magnified XTEM image corresponding to the white dashed box in (a).

the facet geometry, the directions available for lateral growth on the top Si(001) are [0001] and $[1\bar{1}01]$ while that on the bottom Si(001) is $[1\bar{1}01]$ only. Thus the sidewall orientation of the void on the top Si(001) is determined by the lateral growth along the additional [0001].

In Figs. 2, the voids are not at the centers of the Si(001) facets. The gas flow in our horizontal reactor which directs gases from left to right on these figures may result in uneven exposure of the opposing Si{111} facets to the source gases. Since the bottom void forms early in the growth and effectively halts further growth in this region by sealing out any gases, the growth rates depending on facet orientation can be properly compared at the bottom side. In Fig. 2(b), growth proceeded from the Si(111) to the solid white line profile along [0001] and from the dashed white line in a $[1\bar{1}01]$. By applying the same rule to the other side of the groove and averaging the results of the two cases, the growth rate in $[1\bar{1}01]$ is estimated as at least 20-25% of that along [0001]. If migration of Ga atoms from $\{1\bar{1}01\}$ to (0001) on the GaN surface is appreciable in nanoscale dimensions, the actual growth rate along [0001] can be partly due to the narrower width of the upper side of the trapezoidal cross section and could be significantly greater than the nominal overall growth rate of 10 nm/min. Owing to the considerable lateral growth rate in $[1\bar{1}01]$, a hexagonal single phase GaN epilayer can be achieved over the wide area (probably μm-scale area) on Si(001) from a seed layer on a Si{111} facet if the lateral dimension of the top and bottom Si(001) facets were to be extended (e. g. a larger period structure).

Continual growth with fast growth rate along [0001] induces the coalescence between GaN epilayers selectively grown on facing Si{111} and, as a result, generates a void below the coalescence points of Fig. 5(b) where growth is stopped at the inside.

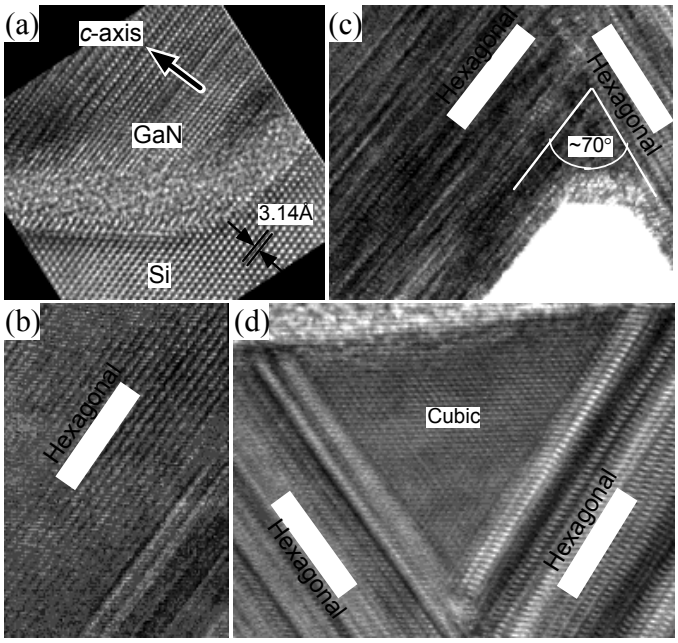


Figure 3. High resolution XTEM images alphabetically labeled (a to d) in the white boxes of Fig. 2. See the text for details.

Although initial deposition begins on localized Si{111} facets, the growth surface is eventually changed to GaN after coalescence, as illustrated in Figs. 5(b) and 5(c). Continual deposition on this GaN surface activates a planarization process leading to the flat surface shown in Figs. 1(d) and 1(e).

In Fig. 3(c), coalescence between adjacent hexagonal GaN nuclei induces a structural instability resulting from the physical contact of these two hexagonal phase GaN epilayers with misaligned *c*-axes. As seen in Figs. 3(b) to 3(d), the hexagonal phase changes to a cubic phase above a void where coalescence proceeds. A growth region with a high density of stacking faults is observed again just as the two opposing hexagonal islands meet. The generation of stacking faults could be understood as phase fluctuation between hexagonal and cubic phases, partly due to a high strain in the region of coalescence that will increase total crystal energy by structural instability.

The appearance of the cubic phase GaN is an important result of this work which suggests a possibility of phase separation or phase control on Si(001) by nanofaceting. Formation of cubic phase GaN which is similar to Fig. 3(d) has been reproduced on a periodic V-grooves fabricated in a Si(001) surface and will be reported elsewhere. In this

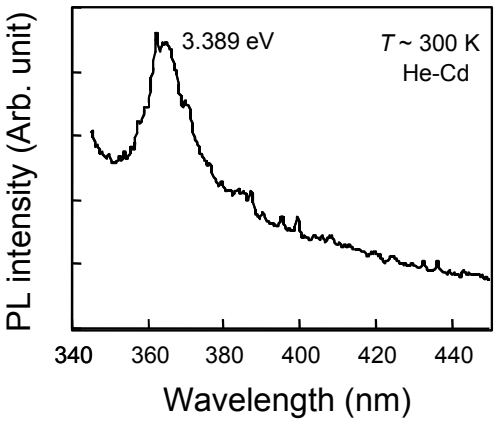


Figure 4. Room-temperature PL of the as-grown sample.

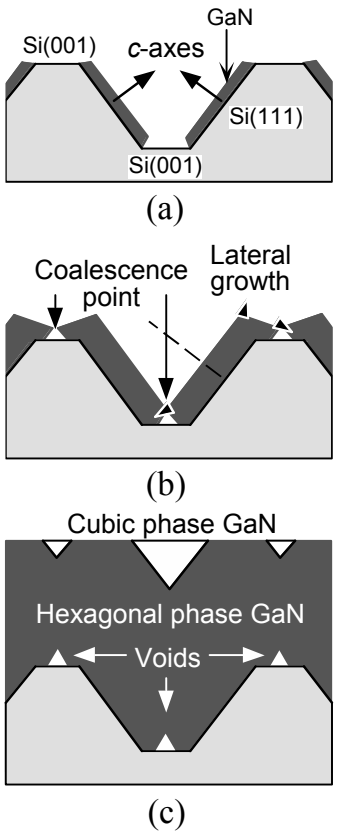


Figure 5. An illustration for orientation-dependent nucleation and lateral growth of GaN on a nanofaceted Si surface.

work, we briefly introduce the formation of cubic phase GaN on a Si(100) substrate by nanofaceting. Further research is required to understand the relation between facet geometry/physical dimension and the relative size of the cubic phase regions on a nanofaceted Si surface. Planarization of the front surface of GaN which is 1D periodic composite of hexagonal and cubic phases also needs more investigation.

4. Summary and conclusions

Growth of GaN on a nanoscale periodic faceted Si(001) substrate by MOVPE has been investigated. On a 355-nm period 1D array of flat-bottom grooves consisting of a bottom Si(001), a top Si(001), and Si{111} sidewall facets in each period, strong selectivity of the nucleation and incorporation of GaN on Si(111) with hexagonal phase has been observed. Orientation-dependent nucleation results in lateral growth from a Si{111} facet toward neighboring bottom and top Si(001) facets with formation of voids, and eventually coalescence. With continued growth, the entire GaN film planarizes to a smooth surface but has an unstable crystal structure resulting from the periodically misaligned *c*-axes. This structural instability causes phase transition from hexagonal to cubic phase on a nanofaceted Si surface. Our experimental results therefore suggests that nanofaceting can be used for GaN phase control or phase separation on a local or a wide area of a Si(001) substrate.

References

Brueck S R J, Zaidi S H, Chen X and Zhang Z 1998 Microelec. Engineering, **41/42**, 145

Godlewski M, Goldys E M, Philips M R, Bergman J P, Monemar B, Langer R and Barski A 1996 MRS Internet J. Nitride Semicond. Res. **3**, 51

Kawaguchi Y, Honda Y, Matsushima H, Yamaguchi M, Hiramatsu K and Sawaki N 1998 Jpn. J. Appl. Phys. **37**, L966

Kung P, Walker D, Hamilton M, Diaz J and Razeghi M 1999 Appl. Phys. Lett. **74**, 570

Lee S C and Brueck S R J 2003 J. Vac. Sci. Technol. (submitted)

Lei T, Fanciulli M, Molnar R J, Woustakas T D, Graham R J and Scanlon J 1991 Appl. Phys. Lett. **59**, 944 .

Nakamura S, Senoh M, Nagahama S, Iwasa N, Yamada T, Matsushita T, Kiyoku H and Sugimoto Y 1996 Jpn J. Appl. Phys. **35**, L217.

Powell A R and Rowland L B 2002 Proc. IEEE, **90**, 942

Yodo T, Ando H, Nosei D, Harada Y and Tamura M 2002 J. Cryst. Growth, **237-239**, 1104

Zamir S, Meyler B, Salzman J, Wu F and Golan Y 2002 J. Appl. Phys. **91**, 1191

Zhang X, Chua S J, Feng Z C, Chen J and Lin J 1999 Phys. Stat. Sol. (a) **176**, 605

Zubia D, Zaidi S H, Brueck S R J and Hersee S D 2000 Appl. Phys. Lett. **76**, 858

HTML ABSTRACT + LINKS

Spatial phase separation of GaN selectively grown on a nanoscale faceted Si surface

S. C. Lee,[a] X. Y. Sun, S. D. Hersee, and S. R. J. Brueck
Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

H. Xu
Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131

(Received 12 November 2003; accepted 23 January 2004)

This letter reports the growth of spatially separated hexagonal and cubic phases of GaN on a patterned Si(001) substrate by metalorganic vapor-phase epitaxy. The substrate surface was patterned with grooves having a 355 nm period. Each groove consisted of two opposed Si{111} facets that were separated by Si(001) surfaces. Epitaxial growth of GaN on this substrate began selectively on the Si{111} facets and yielded the GaN hexagonal phase. With further growth, the two hexagonal phase regions separately grown on the opposed Si{111} facets coalesced, with strongly misaligned *c* axes (~110°). The GaN grown after coalescence was subsequently confirmed, by transmission electron microscopy and photoluminescence, to be of cubic phase. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1687456]

Gallium nitride (GaN) and related compounds exhibit a wide range of emission wavelengths, and are also attractive for high-power and high-speed devices. The lack of a lattice-matched substrate has led to extensive efforts in the development of substrates and growth strategies for these materials. Silicon substrates are inexpensive and readily available in large areas with very high crystalline quality.

Generally, GaN grows with a hexagonal phase on Si(111) and can have a cubic phase on Si(001).[1–7] The formation of the cubic phase is known to depend on growth conditions and on the crystal structure of the substrate and the buffer layer; the material quality is often severely degraded by mixture with or transition to the hexagonal phase, resulting in a polytype structure.[3–7] Despite its expected advantages in electron mobility, *p*-type doping, electrical contact, and cleaving for laser facets,[7,8] high-quality cubic phase GaN has so far proved elusive.

For high-quality cubic phase GaN on Si(001), the strategy followed in this work is to take advantage of the compliance associated with nanoscale growth as shown in Fig. 1.[9] A nanoscale groove [Fig. 1(a)] is etched into a Si(001) surface, revealing two opposed Si(111) faces separated by a bottom Si(001) facet. Deposition of GaN on this patterned surface results in selective growth of GaN only on the Si(111), leading to two hexagonal GaN crystals in the groove.[10] Continued growth leads to the coalescence between these two GaN epilayers, as illustrated in Fig. 1(b). Since GaN has hexagonal phase on Si(111) with its *c* axis normal to it, the two top (0001) surfaces of Fig. 1(b) form a V groove at coalescence, but with misaligned *c* axes. Following the coalescence, incoming flux interacts with this secondary V groove and forms the most stable crystal structure available under the growth conditions. The investigation of the crystal structure of GaN continuously grown on this artificially manipulated surface was the basic motivation for this work.

A 355-nm-period, nanofaceted Si surface, exhibiting {111} and (001) facets, was fabricated on a Si(001) substrate by large-area, *i*-line interferometric lithography and anisotropic wet etching.[10] Figure 2(a) shows a cross-sectional scanning electron microscopy (SEM) image of the nanofaceted Si surface, similar to Fig. 1(a). For convenience, we call a single period of Fig. 2(a) a flat-bottom groove. The widths of the top and the bottom (001) facets are approximately 100–150 nm and 70–100 nm, respectively, and the groove depth is ~100 nm. GaN epilayers were grown by metalorganic vapor-phase epitaxy. A ~1.5-nm-thick GaN buffer was grown at 650 °C and a ~75-nm-thick GaN epilayer was subsequently deposited at 1050 °C. Nominal growth rates were 1.5 nm/min for the buffer and 10 nm/min for the GaN epilayer.

Figure 2(b) shows a cross-sectional transmission electron microscopy (XTEM) image of a single groove of an as-grown sample. In Fig. 2(b), the groove is filled with only a 75-nm-thick deposition. Triangular voids located on the top and bottom Si(001) facets were observed. These voids imply that there was negligible deposition on Si(001).[10] As indicated by the arrows in Fig. 1(b), the initial growth proceeds selectively in the direction normal to the Si{111} planes. The cross section of the GaN at this point was a trapezoid bounded by {1101} sidewalls and a {0001} top facet that was parallel to the underlying Si{111} facet [Fig. 2(c)]. This cross section was maintained until coalescence between opposing GaN nuclei occurred at point X [Fig. 2(b)]. The dashed lines indicate the profile of the secondary V groove at the moment of coalescence. After coalescence, the supplied gas flow no longer accessed the space below point X, and the growth inside the lower void stopped. Continued growth eventually resulted in a second coalescence above the upper Si{001} facet, again with the formation of a void. In this early stage of growth, the crystal structure over this second coalescence

[a]Electronic mail: sclee@chtm.unm.edu

 © 2004 American Institute of Physics
Downloaded 17 Mar 2004 to 64.106.37.74. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

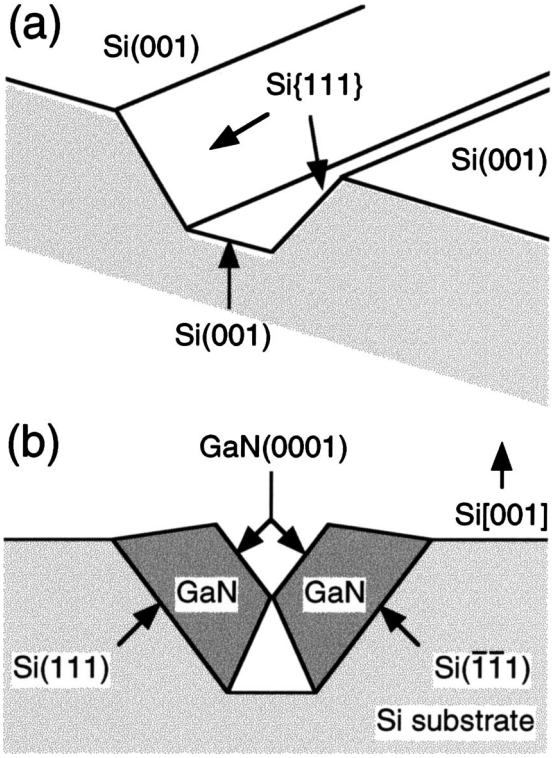


FIG. 1. (a) Concept drawing of the faceted Si surface prepared for this work. (b) Cross section of the faceted Si surface shown in (a) along with selective growth of GaN on the Si(111) faces, leading to the formation of the secondary V groove.

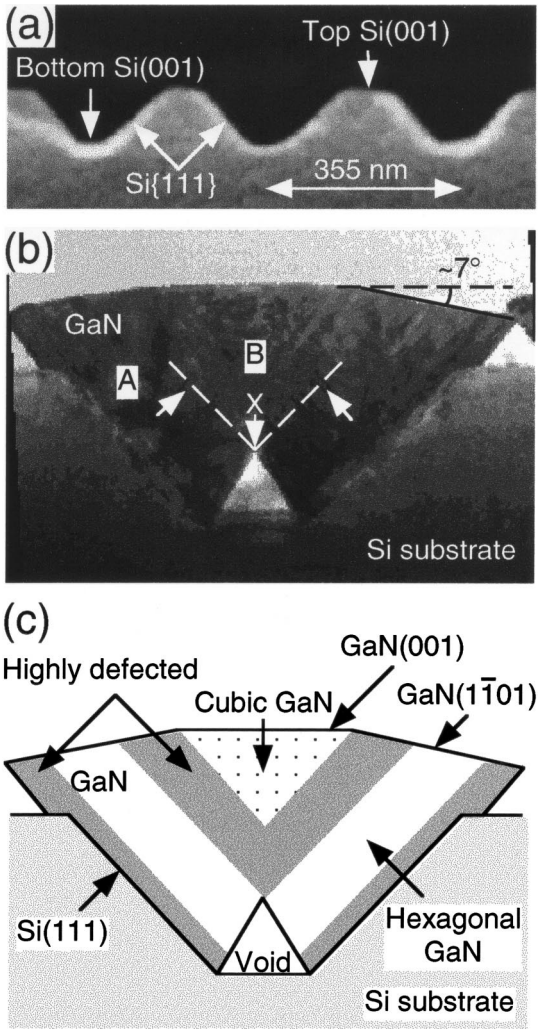


FIG. 2. (a) Cross-sectional SEM image of a 355-nm-period one-dimensional nanofaceted surface fabricated on a Si(001) substrate. (b) XTEM image of a single groove of a 75-nm-thick as-grown GaN sample. The point X corresponds to the initial coalescence point and formation of the secondary V groove. The surface of the GaN exhibits facets that are 7° off from the Si(001). (c) Phase map of the cross section shown in (b), showing regions of hexagonal and cubic crystals separated by highly defected SF regions.

is not yet well defined, and we concentrate on the material properties in the vicinity of point X.

Figure 2(c) illustrates the GaN phase map resulting from analysis of the sample cross section shown in Fig. 2(b). Figure 3 shows high-resolution XTEM images of selected areas of the GaN growth labeled (A,B) in Fig. 2(b). In Fig. 3(a), taken from region A, stacking faults are largely distributed near the GaN/Si interface, but the crystal structure stabilizes to a hexagonal phase that is maintained until coalescence, as shown subsequently [Fig. 3(c)]. Stacking faults (SFs) are often observed in GaN/Si heterostructures, especially when the substrate is patterned at nanoscale.[11] Since a SF in the cubic phase corresponds to the presence of the hexagonal phase (and vice versa), a region of SFs can be regarded as a phase fluctuation or an aperiodic variation of the crystal structure between hexagonal and cubic phases.

Figure 3(b) shows a high-resolution XTEM image from a region near B of Fig. 2(b) that corresponds to the region grown after coalescence of the secondary V groove. The GaN, again, is highly defected with stacking faults in the coalescence region between the two hexagonal crystals, but the resulting crystal structure now is cubic. Selective area electron diffraction (SAD) focused on the GaN region [the inset of Fig. 3(b)] provides evidence of the existence of the cubic phase. The cubic phase diffraction spots, indicated by the black arrows labeled ''C,'' are shown together with those from the two overlapped hexagonal phases (the white arrows labeled ''H1'' and ''H2''). Note that the crystal axes of the cubic phase GaN are parallel to those of the Si substrate. Thus, the planar surface above region B is parallel to the Si{001} surface. The resulting cubic phase region, as illustrated in the phase map of Fig. 2(c), has an isosceles triangle cross-sectional shape that mostly fills the initial grooved surface, and is therefore spatially separated from the hexagonal phase regions on the nanofaceted Si surface.

Figure 3(c) shows details of coalescence revealed in high-resolution XTEM. The coalescence boundary between the two hexagonal phases is clearly resolved. The white arrows in Fig. 3(c) indicate the *c*-axis directions of the two hexagonal GaN epilayers grown from the Si{111} facets. The *c* axes of the misaligned hexagonal nuclei are tilted by ~110° with respect to each other; this situation is analogous to a twin structure. The boundary region shown in Fig. 3(c) is energetically very unstable. The black arrows indicate several lattice points that are significantly displaced along this coalescence boundary. As shown in Figs. 2(c), 3(b), and 3(c), this highly strained coalescence boundary between hexagonal nuclei is gradually overtaken by phase fluctuation (generation of SFs), and finally by the formation of the cubic GaN phase. Experimentally, it is evident from Figs. 3(b) and 3(c) that changing the crystal structure to cubic phase is energetically more favorable than keeping misaligned hexago-

Downloaded 17 Mar 2004 to 64.106.37.74. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

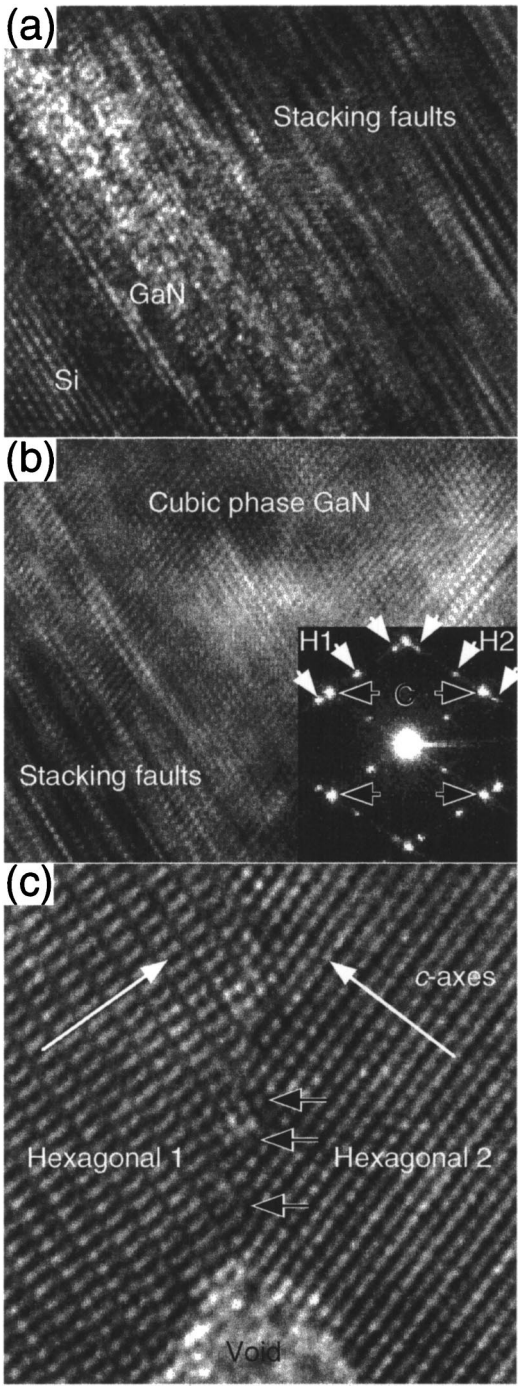


FIG. 3. High-resolution XTEM images taken from the region (a) A, (b) B, and (c) the region near the X, as indicated in Fig. 2(b). The inset of (b) is a SAD pattern obtained from the region above coalescence point X of Fig. 2(b). In the inset, H1 and H2 are the diffraction spots from the two misaligned hexagonal phases, while C marks the cubic phase. In (c), the arrows indicate the *c* axes of the two coalesced, but misaligned, hexagonal phases.

nal phases. A relatively small energy difference between the two phases of GaN can be one of the physical origins driving phase transition to lower the total crystal energy.[12]

Finally, Fig. 4 shows a photoluminescence (PL) spectrum obtained from the as-grown sample at 10 K. In Fig. 4, two dominant PL peaks are observed at 3.439 and 3.276 eV.

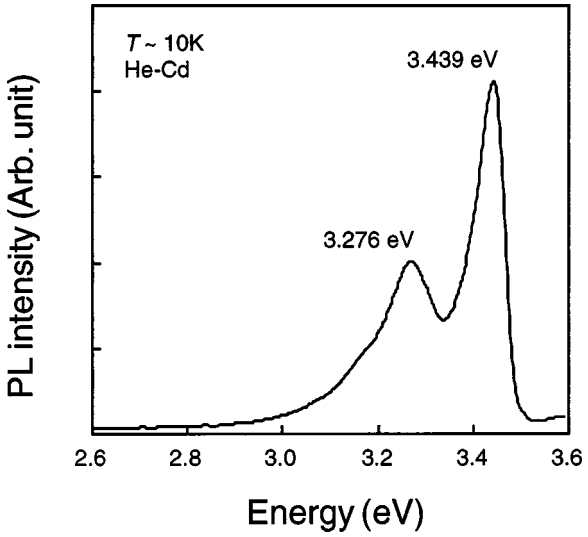


FIG. 4. PL of the as-grown sample.

While 3.439 eV is between donor bound exciton recombination (D^0X) and band–impurity transition in hexagonal phase GaN on a Si(111), 3.276 eV is very close to D^0X recombination of cubic phase GaN on GaAs(001) or Si(001).[5,7] Therefore, the PL of Fig. 4 provides further evidence of the coexistence of hexagonal and cubic phases in a single GaN epilayer and supports the XTEM observation of Figs. 2(b) and 3.

In conclusion, spatial separation between hexagonal and cubic phases of GaN has been observed on a nanoscale secondary V groove that results from selective growth on a nanopatterned Si(001)/Si(111) flat-bottomed groove on a Si(001) substrate. This technique can be applied to the growth of a particular phase GaN on either a local or a wide area of a Si(001) substrate, and therefore to the study of properties of cubic phase GaN.

The authors wish to acknowledge the support of DARPA and of ARO/MURI.

[1] P. Kung, D. Walker, M. Hamilton, J. Diaz, and M. Razeghi, Appl. Phys. Lett. **74**, 570 (1999).
[2] Y. Kawaguchi, Y. Honda, H. Matsushima, M. Yamaguchi, K. Hiramatsu, and N. Sawaki, Jpn. J. Appl. Phys. **37**, L966 (1998).
[3] Y. Honda, Y. Kuroiwa, M. Yamaguchi, and N. Sawaki, J. Cryst. Growth **242**, 77 (2002).
[4] S. Zamir, B. Meyler, J. Salzman, F. Wu, and Y. Golan, J. Appl. Phys. **91**, 1191 (2002).
[5] T. Yodo, H. Ando, D. Nosei, Y. Harada, and M. Tamura, J. Cryst. Growth **237–239**, 1104 (2002); T. Lei, M. Fanciulli, R. J. Molnar, T. D. Woustakas, R. J. Graham, and J. Scanlon, Appl. Phys. Lett. **59**, 944 (1991).
[6] M. Godelwski, E. M. Goldys, M. R. Philips, J. P. Bergman, B. Monemar, R. Langer, and A. Barski, MRS Internet J. Nitride Semicond. Res. **3**, 51 (1996).
[7] J. W. Orton and C. T. Foxon, Rep. Prog. Phys. **61**, 1 (1998).
[8] M. Asif Khan, Q. Chen, J. Yang, Z. Anwar, M. Blasingame, and M. S. Shur, Tech. Dig. Int. Electron Devices Meet. **96**, 27 (1996).
[9] D. Zubia and S. D. Hersee, J. Appl. Phys. **85**, 6492 (1999).
[10] S. C. Lee, X. Y. Sun, S. D. Hersee, and S. R. J. Brueck (unpublished).
[11] D. Zubia, S. H. Zaidi, S. R. J. Brueck, and S. D. Hersee, Appl. Phys. Lett. **76**, 858 (2000).
[12] C.-Y. Yeh, Z. W. Lu, S. Froyen, and A. Zunger, Phys. Rev. B **46**, 10086 (1992).

Downloaded 17 Mar 2004 to 64.106.37.74. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

HTML ABSTRACT + LINKS

Equilibrium crystal shape of GaAs in nanoscale patterned growth

S. C. Lee and S. R. J. Brueck
Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

(Received 5 December 2003; accepted 9 April 2004)

The equilibrium crystal shape (ECS) of GaAs homoepitaxially grown on a nanoscale SiO_2-patterned (001) plane by molecular beam epitaxy is investigated. A GaAs epilayer selectively grown on a nanoscale area bounded by a circular SiO_2 mask undergoes faceting, resulting in a pyramidal shape with {110} sidewalls. Growth is slowed or terminated with the generation of these {110} facets even with a continuing supply of Ga atoms. This implies that the pyramidal shape is energetically very stable. Based on experimental results and the Wulff construction, a {110}-type sidewall pyramid is proposed as an ECS of GaAs on (001) in nanoscale patterned growth. © *2004 American Institute of Physics.* [DOI: 10.1063/1.1757657]

I. INTRODUCTION

A crystal has a certain shape in equilibrium. In crystal growth, the shape of a growing crystal is governed by equilibrium crystal shape (ECS) under a given set of growth conditions. In other words, the shape of a growing crystal evolves to the ECS for the given growth conditions. There has been a theoretical approach to ECS since Wulff proposed a construction rule based on minimization of the total surface free energy more than a hundred years ago.[1] Extensive experimental research work has been reported from the 1980's.[2]

The technology of crystal growth has progressed remarkably over the last three decades. Advanced growth techniques such as molecular beam epitaxy (MBE), providing nearly atomic-layer thickness control, are widely available. In traditional MBE, the substrate is generally centimeter scale while the thickness of the epilayer is typically in the micrometer scale. As a result of the extremely large ratio of the lateral dimension to the thickness and of the very low growth rate [~1 monolayer (ML)/sec], the growing epilayer maintains a quasi-two-dimensional (2D) growth mode and its crystal shape is trivially determined by the shape of the substrate surface. For example, it is always a thin disk if the substrate has a circular shape. If the radius of the substrate opening is reduced from centimeter scale to micrometer or even to nanometer scale so that the lateral dimensions of the epilayer become comparable to or less than its thickness, the ECS will be changed dramatically.

To understand the impact of the variation of the lateral dimension of a growth area on the overall shape of the epilayer, we can imagine selective growth of GaAs on a SiO_2-patterned substrate. If a circular hole with diameter much larger than the thickness of the epilayer to be deposited is fabricated on a SiO_2 mask, growth selectively proceeding within this hole area typically results in a disk shape with a trapezoidal cross section as shown in Fig. 1. If the hole diameter becomes comparable to or smaller than the thickness of the epilayer to be deposited on the hole, as seen in Fig. 1, the overall shape of the epilayer will be strongly affected by the sidewall formed near the boundary of the SiO_2 mask and the epilayer. Depending on the deposition thickness, faceting occurs on the sidewalls and the resulting epilayer shape is governed by ECS.

The condition that the deposition thickness be greater than the lateral dimension of an epitaxial area is easily satisfied in nanoscale patterned growth without requiring unusually thick depositions. In selective patterned growth using a mask to delineate individual nanoscale growth areas, an epitaxial area with a lateral size less than the surface diffusion length of a Ga atom at typical MBE growth conditions allows most of Ga atoms incident on a substrate surface to migrate over the whole nanoscale surface area to find a favorable bonding configuration encouraging the formation of the energetically stable shape.[3]

Evaluation of the ECS is important in determining the surface-free-energy hierarchy among crystal orientations and also in understanding the electronic properties of nanostructures such as quantum dots that are affected by geometrical shape. MBE is generally known as a nonselective deposition technique, but we have demonstrated selective growth for GaAs on a SiO_2-masked GaAs substrate at high growth temperature ($\gtrsim$615 °C) and low deposition flux ($\lesssim$0.1 ML/s).[4] In this work, the ECS of GaAs using selective growth on a nanoscale SiO_2-masked substrate.

II. NANOSCALE PATTERNING AND MBE GROWTH

Figure 2(a) shows a 45°-tilted scanning electron microscopy (SEM) image of a 355-nm period 2D array of circular holes fabricated in a ~50-nm-thick SiO_2 mask on a GaAs(001) substrate by interferometric lithography and dry etching. Three ranges of hole diameter d were prepared on a single substrate: 80–130 nm, 160–200 nm, and 230–260 nm. Figure 2(a) corresponds to the first range for which the smallest hole diameters are less than 100 nm.

On this patterned substrate, we deposited an amount of GaAs corresponding to a 100-nm-thick layer (on an unpatterned wide area) at the growth temperature of 630 °C with deposition flux of ~0.1 ML/s under As-rich conditions (As_2/Ga beam equivalent pressure ~35) where selective growth has been reported.[5] The range of hole diameters

 © 2004 American Institute of Physics

Downloaded 22 Jul 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

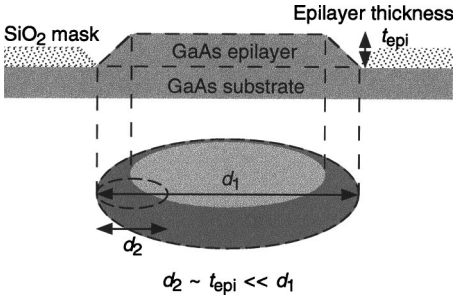


FIG. 1. Cross section (top) and tilted (bottom) schematic views of selective growth of GaAs on a SiO_2-patterned substrate.

spans this deposition amount. Figure 2(b) presents a 45°-tilted SEM image of the GaAs epilayer selectively grown on the SiO_2-patterned substrate shown in Fig. 2(a).

Because of the variations in hole diameter and roughness resulting from the lithography, the GaAs islands show various shapes. Categorization into a few groups is discussed in Sec. III. The insets of both figures show a single circular hole and a nanoscale faceted GaAs epilayer selectively grown in such a hole (GaAs epi-island), respectively.

III. RESULTS AND DISCUSSION

Figure 3(a) shows a top-view SEM image of an as-grown sample for $d \sim 80$ to 130 nm. Various GaAs island shapes are observed: circular, triangular, square, rectangular, and pentagonal. The top-view images in Fig. 3(b) show some of them for mask holes of $d \sim 130$ nm. As indicated by the dashed circle in each image, which represents the original substrate surface opening in the SiO_2 mask, some GaAs epi-islands exhibit lateral growth over the SiO_2 mask, primarily

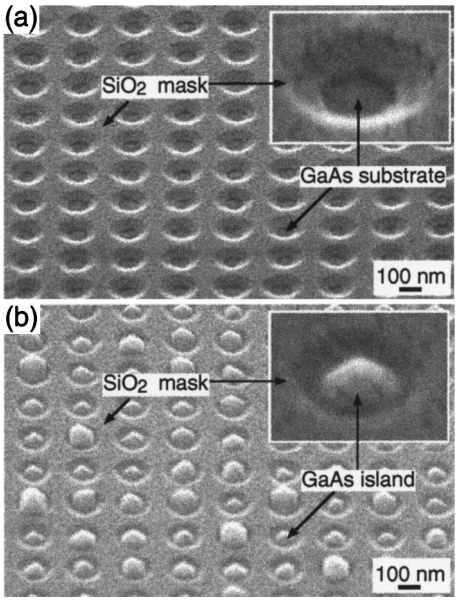


FIG. 2. (a) A 45°-tilted SEM image of a 355-nm period, 2D array of circular holes fabricated in a ~50-nm-thick SiO_2 mask on a GaAs(001) substrate. (b) A 45°-tilted SEM image of a GaAs epilayer selectively grown on the SiO_2-patterned substrate shown in (a). The insets of (a) and (b) show a single hole and an island-type GaAs epilayer selectively grown in a circular hole, respectively.

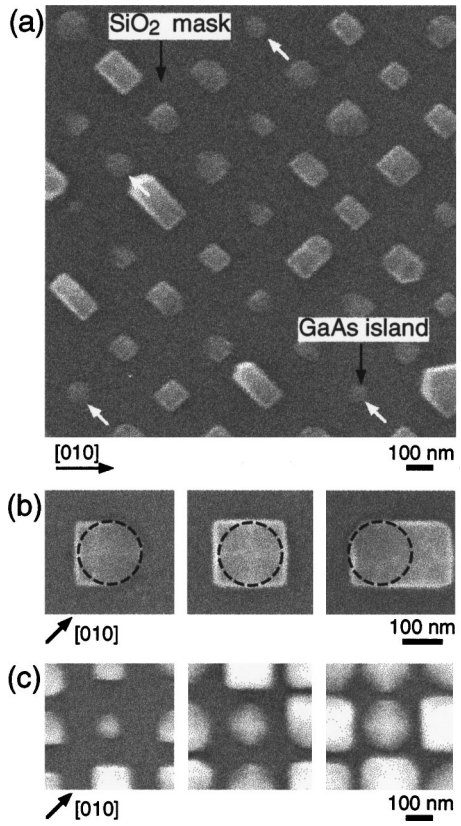


FIG. 3. (a) A top-view SEM image of an as-grown sample of $d \sim 80-130$ nm. White arrows indicate the GaAs islands belonging to the group of islands with minimal lateral growth. (b) A top-view SEM image of GaAs islands of $d \sim 130$ nm, having triangular (left), square (center), and rectangular shapes (right). (c) A top-view SEM image of circular shape GaAs islands of $d \sim 100$ (left), 180 (center), and 260 nm (right).

along [100] and [010] in resulting directions. The three-dimensional (3D) shape of an epi-island similar to that in the center of Fig. 3(b) has been discussed in our previous work.[4] The 3D shape of the GaAs epi-islands shown in Fig. 3(b) is complex and all of the facet orientations are difficult to measure. On the other hand, as shown in Fig. 3(c), growth of some epi-islands retains the roughly circular shape of the initial mask opening without noticeable lateral growth and their simple 3D shape enables characterization of their facet orientations. As indicated by the white arrows, these circular shape islands are relatively common with ~10%–15% of the GaAs islands belonging to this class. Also, as seen in Fig. 3(c), this shape, which exhibits at least twofold symmetry, is observed over the entire range of hole diameters investigated. In this work, we focus on the islands with minimal lateral overgrowth to investigate the ECS of a minimal-volume GaAs epi-island selectively grown in a nanoscale area. In Fig. 3(c), GaAs epi-islands with other shapes are distributed around the circularly shaped ones we are discussing. Additional discussion of these GaAs epi-islands will be presented below.

Figure 4 shows high-magnification top-view and 45°-tilted (top) and side view SEM images (bottom) of a GaAs island for $d \sim 180$ nm taken from different directions. For this figure, the SiO_2 mask was removed by dipping the as-grown sample in diluted HF to allow unobstructed views of the

Downloaded 22 Jul 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

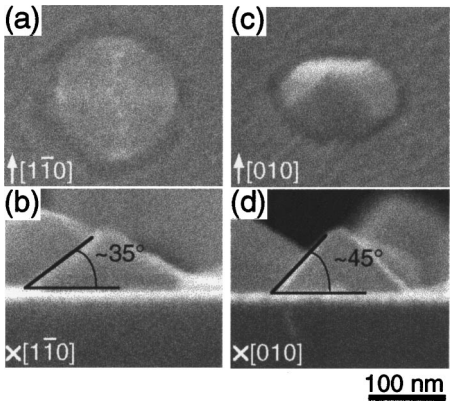


FIG. 4. High-magnification SEM images of a GaAs epi-island for $d \sim 180$ nm taken from several directions with different view angles. The island is similar to that shown in Fig. 3(c) center. Top row are 45°-tilted images; bottom row are side view images. Note the view direction indicated in each part.

GaAs island. The island shown in Fig. 4 belongs to the group of the islands with minimal lateral overgrowth [it is similar to the epi-island in the center image of Fig. 3(c)]. Islands smaller than that of Fig. 4 were too small to obtain highly resolved surface images for clear identification of facet orientation with our field emission SEM. As seen in Fig. 4(a), the top view shows a nearly circular outline, but the side view along $[1\bar{1}0]$ in Fig. 4(b) shows a clear triangular-faceted shape with a ~35° angle of the sidewall measured from the base [or (001)] plane. Additionally, the GaAs island in Fig. 4(c) reveals clear facets on its surface in a 45°-tilted image taken along [010]. As seen in Fig. 4(d), its side-view shape along the same direction is also triangular but with a ~45° angle of the sidewall facet measured from the base. The slight differences in contrast on the island surface of Fig. 4(a) correspond to the different facets.

From the four images of Fig. 4, the 3D shape of the group of GaAs islands we are focusing on is pyramidal. The angle measurement along various directions reveals that the orientation of the sidewall facets on a GaAs island is {110}. A schematic illustration of the {110}-type sidewall pyramid is shown in Fig. 5. Given this shape, we can calculate the volume of the GaAs island and the deposition required for its formation. The maximal and minimal volumes were estimated by equating d alternatively to a diagonal across the pyramid base and to a side of the square pyramid base, respectively. Dividing these volumes by the base area (a circle of $d \sim 180$ nm), a range of the effective deposition thicknesses of 14 to 38 nm was obtained. This rough approximation reflects the actual shape of the GaAs island where clear faceting occurs near the top of the pyramid while the base still seems to be close to circular.[6] Importantly, this range of effective deposition thicknesses is considerably smaller than the supplied amount corresponding to a 100-nm-thick layer. This discrepancy provides a very important experimental result on the incorporation behavior of the incident Ga flux in nanoscale-patterned growth.

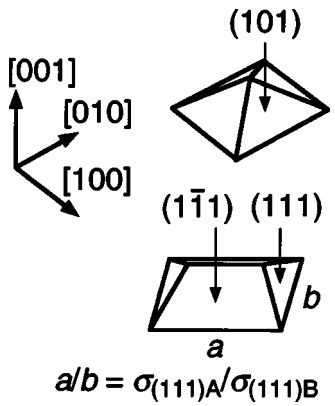


FIG. 5. Schematic diagrams of a {110}-type sidewall pyramid (top) and a {111}-type sidewall pyramid elongated along [110] (bottom).

In conventional MBE (unpatterned, or patterned on micrometer or larger scales) where the deposition thickness is much smaller than the lateral size of the epitaxial area, the deposition amount on any unit area is simply proportional to the growth time and should be the same as the amount supplied if the incorporation rate is unity. At our growth condition, an incorporation rate close to unity was confirmed on a wide-area unpatterned GaAs surface. The larger-volume GaAs islands of Fig. 3(a) are also indirect evidence of this higher incorporation rate; other research groups also have reported unity sticking coefficients of Ga on GaAs (001) around 630 °C.[7] However, as demonstrated above, the effective deposition thickness is significantly smaller than the supplied amount in nanoscale patterned growth for the class of circular epi-islands without significant lateral overgrowth on which we are focusing. This implies that growth on a nanoscale hole area is slowed or terminated once a stable facet such as {110} is generated on the surface of the epilayer. In other words, the formation of an energetically very stable facet provides a thermodynamic barrier for additional incorporation, which would require transformation from the stable facet to another facet (and another geometric shape). Thus, the slowing or termination of growth with the formation of {110}-type sidewalls on a GaAs island suggests that the pyramid observed in Fig. 4 is a very stable geometric shape under our growth conditions. Unincorporated Ga atoms may evaporate from an island if their sticking behavior on {110} is different from that on (001), or they may migrate to and evaporate from an adjacent SiO_2 mask since the Ga sticking coefficient on the SiO_2 surface is negligible under selective growth conditions.

For a given circular base, both orthorhombic and pyramidal shape GaAs islands can be grown with low-index orientation faceting. However, the orthorhombic shape is not available since the slanted sidewall generated at the initial stage of growth as a result of the mask thickness does not allow vertical sidewalls.[5,8] Then, the only possible geometric shape is pyramidal. Both {110} and {111} sidewalls are feasible. Experimentally, however, {111}-type sidewall pyramids were not observed. As seen in Fig. 4, only {110}-type sidewall pyramids are generated.

According to the Wulff construction rule, the crystal shape is determined through the minimization of total surface free energy at a given volume. Once the surface free energy of each orientation of GaAs is known, then the ECS

Downloaded 22 Jul 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

of a GaAs crystal can be built up by the Wulff construction. The distance of a particular facet having surface free energy per unit area of σ from the center of symmetry is proportional to r determined by[9]

$$r(\mathbf{n}) = \min_{\mathbf{m}} \left[\frac{\sigma(\mathbf{m})}{\mathbf{m}\cdot\mathbf{n}} \right], \tag{1}$$

where **m** is a unit vector normal to the given orientation and **n** is a normalized vector directing from the center of symmetry to any point on a given facet surface. With the constraint of a fixed volume of a crystal, Eq. (1) means that r of each facet for $\max[\mathbf{m}\cdot\mathbf{n}]$ is proportional to σ. (For the pyramid we treat in this work, the center of symmetry is on the line perpendicular to the base plane passing through the top vertex, and $\max[\mathbf{m}\cdot\mathbf{n}]$ then has the same value for all four sidewalls regardless of their orientation.) In Eq. (1), however, σ can change with growth conditions such as temperature and As overpressure. To our knowledge, no absolute comparison of surface free energy between different orientations of GaAs under a specific growth condition has been reported. But, as discussed below, Eq. (1) can explain the absence of a {111}-type sidewall pyramid.

For {111}, there are two possible orientations: $(111)A$ and $(111)B$. In general, they have different σ's and then, from Eq. (1), the ECS must be a pyramid elongated along either $[1\bar{1}0]$ or $[110]$ if all the sidewall facets belong to {111}. A theoretical model suggests the order of the lowest available σ as $\sigma_{(111)B} = 43\,\text{meV/Å}^2 < \sigma_{\{110\}} = 45\,\text{meV/Å}^2 < \sigma_{(111)A} = 51\,\text{meV/Å}^2$ under As-rich conditions where σ_j represents the surface free energy of facet j.[10] Then, the ECS might be thought to be a pyramid elongated along $[110]$ and, as shown in Fig. 5, have an expanded $(111)B$ facet and a contracted $(111)A$ facet. This elongated pyramid geometrically may not be probable as the ECS in our experiments since the base area allowed for epitaxial growth is close to circular, requiring equal size and shape sidewalls. In the elongated {111} pyramid shown in Fig. 5, a/b is the ratio of the width to the length of the base, equal to $\sigma_{(111)A}/\sigma_{(111)B}$. Reliable comparison of the total surface free energy ($=\Sigma_j \sigma_j A_j$, where A_j is the area of facet j of the {111} pyramid) with that of the {110} pyramid is not available. But, its rectangular base could explain the absence of a crystal shape consisting of {111} type sidewalls in this work. This suggests that the pattern shape for selective deposition as well as the growth conditions affect the ECS in nanopatterned growth. Indeed, we have already shown that the (113) and (111) orientations appear as the major facets on the sidewalls of GaAs selectively grown in a one-dimensional nanoscale stripe open area fabricated along the $[1\bar{1}0]$ direction.[11] This confirms that the ECS is not unique, but depends on pattern shape as well as on the pattern area and the growth conditions. In the modeling, the calculated $\sigma_{\{110\}}$ is comparable to $\sigma_{(111)B}$ while it has a larger difference from $\sigma_{(111)A}$. Intuitively, the formation of four identical-σ {110} facets could be more favorable than the simultaneous formation of two different high- and low-σ facets.

For our growth conditions, the surface migration length of Ga atoms on (001) is of micrometer order, significantly greater than the typical size of the epi-islands.[3] There has been no quantitative measurement of Ga surface migration length on different orientations with variation of growth temperature. But, it can be conjectured from the data on (001) and the comparison between different orientations that the Ga surface migration lengths on {111} and {110} would be greater than the lateral dimension of the epi-islands in these experiments.[12] There could be several different facets on the sidewalls of an epi-island at the initial stages of growth. Depending on pattern shape and growth conditions, however, the epi-island evolves to the most stable shape having the lowest surface free energy with facet expansion and contraction as growth is continued. By relying on Wulff construction and the observed geometric shape of a GaAs island, therefore, it can be concluded that a {110}-type sidewall pyramid is the ECS of GaAs grown on a nanoscale circular patterned (001) substrate under our growth conditions.

Finally, we comment on the reason that lateral overgrowth proceeds on some GaAs islands [square or rectangular shapes in the top view of Figs. 3(a) and 3(b)] with the formation of additional facets. We do not fully understand this observation but probably this could be due to a distorted shape of the original hole pattern that can induce asymmetric geometrical shapes on an island surface. Irregular initial faceting in a distorted circular hole can result in a growth rate of a particular facet that is considerably faster/slower than other neighboring ones, which asymmetrically modifies the overall shape of the epi-island. Thus, pattern uniformity and substructure even on nanometer and subnanometer scales is an important issue for further study. The investigation of the ECS of GaAs islands undergoing lateral overgrowth is presently underway and will be reported elsewhere.

IV. CONCLUSION

We report the ECS of GaAs grown on a nanoscale SiO_2-patterned substrate by MBE. The overall shape of a GaAs epilayer selectively grown in a nanoscale circular hole is strongly affected by faceting, which results in a pyramidal shape with {110} sidewalls and minimal lateral growth at our growth conditions. Comparison of the actual deposited volume with the supplied amount of material implies that growth on a nanoscale circular hole is slowed or terminated once {110} facets are formed. Generation of these facets across the entire range of hole diameters investigated means that the observed pyramidal shape is energetically very stable. From the experimental results and Wulff construction rule, we conclude that a {110}-type sidewall pyramid is one of the ECS's of GaAs on a (001) plane in nanoscale-patterned growth under our growth conditions.

ACKNOWLEDGMENTS

This work is supported by DARPA and ARO/MURI.

[1] G. Wulff, Z. Kristallogr. **34**, 449 (1901).
[2] For review, see A. Pimpinelli and J. Villian, *Physics of Crystal Growth* (Cambridge University, Cambridge, 1998).
[3] M. Hata, T. Isu, A. Watanabe, and Y. Katayama, J. Vac. Sci. Technol. B **8**, 692 (1990).

Downloaded 22 Jul 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

[4] S. C. Lee, K. J. Malloy, L. R. Dawson, and S. R. J. Brueck, J. Appl. Phys. **92**, 6567 (2002).
[5] S. C. Lee, K. J. Malloy, and S. R. J. Brueck, J. Appl. Phys. **90**, 4163 (2001).
[6] A statistical volume distribution of the GaAs islands similar to that shown in Fig. 4 was obtained from the profiling by atomic force microscopy (AFM) and also gave effective deposition thicknesses in this range. AFM profiling, however, did not clearly reveal the faceting on the surface of the epi-islands for $d \sim 180$ nm because the lateral dimension of an individual-facet was typically less than $d/2$, and the AFM profiles were convoluted with a tip shape of comparable dimensions.
[7] R. Fischer, L. Klem, T. J. Drummond, R. E. Thorne, W. Kopp, H. Morkoc, and A. Y. Cho, J. Appl. Phys. **54**, 2508 (1983).
[8] Vertical sidewalls which also belong to {110} were observed in the GaAs islands having square or rectangular shape in the top view of Fig. 3(b) but they were generated during lateral overgrowth along [100], as seen in Fig. 3(b). Further study of the ECS of these laterally overgrown GaAs islands will be reported elsewhere, but the appearance of vertical {110} sidewalls on those islands is another evidence that {110} is the most favorable orientation for ECS under our growth conditions.
[9] M. Wortis, in *Chemistry and Physics of Solid Surfaces VII*, edited by R. Vanselow and R. F. Howe (Springer, Berlin, 1988).
[10] N. Moll, A. Kley, E. Pehlke, and M. Scheffler, Phys. Rev. B **54**, 8844 (1996).
[11] S. C. Lee, L. R. Dawson, and S. R. J. Brueck, J. Cryst. Growth **240**, 333 (2002).
[12] T. Takebe, M. Fujii, T. Yamamoto, K. Fujita, and T. Watanabe, J. Appl. Phys. **81**, 7273 (1997).

Downloaded 22 Jul 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

HTML ABSTRACT + LINKS

Selective area growth of InAs quantum dots formed on a patterned GaAs substrate

S. Birudavolu, N. Nuntawong, G. Balakrishnan, Y. C. Xin, S. Huang, S. C. Lee, S. R. J. Brueck, C. P. Hains, and D. L. Huffaker[a)]
Center for High Technology Materials, University of New Mexico, 1313 Goddard SE, Albuquerque, New Mexico 87106

(Received 2 April 2004; accepted 14 July 2004)

We describe the growth and characterization of InAs quantum dots (QDs) on a patterned GaAs substrate using metalorganic chemical vapor deposition. The QDs nucleate on the (001) plane atop GaAs truncated pyramids formed by a thin patterned SiO_2 mask. The base diameter of the resulting QDs varies from 30 to 40 nm depending on the size of the mask. With specific growth conditions, we are able to form highly crystalline surface QDs that emit at 1.6 μm under room-temperature photopumped conditions. The crystalline uniformity and residual strain is quantified in high-resolution transmission electron microscopy images and high-resolution x-ray reciprocal space mapping. These strained QDs may serve as a template for selective nucleation of a stacked QD active region. © *2004 American Institute of Physics*. [DOI: 10.1063/1.1792792]

In the strain-driven self-assembled quantum dot (QD) growth process, nonuniformity in the wetting layer gives rise to QD nucleation.[1] The nucleation sites are weakly linked to surface steps that are not uniformly distributed; the QD nucleation sites are thus distributed randomly on the growth surface. The random nucleation results in a nonuniform QD size distribution and a broadened inhomogeneous linewidth, typically >20 meV.

Several methods have been used to define the nucleation sites in the growth plane including electron-beam lithography,[2–4] scanning tunneling microscope-assisted nanolithography,[5–7] and optical lithography.[8] The regrowth of patterned QDs has been accomplished using metalorganic chemical vapor deposition (MOCVD),[2] chemical beam epitaxy,[3] and molecular beam epitaxy.[4–8] The ultimate goal of this research is to demonstrate an inhomogeneously broadened linewidth similar to the homogeneously broadened linewidth (10 meV at 300 K, <1 meV at 4 K).[9] The formation of a large, dense, and uniform QD array is a complex task requiring a stable, uniform patterning process that yields a clean, undamaged surface for regrowth. To date, even room-temperature photoluminescence (RTPL) from regrown QDs has not been demonstrated.

In our work, we have used an optical lithography patterning method for the benefits of fast, large-scale exposure. Two co-authors (S.C.L. and S.R.J.B.) have pioneered this patterning process for homoepitaxial regrowth using molecular beam epitaxy. In this manuscript, we describe patterned QD formation using MOCVD utilizing the H_2 carrier gas to produce an atomically clean postprocessing surface. Furthermore, we have designed the epistructure to reduce nonradiative centers caused by dangling bonds and strain-related defects that result from patterning. In our structures, an InGaAs/GaAs buffer separates the QD active material from the regrowth interface. The buffer thickness is designed to form a truncated pyramid on which the QDs nucleate. The InAs QDs are formed by strain driven processes where the nucleation site is predefined since the In and As atoms preferentially adhere to the available (001) surface atop the GaAs truncated pyramid. This preferential growth on the (001) surface has been noted in previous nanopatterning work.[3]

The resulting patterned QDs demonstrate RTPL measured using a 5 mW HeNe pump laser (500 mW/cm^2), a lock-in amplifier, and an InGaAs detector. This patterning and regrowth methodology results in highly crystalline InAs pyramids confirmed by high resolution transmission electron microscopy (HRTEM) images and high-resolution x-ray reciprocal space mapping (HRXRSM).

The basic structure for processing the growth mask consists of a negative photoresist (PR) film, a 45-nm-thick SiO_2 layer, and a GaAs(001) substrate.[8] A two-dimensional array of holes is formed in the PR using an interferometric lithography technique. The pattern is transferred to the SiO_2 film with dry etching. A very high quality SiO_2 film is necessary to prevent pin-hole formation during the etching process. The patterning technique and subsequent processing to generate the SiO_2 pattern are fully described in Ref. 8. The resulting pattern features 100- to 200-nm-diam holes on a 360 nm pitch along the [110] direction.

The patterned pyramids and QDs are grown in a vertical MOCVD reactor at 60 Torr using TMGa, TMIn, and pure AsH_3 as source materials. Before regrowth, the patterned GaAs sample is deoxidized at 900 °C for 5 min. Hydrogen is flowed during this process to remove carbon, residual oxygen, and other contamination from the surface prior to regrowth. Growth is initiated with a GaAs buffer layer at a growth rate of 1.5 Å/s and V/III ratio of 200 at 800 °C and results in a truncated pyramid ($\sim$30 nm in height). The growth mode is characterized by a low growth rate and high V/III ratio to provide a very smooth GaAs buffer layer. The high V/III ratio is necessary as the group III elements have higher sticking coefficient than group V elements; this effect is amplified by the patterned surface. From initial regrowth experiments on thin stripe patterns in SiO_2, we know that Ga atoms diffuse $\sim$2 μm under these conditions. This diffusion length translates into a very high Ga atom concentration in the patterned holes. The resulting high Ga atom concentra-

[a)]Electronic mail: huffaker@chtm.unm.edu

 © 2004 American Institute of Physics
Downloaded 27 Sep 2004 to 128.253.229.12. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

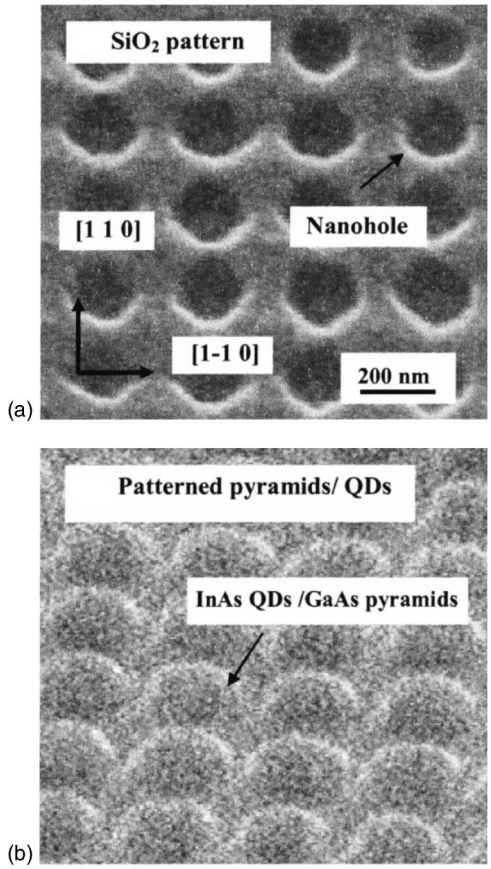


FIG. 1. A. Scanning electron microscopy images of (a) SiO_2 pattern on a GaAs(100) substrate, (b) patterned InAs QDs atop GaAs pyramids.

tion requires a high column V source flow to maintain stoichiometry. After the GaAs truncated pyramids are formed, the temperature is lowered and stabilized at 500 °C. Quantum dot growth is commenced with 5 ML of $In_{0.15}Ga_{0.85}As$ and 3 ML of InAs at a growth rate of 1 ML/s and V/III ratio of 25.

Figure 1 shows SEM images at several stages of patterned growth including (a) patterned holes in the SiO_2 mask, (b) patterned InAs QDs atop GaAs pyramids, and (c) InAs QDs/GaAs pyramids formed under unoptimized growth conditions. From Fig. 1(a), the center-to-center spacing between two holes is ~360 nm in [110] direction and 510 nm in [010]. Figure 1(b) shows a clean SiO_2 masked surface between the InAs pyramidal features demonstrating a highly selective growth. The general shape of pyramid base is circular with a base diameter of ~110 nm. The pyramidal density is $>1\times10^9$ cm^{-2}.

Transmission electron microscopy allows measurement of the limiting crystal planes and process uniformity. Higher resolution images delineate InAs from GaAs materials and confirm crystal integrity especially at the regrowth interface. Figure 2 includes several cross-sectional TEM images looking at the (110) crystal plane to show (a) an array of patterned surface InAs QDs on GaAs pyramids, (b) a single InAs QDs/GaAs pyramid, and (c) an InAs QD that nucleated (opportunistically) at the pyramid base due to a fabrication defect. In Fig. 2(a), the variation in SiO_2 hole diameter from 120 to 150 nm can be observed. The pattern variation results in GaAs pyramid height that varies from 30 to 40 nm and QD diameters ranging 30 to 35 nm. The limiting crystal plane, which is a function of substrate material, appears

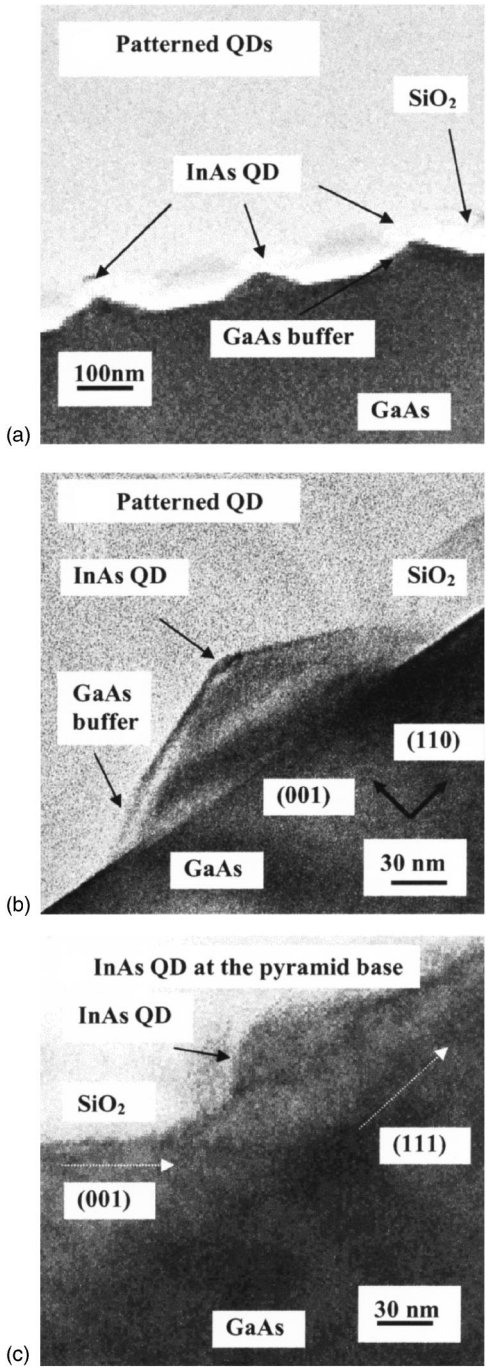


FIG. 2. High-resolution TEM micrographs of (a) patterned surface pyramids with QDs, (b) HRTEM of a single patterned InAs QDs/GaAs pyramid, (c) HRTEM of an InAs QD opportunistically formed at the pyramid base.

highly uniform throughout the observed images. Figure 2(b) shows crystal faceting along the (111) plane near the base and (311) near the apex. As discussed in previous work,[8] the pyramid shape changes from trapezoidal to triangular depending on the size of the nanohole and the amount of material deposition. We have designed our growth sequence to form a GaAs pyramid truncated at the (001) plane. As the (001) plane is the fastest growing plane, the InAs prefers nucleation on this plane. Figure 2(c) focuses at the base of one pyramid and shows the formation of an InAs QD between GaAs buffer and SiO_2 mask where the (001) plane has been exposed due to a processing defect.

Figure 3 shows RTPL spectra of (a) surface QDs on a planar substrate, (b) well-formed patterned surface QDs, and

Downloaded 27 Sep 2004 to 128.253.229.12. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

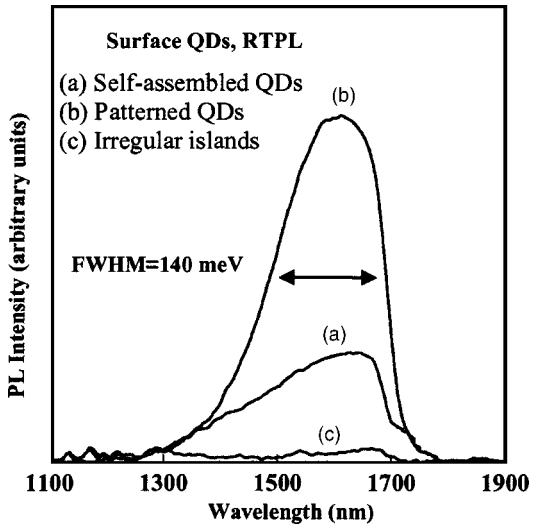


FIG. 3. Room temperature PL spectra from (a) self-assembled QDs, (b) patterned QDs, (c) incompletely nucleated QDs on GaAs pyramids.

(c) incomplete nucleation on a patterned substrate. The self-assembled (a) and the well-formed patterned (b) ensembles emit very similar spectra peaked near ~1610 nm with a full-width at half maximum (FWHM) of ~140 meV. The broad FWHM is a typical characteristic of surface SAQD emission.[10] The peak intensity of the patterned QDs is curiously two times greater than the SAQDs. While this may indicate better material quality, further experimental study is necessary. Finally, there is no PL emission from the irregular InAs/GaAs pyramids suggesting polycrystalline nature.

Figures 4(a)–4(c) show the (044) diffraction space maps for (a) the SAQDs, (b) the patterned InAs QDs, and (c) poorly nucleated QDs. The procedure for collection and calculations from angular units into reciprocal space coordinates of these maps were presented in previous publications.[11,12] The HRXRSM from the SAQDs in Fig. 4(a) shows only the projection of intensity around the GaAs substrate along the *k*-perpendicular [001] direction. This result indicates the wetting layer strain in the growth direction without ordering of dots in the lateral direction. The interesting features appear in the map of patterned InAs QDs, Fig. 4(b), where the periodic peaks, located symmetrically around the GaAs substrate peak in *k*-parallel [110] direction, are observed and indicate lateral ordering of the patterned InAs QDs. The broadening of periodic peaks in the ω (omega) direction indicates crystal homogeneity in each ensemble. The FWHM values are (a) 2.3×10^{-4} Å^{-1} for the SAQDs, (b) 1.9×10^{-4} Å^{-1} from the patterned QDs, and (c) 3.0×10^{-4} Å^{-1} from poorly nucleated QDs. The sharp periodic peaks of Fig. 4(b) suggest uniform crystal plane orientations compare to the broad spectrum of Fig. 4(c) that suggests a highly nonuniform crystallographic structure.

In conclusion, we have demonstrated growth of InAs QDs atop the (001) plane of patterned GaAs truncated pyramids using MOCVD. These QDs present RTPL from patterned InAs QDs. Data from HRTEM and HRXSM analysis shows high crystalline quality material. Of course, the end goal of patterned QD work is to demonstrate an inhomogeneously broadened linewidth that is very similar to the homogeneously broadened linewidth. Using our present method, improved QD uniformity requires improved pattern uniformity.

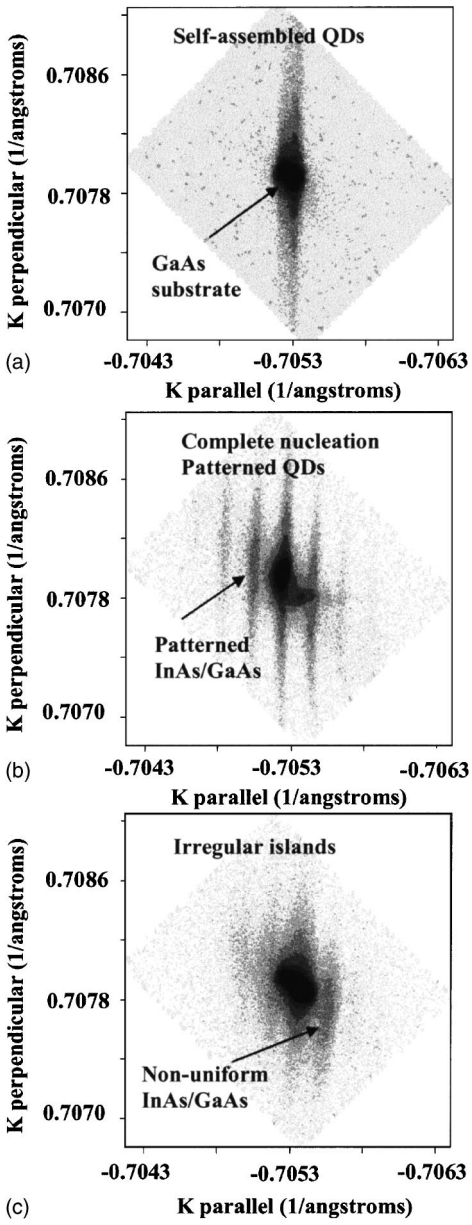


FIG. 4. High-resolution x-ray reciprocal space mapping in the (044) direction of (a) self-assembled QDs, (b) completely nucleated patterned QDs, (c) irregular InAs/GaAs islands.

[1]W. Seifert, N. Carlsson, J. Hohansson, M.-E. Pistol, and L. Samuelson, J. Cryst. Growth **170**, 39 (1997).
[2]T. S. Yeoh, R. B. Swint, A. Gaur, V. C. Elarde, and J. J. Coleman, IEEE J. Sel. Top. Quantum Electron. **8**, 833 (2002).
[3]D. Chithrani, R. L. Williams, J. Lefebvre, P. J. Poole, and G. C. Aers, Appl. Phys. Lett. **84**, 978 (2004).
[4]Y. Nakamura, N. Ikeda, S. Ohkouchi, Y. Sugimoto, H. Nakamura, and K. Asakawa, Physica E (Amsterdam) **21**, 551 (2004).
[5]S. Kohmoto, H. Nakamura, T. Ishikawa, and K. Asakawa, Appl. Phys. Lett. **75**, 3488 (1999).
[6]S. Nishikawa, S. Kohmoto, H. Nakamura, T. Ishikawa, K. Asakawa, and O. Wada, Phys. Status Solidi B **224**, 521 (2001).
[7]T. Ishikawa, S. Kohmoto, S. Nishikawa, T. Nishimura, and K. Asakawa, J. Vac. Sci. Technol. B **18**, 2635 (2000).
[8]S. C. Lee, A. Stintz, and S. R. J. Brueck, J. Appl. Phys. **91**, 3282 (2002).
[9]J. L. Spithoven, J. Lorbacher, I. Manke, F. Heinrichsdorff, A. Krost, D. Bimberg, and M. Dahne-Prietsch, J. Vac. Sci. Technol. B **17**, 1632 (1999).
[10]H. Saito, K. Nishi, and S. Sugou, Appl. Phys. Lett. **73**, 2742 (1998).
[11]P. F. Fewster, Appl. Surf. Sci. **50**, 9 (1991).
[12]A. L. Gray, A. Stintz, K. J. Malloy, T. C. Newell, and L. F. Lester, J. Cryst. Growth **222**, 726 (2001).

Downloaded 27 Sep 2004 to 128.253.229.12. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

HTML ABSTRACT + LINKS

Heteroepitaxial selective growth of $In_xGa_{1-x}As$ on SiO_2-patterned GaAs(001) by molecular beam epitaxy

S. C. Lee, L. R. Dawson, and S. R. J. Brueck
Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

A. Stintz
Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87103

(Received 27 February 2004; accepted 30 June 2004)

Selective growth of $In_xGa_{1-x}As$ on a SiO_2-mask patterned GaAs(001) substrate by molecular beam epitaxy (MBE) is reported. A simple model for selective growth is derived from the general rate equation of thin film growth. Based on this model, heteroepitaxial selective growth of $In_xGa_{1-x}As$ with x up to 0.07 is demonstrated at 595 °C, which is within the range of typical MBE growth temperatures for GaAs but is above the desorption temperature of In atoms from a GaAs surface. An incorporation rate of In atoms of about 0.22% for a Ga flux of 2.0×10^{13} atoms/cm^2 s was measured at this temperature. The concentration of In atoms effectively incorporated into the $In_xGa_{1-x}As$ layer is linearly proportional to the supplied In flux in the range $0.7-9.2\times10^{14}$ atoms/cm^2 s. The extremely low, but finite, In incorporation is utilized for the growth of 5 nm thick $In_xGa_{1-x}As$/GaAs quantum wells and ~300 nm thick $In_xGa_{1-x}As$ layers. Together with selective growth, the surface morphology and optical properties of high-temperature grown $In_xGa_{1-x}As$ are characterized. © *2004 American Institute of Physics*. [DOI: 10.1063/1.1786677]

I. INTRODUCTION

Selective growth is deposition restricted to a desired area on a patterned surface. In group IV molecular beam epitaxy (MBE), selective growth of Si and $Si_{1-x}Ge_x$ utilizing chemical reactions between a SiO_2 mask and incoming Si and Ge atoms has been reported.[1] In III-V MBE, we have demonstrated homoepitaxial selective growth of GaAs on a SiO_2 patterned substrate by a combination of desorption of Ga atoms from the SiO_2 surface and Ga migration from the SiO_2 to a nearby open GaAs surface.[2,3] Low-angle incidence epitaxy[4] and migration enhanced epitaxy[5] also have been investigated to achieve selective growth.

Selective growth of III-V materials has strong potential for electronic and optoelectronic devices, especially those requiring large lattice-mismatched heterostructures. To achieve long wavelength emission and high electron mobility on GaAs, $In_xGa_{1-x}As$ has been widely used as an active region for device structures. The lattice mismatch between InAs and GaAs is 7.2%.[6] For device applications, however, this is too large for InAs or $In_xGa_{1-x}As$ on GaAs to accommodate the excess strain energy with a layer-by-layer Frank-Merwe growth mode, and results in a change to Stranski-Krastanov island growth mode or in the generation of misfit dislocations severely degrading the epilayer quality. Theoretical considerations predict a reduction of the stress of a strained epilayer with decreasing deposition area in the nanoscale regime.[7,8] Selective growth is highly flexible in defining the epitaxial area on a planar surface from macroscale to nanoscale regimes. Then, it may be able to extend the materials available for better device performance to higher In concentration, lattice-mismatched compositions.

In MBE growth, In and Ga atoms show very different sticking behaviors. Experimentally, we found that In atoms begin to desorb from a GaAs surface around 560 °C whereas Ga atoms desorb at ~640 °C.[9] Generally, the optimal growth temperature for $In_xGa_{1-x}As$ is lower than the desorption temperature of In atoms and thus is below ~560 °C while the optimal temperature for GaAs growth is 580–600 °C. To grow a high quality $In_xGa_{1-x}As$/GaAs heterostructure, it is desirable to vary the substrate temperature as different materials are grown. Before increasing the growth temperature for the deposition of GaAs on an $In_xGa_{1-x}As$ layer, the surface is covered with a few nanometers thick low-temperature GaAs layer to avoid the desorption of In atoms.[10]

Although no systematic investigation of the desorption of In and Ga atoms on a SiO_2 surface has been reported, the relationship between their respective desorption temperatures is expected to be similar to that obtained on a GaAs surface. To achieve selective growth on a SiO_2-patterned GaAs substrate, negligible deposition on SiO_2 must be maintained throughout all layers of the heterostructure growth. We have demonstrated selective growth of GaAs and InAs separately at very different temperatures: 630 °C for GaAs and 535 °C for InAs.[3,11] Thus, the requirement of zero deposition on a SiO_2 surface, meaning no nucleation and incorporation under a given growth condition, is another restriction which makes it more difficult to grow GaAs selectively on $In_xGa_{1-x}As$ without In desorption. This is because the low-temperature grown GaAs protection layer must be also prepared in a selective growth mode at a temperature below 560 °C. As seen below, lowering the selective growth temperature below 560 °C may be possible by reducing the Ga flux to a very low level ($\ll2.0\times10^{13}$ atoms/cm^2 s used in this work). However, it is difficult to precisely control the Ga flux at these low values. Also, the Ga vapor pressure at a given

 © 2004 American Institute of Physics

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

temperature must be considered for finite deposition on a GaAs surface if the Ga flux is extremely low.

In this work, we take a different approach to heteroepitaxial selective growth. The optimal growth temperature for GaAs is 580–600 °C. Setting the growth temperature for both GaAs and $In_xGa_{1-x}As$ (with low x) in this range, we examine the sticking behavior of In atoms. It has been reported that In atoms have a nonzero sticking coefficient above 560 °C.[12,13] If selective growth of GaAs can be realized for 580–600 °C, In atoms definitely do not adsorb on a SiO_2 surface at this range, and selective growth of $In_xGa_{1-x}As$ as well as GaAs will be achieved at the same temperature. This has a strong advantage for the subsequent deposition of additional layers on $In_xGa_{1-x}As$ without regrowth.

In this work, we first present a simple model for selective growth derived from the general rate equation of thin film growth. A relation between the incoming flux and growth temperature is proposed in Sec. II. Relying on this model, we demonstrate selective growth of $In_xGa_{1-x}As$ on a SiO_2-patterned GaAs substrate with the variation of growth temperature and In flux. Experimental details are summarized in Sec. III. In Sec. IV, we examine selective growth of GaAs depending on growth temperature for a fixed Ga flux and find the low-temperature limit. We measure an incorporation rate of In atoms at this temperature that is extremely low ($\sim 2\times 10^{-3}$) but sufficient for growth of quantum wells (QWs) and even thick bulk $In_xGa_{1-x}As$ layers. Selective growth of $In_xGa_{1-x}As$ with $x\sim 0.07$ at 595 °C is achieved with variation of the In and Ga fluxes. Material properties of high-temperature grown $In_xGa_{1-x}As$ such as surface morphology and optical properties are reported. Finally, a summary and conclusions are presented in Sec. V.

II. A THEORY OF SELECTIVE GROWTH

Selective growth is achieved when zero deposition occurs on mask surface while finite deposition proceeds on the desired portions of the open substrate surface. First, we focus on zero deposition on the mask surface which is the necessary condition for selective growth. In principle, zero deposition on any ideal surface is always possible in the following way. Here, ideal means free from any extrinsic nucleation source such as impurities and crystal defects or the fluctuation and variation of surface potential by atomic steps. On such an atomically flat perfect surface, a single incident atom has a finite surface residence time τ and eventually desorbs from it. This means that the sequential incidence of a single atom with a time interval longer than τ results in zero deposition on a given surface. Within the residence time τ, the atom migrates on the surface until it desorbs from it. Statistically, the distance of migration is referred to as surface diffusion length L_D and is represented by $\sqrt{D\tau}$, where D is the surface diffusion constant.[14]

In crystal growth, however, multiple atoms are simultaneously incident on a given surface. If the number of atoms incident on and remaining in a *monatomic* state on a unit area at time t is n, then n satisfies the rate equation:[15]

$$\frac{dn}{dt} = F - \frac{n}{\tau} - 2\omega_1 n^2 - n\sum_{i=2}^{\infty} \omega_i n_i, \tag{1}$$

where F is an incident flux on an unit area per unit time, ω_i $(i=2,\ldots)$ is the incorporation rate of a single atom into a cluster consisting of i atoms, and n_i is a number density of i-atom-sized clusters. The third term describes consumption of single atoms through the formation of dimers, and this term with dimerization rate of ω_1 and the fourth term thus correspond to nucleation and clustering. Every n_i for $i\geqslant 2$ will have an equation similar to Eq. (1). We have neglected desorption of single atoms from an existing cluster in Eq. (1) which rarely occurs under typical thin film growth conditions.

Consider deposition of GaAs on a SiO_2-patterned GaAs substrate. Let L_D denote the surface diffusion length of a Ga atom on a SiO_2 surface. The lateral dimension of the SiO_2 mask is assumed to be much greater than L_D of Ga atoms at a given growth condition. Then, n is the density of monatomic Ga atoms on a unit area. The basic requirement of selective growth is zero deposition, that is, the absence of nucleation or clustering on the SiO_2 surface. That is, $n_i=0$ for $i\geqslant 2$ and the absence of dimer formation on the SiO_2 surface. Then, the rate equations for n_i for $i\geqslant 2$ do not need to be considered. By applying these conditions to Eq. (1), the rate equation for n of the SiO_2 surface for zero deposition can be written as

$$\frac{dn}{dt} = F - \frac{n}{\tau}, \tag{2}$$

which is the equation describing zero deposition on a SiO_2 surface. Theoretically, some clusters of which the size is below a certain critical value can migrate on the surface. Also, not only Ga atoms, but also other migrating species such as GaAs molecules could be involved in the equation for n_i similar to Eq. (1) for III-V semiconductors. In a more rigorous approach, the rate equations for n_i for $i\geqslant 2$ should be considered. In this work, migrating dimers or clusters for $i\geqslant 2$ are not considered because they ultimately play the role of immobile nucleation centers as growth is continued. Also, we do not consider migration of Ga atoms across the boundary between the SiO_2 mask and the open GaAs substrate surface, since that effect is not significant in achieving selective growth mode if the lateral dimension of the SiO_2 mask is much greater than L_D. Monatomic Ga atoms must be the major migrating species in deposition of GaAs on a SiO_2 surface and their behavior described by Eq. (2) is very important in understanding zero deposition. Again, it should be emphasized that these conditions are for the Ga atoms migrating on a wide-area unpatterned SiO_2 surface.

In steady state, Eq. (2) simplifies to $F=n/\tau$. To keep zero deposition on the mask, the surface diffusion length of a Ga atom from its initial point of incidence without encountering other Ga atoms on the SiO_2 surface before their desorption must be less than a certain characteristic length $L_c\sim n^{-1/2}$ obtained from the assumption of uniform distribution of n atoms on the unit area of the mask. Then, Eq. (2) in steady state can be rewritten as

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

$$F_c \sim \frac{1}{L_c^2 \tau} \lesssim \frac{1}{D\tau^2} \tag{3}$$

under the condition of $L_D \lesssim L_c$. In Eq. (3), F_c is a critical flux explained below. In Eq. (3), $D = C_d \exp(-E_{\text{diff}}/k_B T)$ where E_{diff} is the activation energy of a Ga atom for surface diffusion on a SiO_2 surface, C_d is a proportionality constant, k_B is Boltzmann's constant, T is the growth temperature, and $\tau = \nu_0^{-1} \exp(E_{\text{des}}/k_B T)$ where ν_0 is a desorption rate constant and E_{des} is an activation energy of a Ga atom for desorption from a SiO_2 surface. Then, F_c of Eq. (3) roughly satisfies the following condition:

$$F_c \sim \frac{\nu_0^2}{C_d} \exp[-(2E_{\text{des}} - E_{\text{diff}})/k_B T_s]. \tag{4}$$

It should be noted that E_{des} and E_{diff} are only weakly dependent on T and temperature dependence of the growth is determined largely by the explicit exponential dependence. In Eq. (4), T_s is the growth temperature above which a selective growth mode is available for a given F_c. If $2E_{\text{des}} > E_{\text{diff}}$ (which is valid in most thin film growth) is assumed, zero deposition on the mask at lower T is possible by reducing F and conversely, zero growth can be achieved at a higher flux if the growth temperature is raised. Thus, F_c gives the critical flux below which zero deposition on the SiO_2 mask is achieved at a given T_s.

Selective growth requires finite deposition on the desired substrate surface as well as zero deposition on the mask surface. At a Ga flux below F_c, therefore, the Ga atoms incident on a substrate surface must satisfy Eq. (1) with different τ and different ω_i's and also with *nonzero* n_i's ($i \geq 2$) for finite nucleation to lead to epitaxial growth. Thus, the actual growth conditions must be properly set to satisfy this requirement. In previous work, we have reported the homoepitaxial selective growth condition of GaAs on a SiO_2-patterned substrate.[2] Basically, Eq. (4) assumes that every atom incident on a patterned substrate undergoes a process of finite-time residence for subsequent incorporation/desorption. Although Eq. (4) is a rough approximation, it consistently agrees with the experimental results as shown in Sec. IV.

III. EXPERIMENT

Semi-insulating GaAs(001) wafers were used as substrates. On each 10×10 mm^2 substrate, a 45 nm thick SiO_2 film was deposited by electron-beam evaporation. With conventional photolithography and chemical etching process, a half of the SiO_2 film ($\sim 5 \times 10$ mm^2) was removed for examination of selective growth. All substrates were cleaned in an oxygen plasma to remove residual organic contaminants and by diluted HF ($HF:H_2O = 1:400$) to expose a fresh SiO_2 surface. As seen later, the HF treatment removing at least a 5–6 nm thick layer from the surface is very critical for avoiding deposition on the SiO_2 surface.

According to the model proposed in Sec. II, T_s and group III fluxes are important for selective growth. To characterize T_s for GaAs, a fixed Ga flux, F_{Ga}, $\sim 2.0 \times 10^{13}$ atoms/cm^2 s was employed. By changing growth temperature, T_s for GaAs of about 595 °C at this F_{Ga} was measured. A cracker cell was used to supply As_2 as a major component of As flux. Beam equivalent pressure of each element is denoted by P with subscripts Ga, In, and As_2. The ratio of Ga to As_2 in beam equivalent pressure, P_{As_2}/P_{Ga}, was changed from 25 to 100 to characterize the effect of As overpressure on selective growth. For $T_s = 595$ °C and F_{Ga} given above, In flux, F_{In}, was varied from 0.7–9.2 $\times 10^{14}$ atoms/cm^2 s to examine selective growth of $In_xGa_{1-x}As$. An InAs substrate was employed to calibrate F_{In} with reflection high-energy electron diffraction (RHEED) oscillations at $T = 510$ °C where the incorporation rate of In atoms is unity. The substrate temperature was monitored by optical pyrometry.

As-grown samples were examined with field-emission electron microscopy [scanning electron microscopy(SEM)], 77 K photoluminescence (PL), and double crystal x-ray diffractometry (DXRD). Indium composition x of $In_xGa_{1-x}As$ layers was measured by DXRD and PL. The incorporation rate of In atoms at high temperature was measured with 5 nm thick $In_xGa_{1-x}As/GaAs$ QWs by PL. Selective growth of $In_xGa_{1-x}As$ on a SiO_2-patterned GaAs(001) substrate was examined with ~300 nm thick layers through SEM. The In composition and strain relaxation were evaluated from the (004) symmetric and (115) asymmetric reflections of DXRD.

IV. RESULTS AND DISCUSSION

A. Selective growth of GaAs

Previously,[2] we have reported homoepitaxial selective growth of GaAs on a SiO_2-patterned substrate depending on growth temperature for a fixed F_{Ga}. For $F_{\text{Ga}} = 6.5 \times 10^{13}$ atoms/cm^2 s, the apparent sticking coefficient of Ga atoms on a SiO_2 surface, $S_{\text{Ga-SiO}_2}$, has a negligible value of less than 0.01 for $T \gtrsim 615$ °C.[2] Here, $S_{\text{Ga-SiO}_2}$ is defined as the fractional area covered by GaAs on a SiO_2 surface for a given set of growth conditions. In Sec. II, an ideal surface was assumed for simplicity, but this surface is not available in actual growth. Generally, external nucleation centers always exist on a SiO_2 surface. Such nucleation centers have a nonzero probability of being activated during growth, which increases with both growth time and supplied fluxes. In other words, nucleation is unavoidable as the growth time is extended and ultimately selective growth is not available in actual growth. Then, selective growth requires a certain criterion on $S_{\text{Ga-SiO}_2}$ and must be characterized under specific growth conditions including growth time or supplied amount. In this work, for the consistency with the previous work, we set $S_{\text{Ga-SiO}_2} \sim 0.01$ on a SiO_2 surface as the upper limit of selective growth at a given growth condition.

By reducing F_{Ga} from 6.5 to 2.0×10^{13} atoms/cm^2 s, T_s of GaAs was lowered from ~615 to ~595 °C. This experimental result is consistent with Eq. (4). Figure 1 shows the 45°-tilted SEM images taken near the boundary between the SiO_2 mask and the substrate of the samples grown at 585, 591, and 595 °C with $P_{As_2}/P_{\text{Ga}} \sim 25$. To enhance the precision of the temperature measurement, deposition was examined at a given growth temperature at least two or three times. The temperatures referred in this work were measured

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

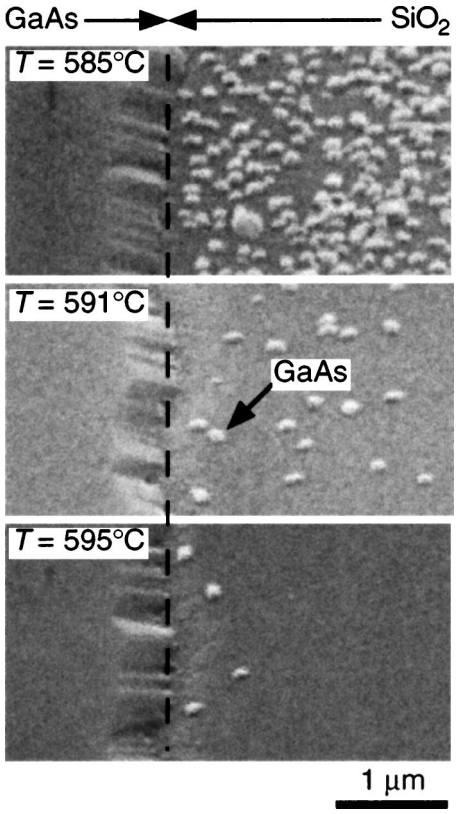


FIG. 1. 45° tilted SEM images of GaAs grown on a SiO_2-patterned substrate at 585, 591, and 595 °C. A dashed line indicates the boundary between GaAs and SiO_2.

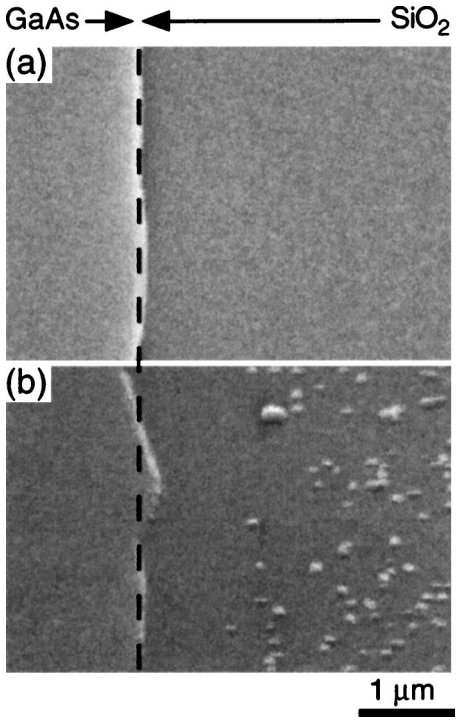


FIG. 2. (a) A 45° tilted SEM image of GaAs grown on a SiO_2-patterned substrate at 595 °C with $P_{As_2}/P_{Ga} \sim 100$. (b) A 45° tilted SEM image of GaAs grown on a SiO_2-patterned substrate at 630 °C with $P_{As_2}/P_{Ga} \sim 40$ without HF treatment. A dashed line indicates the boundary between GaAs and SiO_2.

before starting growth. More than an hour preheating at a given growth temperature under an As-rich environment provides a well-stabilized temperature. Run-to-run variation of growth temperature between before and after deposition was about ±1 °C by optical pyrometry. The deposition thickness for each sample was 200 nm. The dashed line roughly indicates the boundary between a GaAs epilayer and a SiO_2 mask. In each image of Fig. 1, the islands on the SiO_2 surface are GaAs. As seen in Fig. 1, the SiO_2 surface is almost covered by GaAs islands at $T=585$ °C. As T is increased, the number and size of GaAs islands are reduced and finally become negligible ($S_{Ga\text{-}SiO_2} < 0.01$) at $T=595$ °C. This growth condition was used for selective growth of $In_xGa_{1-x}As$ in the following experiments.[16]

In selective growth, F_{Ga} has been considered as a major factor but the As flux, F_{As} can be also important in As-based MBE because of the formation of GaAs molecular species which may play the role of a nucleation center on a SiO_2 surface. To examine whether As overpressure influences the selective growth mode, P_{As_2}/P_{Ga} was increased from 25 to 100. Figure 2(a) shows a 45° tilted SEM image from the sample grown at 595 °C under $P_{As_2}/P_{Ga} \sim 100$. As seen in this figure, selective growth mode is kept without nucleation on the SiO_2 mask. Thus, in the range examined in this work, high As overpressure does not induce nucleation on a SiO_2 surface confirming that F_{Ga} is the major parameter for selective growth of GaAs.

All the SiO_2-patterned substrates used for selective growth were treated with diluted HF to remove a very thin layer (~5–6 nm in thickness) from the SiO_2 surface immediately before loading them into the MBE growth chamber. This process is very critical for selective growth. Figure 2(b) shows a 45° tilted SEM image of the sample grown at 630 °C with the same F_{Ga} but without the HF treatment. In Fig. 2(b), significant nucleation of GaAs occurs on the SiO_2 mask at a temperature considerably higher than T_s. Before being loaded into the MBE chamber, the SiO_2-patterned substrate undergoes a wafer patterning and cleaning process. The comparison of Fig. 2(b) with Fig. 1 suggests that the SiO_2 surface is contaminated or damaged and some external nucleation centers are generated during the patterning process. Therefore, removal of such nucleation sources by diluted HF is very important to achieve selective growth relying on desorption of Ga atoms from the SiO_2 surface.

B. Incorporation of In atoms at high temperature

Selective growth of $In_xGa_{1-x}As$ on GaAs is the major goal of this work and the desorption temperature of InAs ~560 °C is referred to as T_{des}. In Sec. IV A, T_s for selective growth of GaAs was determined as 595 °C. This is in the optimal range of GaAs growth but higher than T_{des}. From Eq. (4), T_s can be lowered further by decreasing F_{Ga}. However, the accuracy in control of F_{Ga} at very low range and the required growth time for a given thickness which is inversely proportional to F_{Ga} make this strategy problematic. Instead, it is useful to investigate the incorporation of In atoms on a GaAs substrate which is significantly affected by temperature near and above T_{des}. Then, it is desirable to fix T at 595 °C and to examine selective growth of $In_xGa_{1-x}As$ with

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

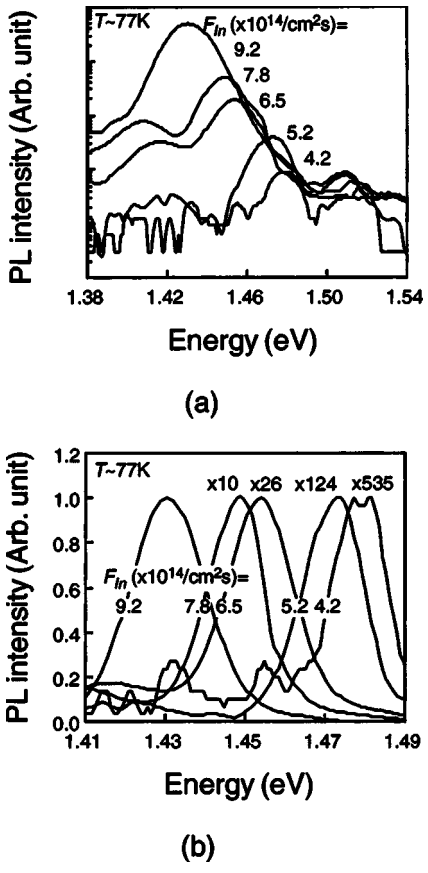


FIG. 3. (a) 77 K PL spectra of 5 nm thick $In_xGa_{1-x}As/GaAs$ QWs with variation of F_{In} from 4.2 to 9.2×10^{14} atoms/cm^2 s. (b) PL spectra shown in (a) with focusing on peak 1. PL intensity was normalized with respect to that of $F_{In}=9.2\times10^{14}$ atoms/cm^2 s.

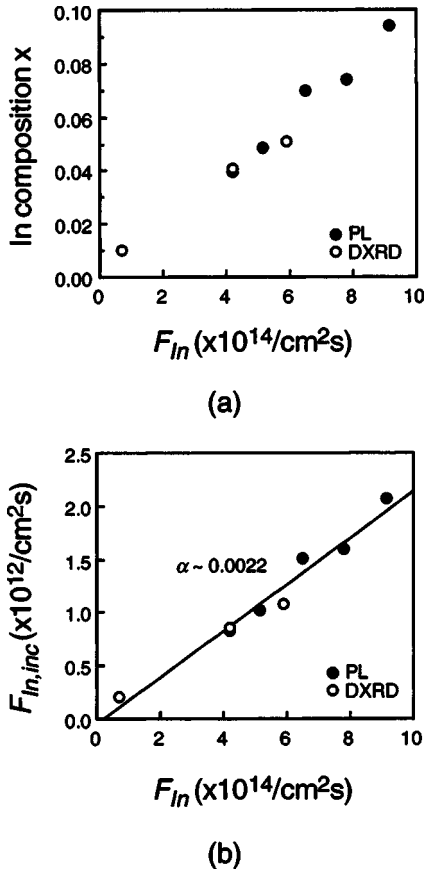


FIG. 4. (a) Relation between F_{In} and x measured from PL shown in Fig. 3. (b) Relation between F_{In} and $F_{In,inc}$.

variation of F_{In}. A major concern is then to determine the incorporation rate of In atoms on a GaAs surface at 595 °C. Another issue in these high temperature grown $In_xGa_{1-x}As$ layers is the quality of the epilayer such as crystallinity, surface morphology, and optical properties. First, we examined the sticking behavior of In atoms at this temperature and applied the results to selective growth of $In_xGa_{1-x}As$ on a SiO_2-patterned GaAs(001) substrate. While the In incorporation at this temperature is extremely low as seen below, we keep $P_{As_2}/P_{Ga}\sim100$ considering the addition of F_{In} to F_{Ga} for growth of $In_xGa_{1-x}As$.

With the variation of F_{In}, the incorporation rate of In atoms at $T=595$ °C was measured for the growth of $In_xGa_{1-x}As/GaAs$ QWs. The In flux was changed from 4.2 to 9.2×10^{14} atoms/cm^2 s. Based on the preliminary data about In incorporation, the width of the QW was controlled to 5 nm by growth time determined by both F_{Ga} and F_{In}. On every $In_xGa_{1-x}As$ layer, a 150 nm thick GaAs capping layer was deposited. The range of F_{In} used in this section is for the measurement of the In incorporation rate at 595 °C. As discussed in the following section, selective growth of $In_xGa_{1-x}As$ at the given growth conditions ($F_{Ga}=2.0\times10^{13}$/cm^2 at 595 °C) is achieved only for $F_{In}<6.0\times10^{16}$/cm^2.

Figure 3 shows PL spectra obtained from 5 nm thick $In_xGa_{1-x}As/GaAs$ QWs grown with different F_{In}'s. All the spectra were taken at 77 K with a 632.8 nm He-Ne laser and a room temperature Ge detector. In Fig. 3, three PL peaks were observed. Among them, the PL peak at 1.51–1.52 eV is very weak and is almost independent of F_{In} and is thus from the GaAs substrate. As F_{In} is increased, the remaining two major peaks are shifted to lower energy but with a nearly constant peak separation. The peak separation shows a little fluctuation and is about 38–45 meV without particular trend depending on F_{In}. For convenience, the high and the low intensity PL peak in each spectrum are denoted by peak 1 and 2, respectively. Figure 4(b) shows the variation of peak 1 energy depending on F_{In} with PL intensity normalized with respect to that of $F_{In}=9.2\times10^{14}$ atoms/cm^2 s.

As F_{In} is increased from 4.2 to 9.2×10^{14} atoms/cm^2 s, the energy of peak 1 varies from 1.4813 to 1.4305 eV. The shift of PL peak 1 to low energy with increase of F_{In} means an increase in the In composition of the QW. Also, as seen in Fig. 4, the PL intensity is rapidly increased with F_{In}. These results suggest that quantum confinement is proportionally enhanced by a potential energy barrier which depends on the In amount in a QW structure, and therefore, imply that peak 1 originates from the transition between the lowest conduction subband and the lowest valence subband of the QW. Then, peak 2 in each spectrum could be the transition from the lowest conduction subband to impurities.

With the empirical formula for the relation between PL peak energy and x of a 5 nm thick $In_xGa_{1-x}As/GaAs$ QW reported in our previous work, the In composition was estimated from the PL peak energies and is summarized in Table

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

TABLE I. Summary of experimental results obtained from PL and DXRD.

		DXRD Peak separation (arc sec)			
$F_{In}(\times 10^{14}/cm^2\,s)$	PL (77 K) Peak 1 (eV)	(004)	(115)	In composition x	$F_{In,inc}(\times 10^{12}/cm^2\,s)$
0.7		180	324	0.010	0.2
4.2	1.4813			0.039	0.8
4.2		544	1163	0.041	0.8
5.2	1.4730			0.048	1.0
6.0		746	1310	0.051	1.1
6.5	1.4531			0.070	1.5
7.8	1.4489			0.074	1.6
9.2	1.4305			0.094	2.1

I.[17] As F_{In} is increased from 4.2 to 9.2×10^{14} atoms/cm^2 s, x varies linearly from 0.039 to 0.094 as shown in Fig. 4(a). For this range of misfit, the critical thickness of $x=0.094$ $In_xGa_{1-x}As$ is at least three times 5 nm and the QWs of Fig. 4(a) should be assumed to be strained even with the high growth temperature justifying the use of the empirical formula to evaluate the In mole fraction in Table I.[18]

The incorporation rate of In atoms on a GaAs surface at 595 °C, α, defined as

$$\alpha = \frac{F_{In,inc}}{F_{In}}, \tag{5}$$

can be extracted from Table I or Fig. 4(a). In Eq. (5), $F_{In,inc}$ is an amount of an effective In flux which is incorporated into an $In_xGa_{1-x}As$ layer. From the In compositions measured by PL, $F_{In,inc}$ of Table I can be calculated and is plotted in Fig. 4(b). The least square fit of the experimental results with linear dependence of $F_{In,inc}$ on F_{In} is indicated by the solid line. At 595 °C with $F_{Ga}=2.0\times10^{13}$ atoms/cm^2 s and $P_{As_2}/P_{Ga}\sim100$, α is ~0.0022.

The measured α is extremely low compared with the reported data which is about 0.3–0.4 at similar growth temperatures.[12,13] The previous work used mass spectrometry and RHEED oscillations to measure the desorption In flux during growth of an $In_xGa_{1-x}As$ layer. These experimental techniques are very sensitive to surface state such as In coverage and segregation. On the other hand, we characterized α through the bulk $In_xGa_{1-x}As$ layers by PL and DXRD.

From *in situ* RHEED, we have observed that the (2×4) reconstruction pattern from the GaAs surface is changed to (4×2) when InAs begins to grow on GaAs at 595 °C. This transition was observed even for the lowest F_{In} examined in this work, where $P_{As_2}/P_{In}\sim20$. Depending on F_{In} there is a finite time delay from the starting point of InAs deposition to the buildup of the (4×2) pattern. The higher F_{In} the shorter the time delay. The time delay changes from less than 1 to ~20 s as F_{In} is reduced from 9.2 to 0.7×10^{14} atoms/cm^2 s. When the In supply is stopped, the (4×2) pattern immediately changes back to the (2×4) pattern. A similar variation also occurs for growth of $In_xGa_{1-x}As$ (F_{In} and F_{Ga}) with a time delay shorter than that observed with only F_{In}. This means that the change of RHEED pattern is related to the presence of In atoms on the surface, but that the Ga atoms also play a role, perhaps in assisting the In-atom adsorption.

At $F_{In}=7.0\times10^{13}$ atoms/cm^2 s and $P_{As_2}/P_{In}\sim20$, the change of RHEED pattern mentioned above does not occur for $T<T_{des}$ in growth of InAs on GaAs. On the contrary, RHEED exhibits a change to a spotty pattern indicating the formation of quantum dots. Also, growth of GaAs at the same As overpressure with the F_{Ga} equal to $F_{In}=7.0\times10^{13}$ atoms/cm^2 s proceeds keeping (2×4) pattern at 595 °C without change of a RHEED pattern. These suggest that the change of RHEED pattern in the $In_xGa_{1-x}As$ growth is not due to As deficiency. As measured above, the sticking coefficient of In atoms is extremely low at 595 °C, and $F_{In,inc}$ is far below F_{In} as shown in Fig. 4(b) or equivalently much less than F_{Ga} under the given growth conditions.

The change of RHEED pattern in the $In_xGa_{1-x}As$ growth at 595 °C implies that the front growth surface of InAs growth becomes In rich. In other words, like InAs growth below T_{des} where an In floating layer formed by segregation effect on a GaAs surface, an In adlayer is also generated on the surface even at 595 °C.[19] Unlike the growth below T_{des}, however, most of In atoms trying to form the adlayer evaporate from the surface as the In supply stops because of high growth temperature. Quantitative analysis about the incorporation of In atoms above T_{des} will be reported elsewhere but the change of RHEED pattern and the desorption of the In adlayer from the front growth surface suggest that a significant amount of In atoms resides on the front surface during growth and that the In amount at the surface of $In_xGa_{1-x}As$ could be very different from the incorporation of In atoms effectively measured from bulk $In_xGa_{1-x}As$. Even if In atoms are incorporated into an $In_xGa_{1-x}As$ layer, they can diffuse out to or segregate to and subsequently desorb from the front growth surface if they are very close to it. The comparison of a bulk with a surface in the measurement of In incorporation rate must consider this difference for consistent interpretation.

Another possibility for the large discrepancy in α between our measurement and the reported data is the difference of Ga flux. As mentioned above, Ga atoms seem to positively influence In incorporation. In this work, F_{Ga} is very low and less than 10% of that in one of the reported works.[12] If Ga atoms having almost unit sticking coefficient at our growth condition play a certain role in the incorpora-

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

tion of In atoms into an $In_xGa_{1-x}As$ layer, the change of F_{Ga} may affect α of In atoms. Also, some difference in the calibration of growth temperature could partly explain the discrepancy in α. Further study is required to understand the incorporation mechanism of In atoms above T_{des}.

C. Selective growth of $In_xGa_{1-x}As$

Based on the incorporation rate measured in the preceding section, selective growth of $In_xGa_{1-x}As$ on a SiO_2-patterned GaAs(001) substrate was examined. According to Fig. 1, the incorporation of Ga atoms is considerably reduced with a few degrees difference in growth temperature (e.g., between 591 and 595 °C). Then, from Eq. (4), the addition of small amount of F_{In} to $F_{Ga}=2.0\times10^{13}$ atoms/cm^2 s may result in incomplete selective growth at 595 °C. For this reason, the range of F_{In} was lowered to $0.7-6.0\times10^{14}$ atoms/cm^2 s. As previously mentioned, selective growth relying on desorption can be affected by supplied amount or growth time. In the consideration of the application to QW growth or a strain buffer, selective growth was examined with ~300 nm thick $In_xGa_{1-x}As$ layers.

Figure 5 shows 45°-tilted SEM images of the $In_xGa_{1-x}As$ layers grown on a SiO_2-patterned GaAs(100) substrate with $F_{In}=0.7$, 4.2, and 6.0×10^{14} atoms/cm^2 s. Also, Fig. 6 shows DXRD (004) and (115) (glancing incidence) reflection rocking curves obtained from the samples shown in Fig. 5. In Fig. 6(a), the angular separation between the peaks from GaAs and $In_xGa_{1-x}As$ in both (004) and (115) glancing incidence rocking curves is increased with F_{In}. In (004) [(115) glancing incidence] reflection rocking curves, the peak separation is varied from 180 to 746 arc sec (324 to 1310 arc sec) as F_{In} is changed from 0.7 to 6.0×10^{14} atoms/cm^2 s. This means that In composition is enhanced with F_{In} as observed in the PL results shown in Fig. 3. From the peak separations, the corresponding x varying from 0.01 to 0.05 was obtained. Relaxation is also increased from 19.2% to 39.7%.[20] According to a theoretical calculation based on the Matthews and Blakeslee model, the critical thickness of $In_xGa_{1-x}As$ for $x=0.01$ is less than 300 nm.[18] Thus, partial relaxation of the $In_xGa_{1-x}As$ layers of Fig. 5 was expected. The x-F_{In} relation obtained from DXRD was plotted in Fig. 4 together with the results of PL. As seen in Fig. 4, the DXRD results are consistent with the PL results in measurement of x and the incorporation rate of In atoms. The measured In composition and relaxation of each sample were summarized in Table I.

The surface of the $In_xGa_{1-x}As$ layers shown in Fig. 5 is severely degraded except for that of $x=0.01$. The surfaces of both samples shown in Figs. 5(b) and 5(c) were hazy in visual inspection and revealed island-type morphology in SEM. In Fig. 5, higher x results in a more severe islanding surface by three-dimensional growth mode. This island growth mode is partly affected by low growth rate employed in this work. Compared with other heterostructures (e.g., $Si_{1-x}Ge_x/Si$), the surface degradation by islanding growth at the same range of misfit ~0.0035 for $x=0.05$ is more severe in $In_xGa_{1-x}As/GaAs$.[21] This could be due to high growth temperature above T_{des}.[22,23]

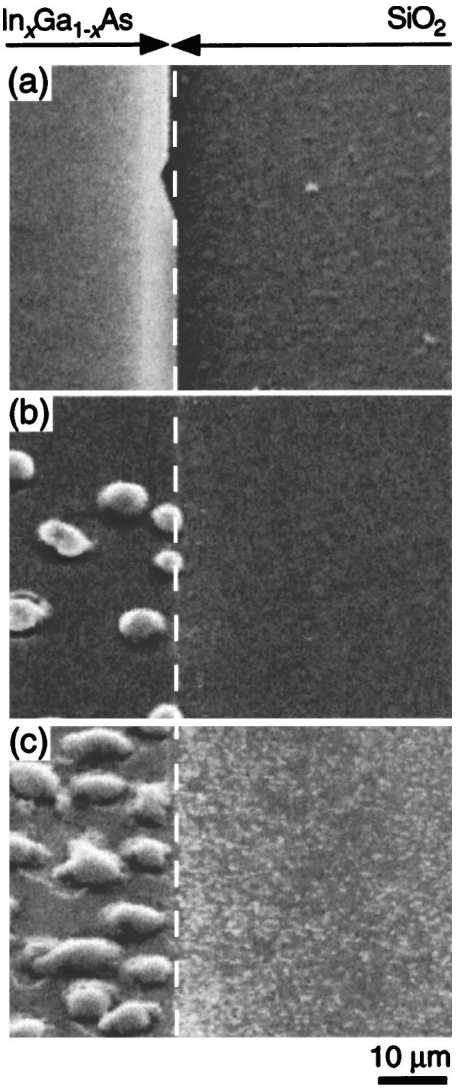


FIG. 5. 45° tilted SEM images of $In_xGa_{1-x}As$ grown on a SiO_2-patterned GaAs substrate at 595 °C with $F_{In}=0.7$ ($x=0.01$) (a), 4.2 ($x=0.04$) (b), and 6.0×10^{14} atoms/cm^2 s ($x=0.05$) (c). A dashed line indicates the boundary between $In_xGa_{1-x}As$ and SiO_2.

As seen in Fig. 5, negligible deposition occurs on a SiO_2 surface for $x=0.04$ while noticeable nucleation was observed on a SiO_2 mask for $x=0.05$. Figure 7 presents a magnified SEM image of the sample shown in Fig. 5(c). As seen in Fig. 7, $In_xGa_{1-x}As$ islands were grown on the SiO_2 surface but their physical size is less than 1 μm which is much smaller than those on the GaAs surface. Therefore, selective growth of $In_xGa_{1-x}As$ for $F_{Ga}=2.0\times10^{13}$ atoms/cm^2 s at 595 °C is achieved by reducing F_{In} from 6.0 to 4.2×10^{14} atoms/cm^2 s. This also agrees with Eq. (4) in that selective growth occurs by reduction of F for a given T. With corresponding $F_{In,inc}$, F_c at 595 °C is increased to 2.08×10^{13} atoms/cm^2 s which is the sum of F_{Ga} and $F_{In,inc}$. But it should be noted that more than 4% increment of F results in incomplete selective growth. This means that the margins in F and T from F_c and T_s for selective growth are not very much (~4% in F and ~4–5 °C in growth temperature).

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

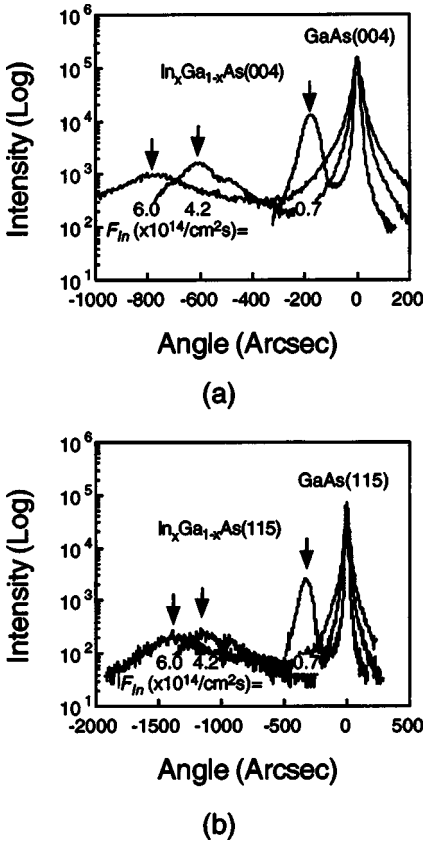


FIG. 6. (004) (a) and (115) glancing incidence (b) x-ray rocking curves of the samples shown in Fig. 5.

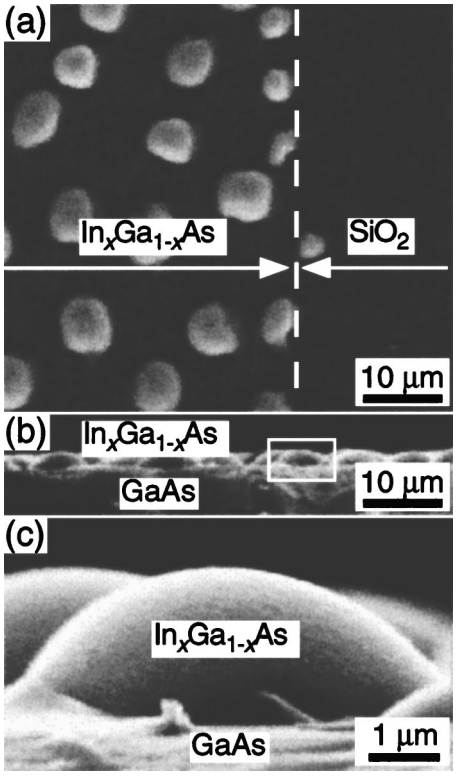


FIG. 8. A top (a) and a side view (b) of the selectively grown $In_xGa_{1-x}As$ ($x=0.07$) on a SiO_2-patterned GaAs substrate at 595 °C with $F_{Ga}=1.4\times10^{13}$ atoms/cm^2 s and $F_{In}=6.5\times10^{14}$ atoms/cm^2 s. (c) A side-view SEM image of a single $In_xGa_{1-x}As$ epi-island indicated by a dashed box in (b).

To increase x while keeping a selective growth mode, one possible way is to reduce F_{Ga} and increase F_{In} with keeping their effective sum below 2.08×10^{13} atoms/cm^2 s. For higher x, F_{Ga} is reduced to 1.4×10^{13} atoms/cm^2 s with $F_{In}=6.5\times10^{14}$ atoms/cm^2 s of which the effective sum is about 1.55×10^{13} atoms/cm^2 s. Figure 8 shows a top- and a side-view SEM image of the sample grown under these conditions. Figure 8(a) was taken near the boundary between the $In_xGa_{1-x}As$ layer and the SiO_2 mask. As seen in this figure, no deposition occurs on the SiO_2 surface and selective growth is achieved at the given fluxes, as expected. Figures 9(a) and 9(b) show a (004)and (115) glancing incidence reflection x-ray rocking curve of the 300 nm thick $In_xGa_{1-x}As$ on GaAs and a 77 K PL spectrum of a 5 nm thick $In_xGa_{1-x}As$/GaAs QW, respectively. Both $In_xGa_{1-x}As$ layers in these samples were grown under the same growth conditions. The angular peak separation between GaAs and $In_xGa_{1-x}As$ in a (004) and a (115) glancing incidence reflection rocking curve are 939 and 1613 arc sec, respectively. In the PL spectrum, like those in Fig. 3(a), three PL peaks were observed. The PL peak energy from the $In_xGa_{1-x}As$/GaAs QW corresponding to peak 1 of Fig. 3 is 1.445 eV. Thus selective growth is achieved with x increased to 0.069 in DXRD (with relaxation of 56%) and 0.078 in PL. Such an increment of x is consistent with the results shown in Fig. 4 with similar α. Enhancement of x by reduction of F_{Ga}, however, may have a restriction since, as previously mentioned, Ga atoms have a certain role in incorporation of In atoms onto GaAs.

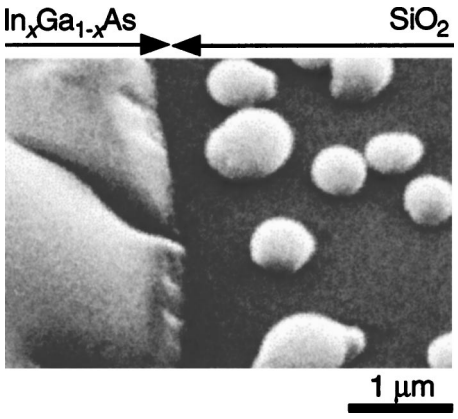


FIG. 7. A magnified 45° tilted SEM image of the sample shown in Fig. 5(c).

In Fig. 8, the three-dimensional shape and size of an epitaxially grown $In_xGa_{1-x}As$ island (epi-island) is clearly revealed. While the lateral dimension is about 3–7 μm, its height in Fig. 8(c) is about 2.5 μm which is considerably greater than the deposition amount per unit area. Also, inter-island distance (center-to-center) is about 10 μm. The overall shape of an $In_xGa_{1-x}As$ epi-island seems to be a spherical section without faceting. The observation of (004) and (115) reflection peaks, however, shows that this rounded shape $In_xGa_{1-x}As$ epi-island is epitaxially grown on a GaAs surface. Formation of islands suggests the redistribution of Ga and In atoms on the front growth surface by migration. According to the measurement of surface diffusion length of Ga

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

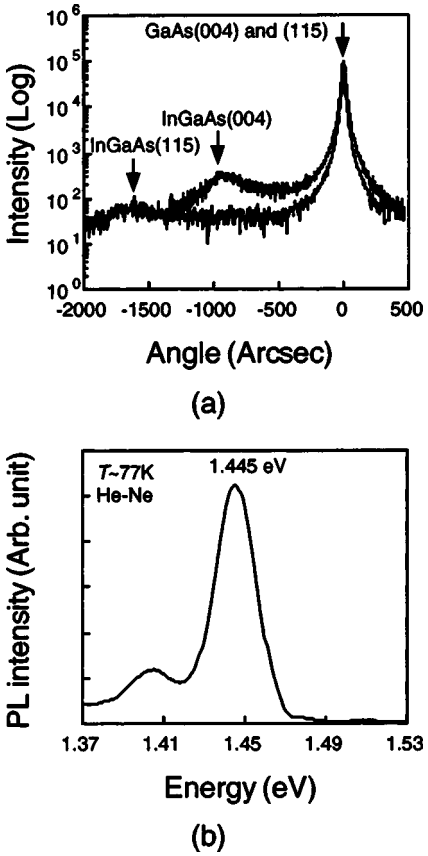


FIG. 9. (a) (004) (a) and (115) glancing incidence x-ray rocking curves of the sample shown in Fig. 8. (b) 77 K PL spectrum of a 5 nm thick $In_xGa_{1-x}As/GaAs$ QW of which the $In_xGa_{1-x}As$ layer was grown at the same growth condition as that shown in Fig. 8.

atoms on a GaAs surface, Ga atoms can migrate a distance greater than $\sim 3\ \mu m$ at 595 °C for $F_{Ga} \sim 2 \times 10^{14}$ atoms/cm^2 s which is almost ten times the flux used in this work.[24] Reduction of F_{Ga} enhances the surface diffusion length and may enable Ga atoms to migrate the interisland distance.[25] As previously mentioned, the islanding mode shown in Figs. 5 and 8 is unusual at the given range of misfit but such a mode is feasible as a partial release of strain energy due to lattice mismatch like self-assembly of quantum dots.[26]

Stress release in lattice mismatched heterostructures is one of the important issues in selective growth. As seen in Fig. 8, the surface morphology of the selectively grown $In_{0.07}Ga_{0.93}As$ layer which was severely degraded is critical for device applications. The size of the $In_xGa_{1-x}As$ epi-island is in micrometer scale. If a nanoscale patterning of which the period is much smaller than the lateral dimension of the epi-island is applied for the growth of such an $In_xGa_{1-x}As$ island, it significantly improves surface morphology and crystallinity by relieving strain without relying on the generation of misfit dislocations.[27,28] Detailed experimental results of strain-relieved dislocation-free $In_xGa_{1-x}As$ selectively grown on a nanoscale SiO_2-patterned GaAs will be reported elsewhere.

In this work, selective growth was demonstrated up to a 300 nm thick layer with $x \sim 0.07$. Thus, it is possible to grow a very thin layer such as a QW layer with selective growth mode. Assisted by surface migration,[3] selective growth can be achieved at lower temperature or higher In composition with better crystallinity and surface morphology.

V. SUMMARY AND CONCLUSIONS

Heteroepitaxial selective growth of $In_xGa_{1-x}As$ on SiO_2-patterned GaAs by molecular beam epitaxy (MBE) has been reported. A simple model for selective growth has been derived from the general rate equation for thin film growth. Based on this model, selective growth of $In_xGa_{1-x}As$ has been achieved on a SiO_2-patterned GaAs(001) substrate with variation of growth temperature and Ga and In fluxes. Growth temperature is set to 595 °C which is in the range of typical growth temperature of GaAs in MBE but is above the desorption temperature of In atoms from GaAs surface. At this temperature, the incorporation rate of In atoms is about 0.22%. The incorporation rate which is extremely low but finite is utilized for the growth of 5 nm thick $In_xGa_{1-x}As/GaAs$ quantum wells and ~300 nm thick $In_xGa_{1-x}As$ layers. The amount of In atoms effectively incorporated into an $In_xGa_{1-x}As$ layer is linearly proportional to the supplied In flux in the range of 0.7–9.2 $\times 10^{14}$ atoms/cm^2 s. The QWs show a systematic variation of PL peak energy and intensity depending on In flux. The high-temperature grown $In_xGa_{1-x}As$ has a severely degraded surface morphology due to islanding growth mode but clearly shows selective growth on a SiO_2-patterned GaAs substrate. The experimental results are consistent with the model for the relation between growth temperature and incoming flux. In future work, we will report selective growth of high In composition $In_xGa_{1-x}As$ on GaAs in the nanoscale regime to investigate the effect of a finite epitaxial area on the stress distribution in large lattice-mismatched heterostructures.

ACKNOWLEDGMENTS

Support for this work was provided by the ARO/MURI in Deep Subwavelength Optical Nanolithography and by Sandia National Laboratory.

[1] S. J. Yun, S. C. Lee, B. W. Kim, and S. W. Kang, J. Vac. Sci. Technol. B **12**, 1167 (1994).
[2] S. C. Lee, K. J. Malloy, L. R. Dawson, and S. R. J. Brueck, J. Appl. Phys. **92**, 6567 (2002).
[3] S. C. Lee, K. J. Malloy, and S. R. J. Brueck, J. Appl. Phys. **90**, 4163 (2001).
[4] T. Nishinaga, J. Cryst. Growth **237–239**, 1410 (2002).
[5] K. Suzuki, M. Ito, and Y. Horikoshi, Jpn. J. Appl. Phys., Part 1 **38**, 6197 (1999).
[6] *Semiconductors, Group IV Elements and III-V Compounds*, Data in Science and Technology, edited by O. Madelung, M. Schulz, and A. Goldmann (Springer, Berlin, 1991).
[7] A. Fischer and H. Richter, Appl. Phys. Lett. **61**, 2656 (1992).
[8] D. Zubia and S. D. Hersee, J. Appl. Phys. **85**, 6492 (1999).
[9] R. Fischer, L. Klem, T. J. Drummond, R. E. Thorne, W. Kopp, H. Morkoc, and A. Y. Cho, J. Appl. Phys. **54**, 2508 (1983).
[10] A. Stintz, G. T. Liu, A. L. Gray, R. Spillers, S. M. Delgado, and K. J. Malloy, J. Vac. Sci. Technol. B **18**, 1496 (2000).
[11] S. C. Lee, A. Stintz, and S. R. J. Brueck, J. Appl. Phys. **91**, 3282 (2002).
[12] K. R. Evans, C. E. Stutz, E. N. Taylor, and J. E. Ehret, J. Vac. Sci. Technol. B **9**, 2427 (1991).
[13] M. Mashita, Y. Hiyama, K. Arai, B.-H. Koo, and T. Yao, Jpn. J. Appl. Phys., Part 1 **39**, 4435 (2000).

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

[14]B. Lewis and J. C. Anderson, *Nucleation and Growth of Thin Films* (Academic, New York, 1978).
[15]B. Mutaftschiev, *The Atomistic Nature of Crystal Growth* (Springer, Berlin, 2001).
[16]This work and Refs. 2 and 3 were performed with different MBEs having different optical pyrometry systems. There could be 5–10 °C difference between these two systems in the measurement of growth temperature.
[17]S. C. Lee, L. R. Dawson, K. J. Malloy, and S. R. J. Brueck, IEEE J. Sel. Top. Quantum Electron. **JSTQE-8**, 972 (2002).
[18]J. Tsao, *Materials Fundamentals of Molecular Beam Epitaxy* (Academic, San Diego, 1993).
[19]J. M. Garcia, J. P. Silveira, and F. Briones, Appl. Phys. Lett. **77**, 409 (2000).
[20]C. W. Wie, J. Appl. Phys. **65**, 2267 (1989).
[21]J. C. Bean, L. C. Feldman, A. T. Fiory, S. Nakahara, and I. K. Robinson, J. Vac. Sci. Technol. A **2**, 436 (1984).
[22]S. F. Yoon, J. Vac. Sci. Technol. B **11**, 562 (1993).
[23]A. G. Cullis, A. J. Pidduck, and M. T. Emeny, J. Cryst. Growth **158**, 15 (1996).
[24]M. Hata, T. Isu, A. Watanabe, and Y. Ktayama, J. Vac. Sci. Technol. B **8**, 692 (1990).
[25]A. Pimpinelli and J. Villian, *Physics of Crystal Growth* (Cambridge University Press, Cambridge, 1998).
[26]J. Tersoff and R. M. Tromp, Phys. Rev. Lett. **70**, 2782 (1993).
[27]S. Luryi and E. Suhir, Appl. Phys. Lett. **49**, 140 (1986).
[28]S. C. Lee, L. R. Dawson, B. Pattada, and S. R. J. Brueck, *46th Electron. Mat. Conf.* (June 23–25, 2004, Indiana).

Downloaded 29 Oct 2004 to 64.106.37.66. Redistribution subject to AIP license or copyright, see http://jap.aip.org/jap/copyright.jsp

HTML ABSTRACT + LINKS

Strain-relieved, dislocation-free $In_xGa_{1-x}As/GaAs(001)$ heterostructure by nanoscale-patterned growth

S. C. Lee,[a] L. R. Dawson, B. Pattada, and S. R. J. Brueck
Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

Y.-B. Jiang and H. F. Xu
Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131

(Received 18 May 2004; accepted 31 August 2004)

A strain-relieved, dislocation-free $In_xGa_{1-x}As$ layer is selectively grown on nanoscale SiO_2-patterned GaAs(001) by molecular beam epitaxy. By localizing the epitaxial area to a periodic array of nanoscale circular holes opened in a SiO_2 mask and allowing the $In_xGa_{1-x}As$ epilayers selectively grown on adjacent holes to coalesce over the SiO_2 mask by lateral overgrowth, the strain of the resulting $In_xGa_{1-x}As$ layer ($x=0.06$) is relieved with a dramatically decreased generation of misfit dislocations. These experimental results qualitatively support the basic idea of the Luryi-Suhir proposal [Appl. Phys. Lett. **49**, 140 (1986)]. © *2004 American Institute of Physics*. [DOI: 10.1063/1.1811799]

Lattice mismatch is an inherent problem in heteroepitaxial growth. As a solution for this classic conundrum, theoretical models have proposed the possibility of strain relief by decreasing the initiating epitaxial areas to the nanoscale regime.[1–4] Among these, the Luryi–Suhir model[1] derives an effective critical thickness depending on the characteristic linear dimension of individual epitaxial areas. According to this continuum mechanics model, the effective critical thickness becomes infinite as the linear dimension of individual nanoscale areas is reduced to the point where the stress is relieved before sufficient energy is accumulated to create a defect such as a misfit dislocation. Thus, there is no thickness limit to the growth of dislocation-free heterostructures for sufficiently small nanoscale epitaxial areas. Recently, we introduced nanoscale-patterned growth (NPG) as a frontier of epitaxial growth for semiconductor nanostructures.[5,6] In this work, we apply NPG to the growth of $In_xGa_{1-x}As/GaAs(001)$ and examine the strain relief for a lattice-mismatched heterostructure selectively grown and laterally coalesced on a nanoscale SiO_2-patterned substrate by molecular beam epitaxy (MBE).

Figure 1 shows 45°-tilted scanning electron microscope (SEM) images of the sample at various stages of the process. The inset to each image is a schematic illustration of the corresponding process step viewed in cross section. First, as shown in Fig. 1(a), a two-dimensional (2D), periodic, nanoscale circular-hole pattern was fabricated in a 45-nm-thick SiO_2 film that had been deposited on a semi-insulating GaAs(001) substrate. The patterning was performed by large-area interferometric lithography and dry etching.[5,7] The pattern period is 355 nm and the hole diameter is about 300 nm.

On this SiO_2-patterned substrate, a 50-nm-thick GaAs buffer layer was selectively deposited at 595 °C by MBE. The resulting surface is shown in Fig. 1(b). For this 50 nm deposition, the GaAs epilayer selectively grown within each hole (GaAs epi-island) does not show clear faceting but has a roughly triangular cross section as seen below.

On the surface shown in Fig. 1(b), a 200-nm-thick $In_xGa_{1-x}As$ layer was then selectively deposited at the same growth temperature by MBE. The deposition conditions for selective growth of $In_xGa_{1-x}As/GaAs(001)$ have been reported elsewhere.[8] At this high temperature, the incorporation rate of In is very low (~0.2%); measurements reported below established the In composition of the film at $x\sim0.06$. While the starting GaAs buffer surface of Fig. 1(b) consisted of a periodic array of discontinuous nonplanar growth seeds, the surface of the as-grown $In_xGa_{1-x}As$ is flat and smooth with only a few incompletely planarized areas that correspond to imperfections/contamination in the original pattern [Fig. 2(c)]. The surface morphology of this $In_xGa_{1-x}As$ layer is significantly improved as compared with comparable thickness $In_xGa_{1-x}As$ layers grown on unpatterned wide-area GaAs(001) substrates under the same growth conditions, which are severely degraded by islanding growth.[8]

Figure 2 shows the details of the $In_xGa_{1-x}As$ layer. Figure 2(a) is a cross sectional transmission electron micrograph (TEM) of the as-grown sample. The inset of Fig. 2(a) is a material map obtained from this TEM image. As seen again in this figure, the $In_xGa_{1-x}As$ layer has a flat surface that completely planarizes the GaAs epi-island buffer and the SiO_2 mask. Figure 2(b) is a cross sectional scanning TEM image of a single period shown in Fig. 2(a). A slight contrast change indicates the boundary between the GaAs buffer and the $In_xGa_{1-x}As$ layer and demonstrates that the $In_xGa_{1-x}As/GaAs$ heterointerface is confined to the original circular holes in the SiO_2 mask. Figure 3 is a high resolution cross sectional TEM image corresponding to the white box of Fig. 2(a). As seen in Fig. 3, the $In_xGa_{1-x}As$ layer initiated from the GaAs buffer surface grows laterally over the SiO_2 mask. The plan-view TEM image of Fig. 2(c) shows an outline of the circular GaAs epi-island buffer bounded by the SiO_2 mask which is buried under the $In_xGa_{1-x}As$ layer after coalescence and planarization. Figures 1(c) and 2 demonstrate the achievement of epitaxial lateral growth of an

[a]Author to whom correspondence should be addressed; electronic mail: sclee@chtm.unm.edu

 © 2004 American Institute of Physics
Downloaded 03 Nov 2004 to 64.106.37.253. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

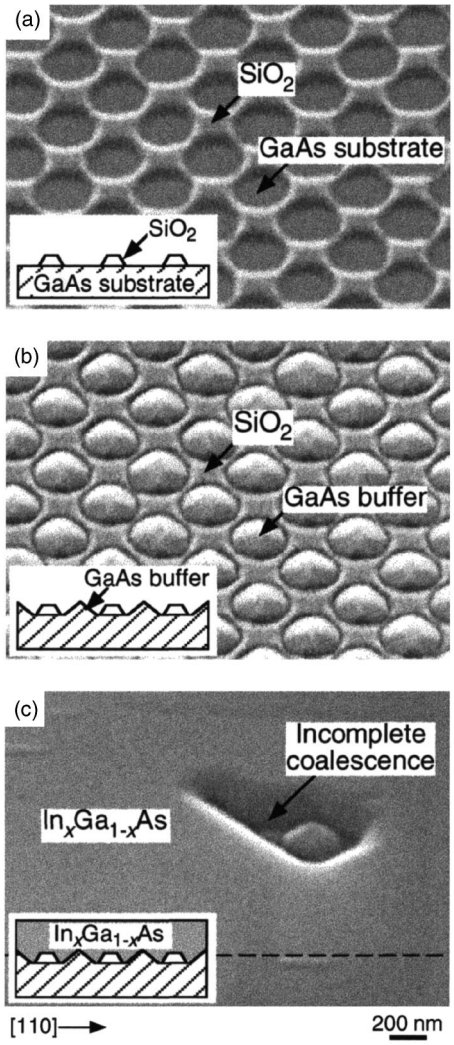


FIG. 1. 45°-tilted SEM images of (a) a 355 nm period 2D array of circular holes fabricated in a 45-nm-thick SiO_2 film, (b) the surface of the sample shown in (a) after selective deposition of a 50 nm GaAs layer, and (c) planarized surface of an as-grown sample after 200 nm deposition of $In_xGa_{1-x}As$ on the surface shown in (b). The inset of each figure is a schematic illustration of the corresponding process step.

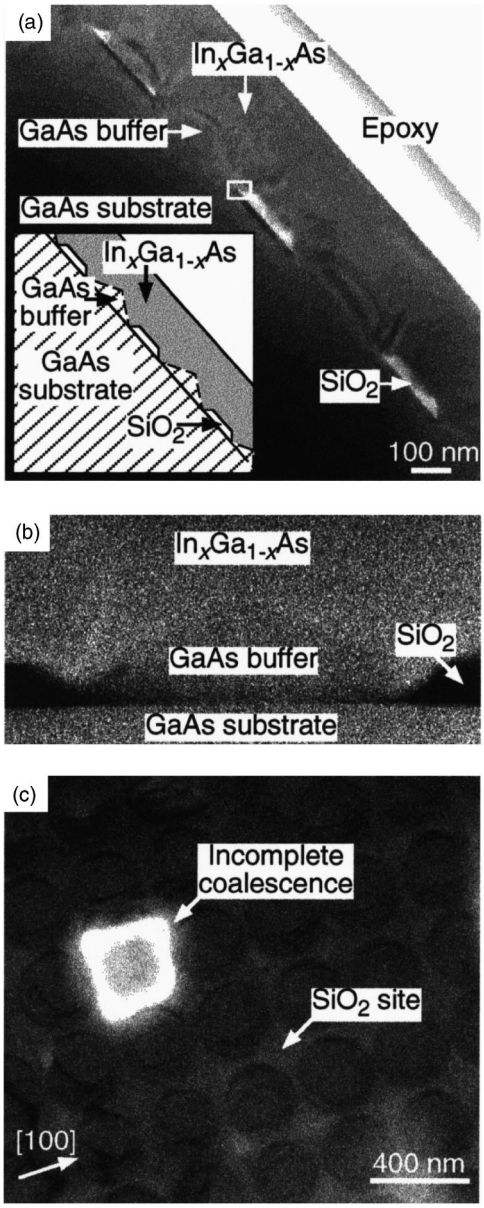


FIG. 2. (a) A cross sectional TEM image of the $In_xGa_{1-x}As$ layer taken along the direction parallel to the dashed line in Fig. 1(c). The inset shows the material map obtained from this image. (b) A cross sectional scanning TEM image of a single period. (c) A plan-view TEM image of the as-grown sample.

$In_xGa_{1-x}As/GaAs$ heterostructure over a SiO_2 mask by MBE.

The thickness of the $In_xGa_{1-x}As$ layer ($\sim$200 nm) is significantly greater than the critical thickness of an $In_xGa_{1-x}As$ layer of the same In composition grown on unpatterned large-area GaAs(001).[9] As seen in Fig. 2(c), however, no dislocations were observed in the plan-view TEM over an area of about 20×10 μm^2. Also, no other crystal defects are observed in the $In_xGa_{1-x}As$ layer of Fig. 2(a). Additional experiments providing a more accurate measure of the dislocation density are presently under way; nonetheless, these results imply that the 200-nm-thick $In_xGa_{1-x}As$ layer grown on the nanoscale SiO_2-patterned GaAs(001) substrate has a dislocation density too low ($<10^6$ cm^{-2}) to estimate in TEM and thus is substantially free from misfit dislocations.

Double-crystal x-ray diffractometry (DXRD) provides critical information on strain relaxation in this heterostructure. From the (004) symmetric and (115) asymmetric reflection rocking curves shown in Fig. 4(a), the In composition, x, and the relaxation, R, can be calculated. The separations between the $In_xGa_{1-x}As$ peak and the GaAs peak are 614.3 and 979.8 arc sec, respectively. From peak separation analysis, an In composition, x, of 0.06 (Misfit $f\sim0.0042$) is obtained. The relaxation, R, which is defined as a ratio of the difference between the in-plane lattice constant of the $In_xGa_{1-x}As$ layer, $a_{\parallel,\mathrm{InGaAs}}$, and the lattice constant of the GaAs substrate, a_{GaAs}, to the difference between the lattice constant of an unstrained $In_xGa_{1-x}As$ layer, a_{InGaAs}, and a_{GaAs} is 0.90.[8] Thus $a_{\parallel,\mathrm{InGaAs}}$ is very close to a_{InGaAs} and the $In_xGa_{1-x}As$ layer selectively grown on the nanoscale SiO_2-patterned GaAs(001) is substantially strain relaxed.

In TEM, as mentioned above, the dislocation density of the $In_xGa_{1-x}As$ layer is significantly lowered as compared with that of a relaxed heterostructure which has a similar thickness and misfit.[10] From DXRD, however, the $In_{0.06}Ga_{0.94}As$ layer is almost fully relaxed. From these experimental results, it can be concluded that the $In_{0.06}Ga_{0.94}As$ layer grown on the nanoscale SiO_2-patterned GaAs(001) substrate is dislocation free but, is also strain relieved.

Downloaded 03 Nov 2004 to 64.106.37.253. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

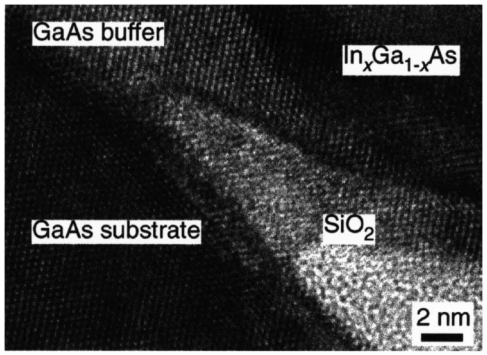


FIG. 3. A high resolution TEM image of the region indicated by a white box in Fig. 2(a).

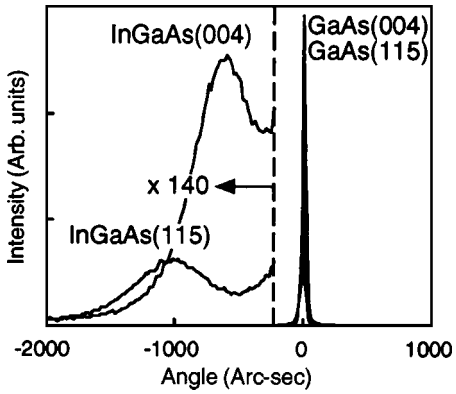


FIG. 4. DXRD rocking curves from the as-grown sample in (004) symmetric and (115) asymmetric reflections.

By growing the GaAs buffer, the structure of our as-grown sample becomes analogous to that of Luryi–Suhir model. The model thickness beyond which the stress exponentially decays for the given hole diameter (~300 nm) and $In_{0.06}Ga_{0.94}As$ is ~50 nm. This is considerably smaller than the $In_{0.06}Ga_{0.94}As$ thickness of our film. Although this comparison provides qualitative justification that nanoscale patterning contributes to the observed strain relief, our sample structure has several differences from that suggested in the model. As seen in Figs. 1(b) and 2(b), the convex shape of the GaAs epi-island buffer results in an expansion of the effective area of the GaAs buffer/$In_{0.06}Ga_{0.94}As$ heterointerface within each circular hole. Also, as seen in the inset of Fig. 2(a), this further leads to the modulation of the $In_{0.06}Ga_{0.94}As$ thickness across each pattern unit cell. These effects could partly contribute to strain relief of the $In_{0.06}Ga_{0.94}As$ layer observed in DXRD.

There have been several studies of strain relaxation of limited-area, lattice-mismatched heterostructure growth.[2–4,11,12] While the Luryi–Suhir model focuses on the variation of the biaxial strain as the growth thickness increases on a periodically patterned substrate, other models concentrate on the variation of the stress across the growth area by considering edge relaxation.[2,4,12] According to all of these models, the isolated $In_xGa_{1-x}As$ epilayer grown on each GaAs epi-island buffer ($In_xGa_{1-x}As$ nano-island) of Fig. 1(b) undergoes stress relaxation leading to an exponential decay along the entire surface. If this stress relaxation is completed before coalescence of the individual nano-islands, the resulting film could be also strain relieved after coalescence. The stress decay length, which is a characteristic length effective only in nanoscale patterned growth, undoubtedly depends on the geometrical shape and orientation of each nano-island as well as its elastic properties and the lattice mismatch. The stress decay length of ~50 nm for the $In_{0.06}Ga_{0.94}As$ on our nanopatterned GaAs(001) estimated from the Luryi–Suhir model is comparable to the growth amount of the $In_{0.06}Ga_{0.94}As$ required for coalescence which is 50–100 nm deposition under the present growth conditions. This implies that the strain relief before coalescence is likely a major relaxation process for this growth.

Our experimental results include coalescence effects that have not yet been theoretically evaluated in detail. Coalescence between fully or partially relaxed nano-islands and associated quenching/healing of crystal imperfections as two (or more) nano-islands conjoin is a complex process deserving of significant further investigation. Nonetheless, it is experimentally evident that nanoscale patterning plays an important role in the strain relief of a lattice-mismatched heterostructure resulting in a very significant reduction in the density of misfit dislocations.

We have demonstrated MBE growth of strain-relieved dislocation-free $In_{0.06}Ga_{0.94}As$ on nanoscale SiO_2-patterned GaAs(001) with consistent material characterizations. The experimental results are qualitatively consistent with the Luryi–Suhir model. Once coalescence and planarization are achieved over a SiO_2 mask in $In_xGa_{1-x}As$ deposition, the growth temperature and growth rate can be adjusted to the optimal conditions for the growth of a device structure to suppress contamination issues and to improve the quality of the epilayers. These conditions are typically very different from those for selective growth. Growth of InAs quantum dots on such a strain-relieved dislocation-free $In_xGa_{1-x}As$ layer is a clear future research direction, aimed at long wavelength emission on GaAs. Additionally, residual strain and its dependence on the $In_xGa_{1-x}As$ thickness and on pattern parameters such as period, hole size, pattern symmetry, etc., as well as on thermal expansion mismatch, requires further investigation.

Support for this work was provided by the DOE and by DARPA.

[1] S. Luryi and E. Suhir, Appl. Phys. Lett. **49**, 140 (1986).
[2] A. Fischer and H. Richter, Appl. Phys. Lett. **61**, 2656 (1992).
[3] D. Zubia and S. D. Hersee, J. Appl. Phys. **85**, 6492 (1999).
[4] F. Y. Huang, Phys. Rev. Lett. **85**, 784 (2000).
[5] S. C. Lee, K. J. Malloy, and S. R. J. Brueck, J. Appl. Phys. **90**, 4163 (2001).
[6] S. C. Lee, K. J. Malloy, L. R. Dawson, and S. R. J. Brueck, J. Appl. Phys. **92**, 6567 (2002).
[7] S. R. J. Brueck, S. H. Zaidi, X. Chen, and Z. Zhang, Microelectron. Eng. **41/42**, 145 (1998).
[8] S. C. Lee, L. R. Dawson, A. Stintz, and S. R. J. Brueck, J. Appl. Phys. (to be published).
[9] J. W. Matthews and A. E. Blakeslee, J. Cryst. Growth **27**, 118 (1974).
[10] Y. Fukuda, Y. Kohama, M. Seki, and Y. Ohmachi, Jpn. J. Appl. Phys., Part 1 **27**, 1593 (1988).
[11] Y. H. Xie and J. C. Bean, J. Vac. Sci. Technol. B **8**, 227 (1990).
[12] E. Suhir, J. Appl. Mech. **53**, 657 (1986).

Downloaded 03 Nov 2004 to 64.106.37.253. Redistribution subject to AIP license or copyright, see http://apl.aip.org/apl/copyright.jsp

Nanoscale two-dimensional patterning on Si(001) by large-area interferometric lithography and anisotropic wet etching

S. C. Lee and S. R. J. Brueck
Department of Electrical and Computer Engineering and Center for High Technology Materials, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

(Received 18 December 2003; accepted 17 May 2004; published 13 August 2004)

A fabrication technique for nanoscale two-dimensional (2D) patterning by large-area interferometric lithography (IL) and anisotropic wet etching (AWE) is reported. On a Si(001) substrate corrugated by one-dimensional (1D) periodic V-grooves in a first IL/AWE step, an additional 1D IL/AWE patterning perpendicular to the groove direction is performed. In this process, {111} facets—the sidewalls of a V-groove having an extremely low etch rate in AWE—are utilized as an etch barrier to confine AWE to a nanoscale-wide (001) stripe facet between V-grooves. IL on the corrugated Si surface results in an array of photoresist (PR) lines localized on each (001) stripe. The subsequent AWE with a Cr etch mask fabricated from this unusual PR pattern thus begins from the inner side of (001) stripe but slows down as it reaches the (001)-{111} facet intersection of the V-groove. The {111} planes at each facet intersection play the important role of an etch barrier laterally confining the etch area within each (001) stripe. The direction along the stripe is also confined with the formation of 2D {111} inverted pyramids. A 355 nm period, 2D array of Si mesas having a rectangular top (001) facet of area ~150 nm×110 nm and surrounded by various (111) facets in a unit period is fabricated on a Si(001) substrate. © *2004 American Vacuum Society.* [DOI: 10.1116/1.1771663]

I. INTRODUCTION

Recent progress in nanoscale semiconductor technology shows potential for realizing the novel characteristics and high performance required for the next-generation of electronic/optoelectronic devices. By scaling feature sizes to the nm-range in patterned growth of semiconductor thin films, we have demonstrated various growth phenomena which are not available with μm- or larger-scale patterning.[1–3] One such phenomenon is orientation-dependent migration and incorporation which utilizes surface migration of atoms incident on a nanoscale faceted (nanofaceted) surface and crystal orientation-dependent adatom incorporation. Reducing the lateral dimension of a facet to less than the surface migration length of an adatom allows new regimes of crystal growth that will lead to new device capabilities.

In our previous work, we reported a deposition technique for the fabrication of nanofaceted surfaces (nanofaceting).[1,2] Nanofaceting can be used for the fabrication of one- to three-dimensional (1D to 3D) periodic structures or particular shaped features required for nano- and μm-scale Si electronics and machining.[4,5] Anisotropic wet etching (AWE) is a reliable and cost-effective alternative for nanofaceting.[4–8] In this work, we present a fabrication technique using a {111} facet which has a significantly lower etch rate than the (001) orientation in AWE, as an etch barrier. The first step is formation of a nanoscale 1D pattern by large area interferometric lithography (IL) and etching to form a traditional V-groove pattern revealing alternate (100) flats and {111} V-grooves. After a second orthogonal IL step, AWE is confined to a nanoscale-wide (001) stripe facet that is already bounded by neighboring {111} facets in one direction. This effectively results in the confinement of AWE to the formation of inverted {111} pyramids inside the etch area and results in a 2D pattern restricted to the (001) stripe. This is the basic motivation of this work. With this technique, we have achieved a 355 nm period 2D array of Si mesas with a rectangular top (001) facet having an area ~150 nm×110 nm surrounded by various (111) type facets in a unit period. To our knowledge, this is the smallest period 2D pattern ever made by AWE.

In Sec. II, 1D PR patterning on a corrugated Si(001) surface by IL is described. This is a critical step for the confinement of AWE to each (001) stripe. IL on a nanoscale nonplanar surface is very different from that on a wide-area planar surface and is utilized for the localized photoresist (PR) patterning on a (001) stripe. The intensity distribution of UV light on a corrugated Si surface leading to this unusual PR pattern is discussed. In Sec. III, the fabrication of a Cr etch mask from the PR pattern is described. In Sec. IV, the results of AWE performed on the periodic V-grooved Si(001) with the Cr etch mask is presented. Finally, we summarize all of the results in Sec. V.

II. PR PATTERNING ON PERIODIC V-GROOVED Si(001)

Figure 1 shows a 45°-tilted scanning electron microscopy (SEM) image of a 355 nm period 1D PR pattern generated atop a periodically V-grooved Si surface by large-area IL which is a low-cost technology for sub- 100 nm-scales.[9] A 355 nm wavelength, frequency-tripled YAG laser was used as an ultraviolet (UV) light source. A boron-doped Si(001) substrate having a resistivity of about 8 to 15 Ω cm was used. Aqueous KOH, which exhibits strongly orientation dependent etching on a Si surface, was employed.[6] Each

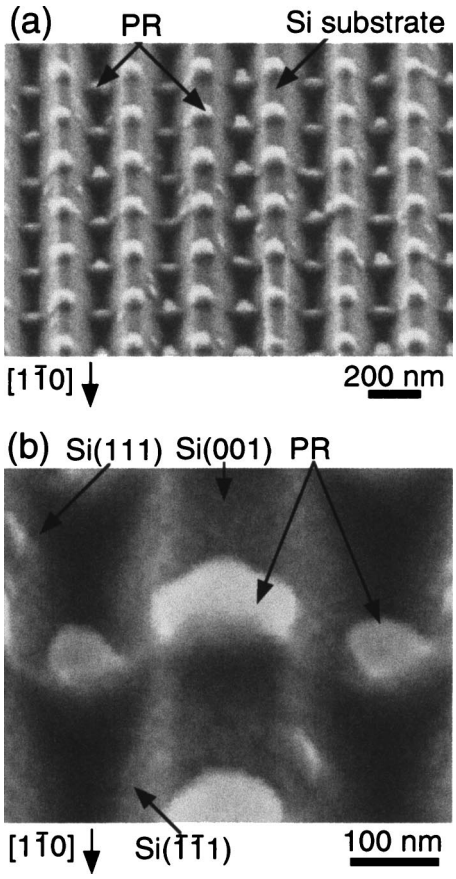


FIG. 1. (a) A 45°-tilted SEM image of a 1D stripe PR pattern generated on a periodic V-grooved Si surface. (b) Magnification of (a). The discontinuity of the PR stripes near the (001)-{111} facet intersections is evident.

V-groove consists of (111) and $(\bar{1}\bar{1}\bar{1})$ sidewalls oriented along a [110] direction. These grooves were fabricated by IL and AWE using a Cr etch mask. The width of the (001) stripes between V-grooves is about 130–170 nm. The direction of the PR lines was chosen to be perpendicular to that of the prepatterned (001) stripes. The PR was Shipley 5206E diluted to provide a thickness of about 200 nm.

The PR pattern of Fig. 1(a) is very different from that on a planar surface. There are voids in the PR lines at every (001)-{111} intersection. Here, a (001)-{111} intersection means the line along which each (001) stripe facet meets the neighboring {111} facets; each period has two such intersections. Figure 1(b) clearly reveals the discontinuity of the PR stripe near the (001)-{111} intersection. Thus, the PR film separately remains on top of each (001) facet and inside each V-groove.

The cross section SEM image of Fig. 2(a) shows the photoresist film before the development step; both the top (001) faces and the V-grooves are filled with PR. The PR thickness near the (001)-{111} intersections is comparable to that on the (001) stripes. As seen in Fig. 2(b), development mainly removes the PR at the intersections. Thus, the unusual discontinuity in the PR lines of Fig. 1 is not due to the thickness variation across the V-groove.

Generally, IL has been performed on a planar surface and

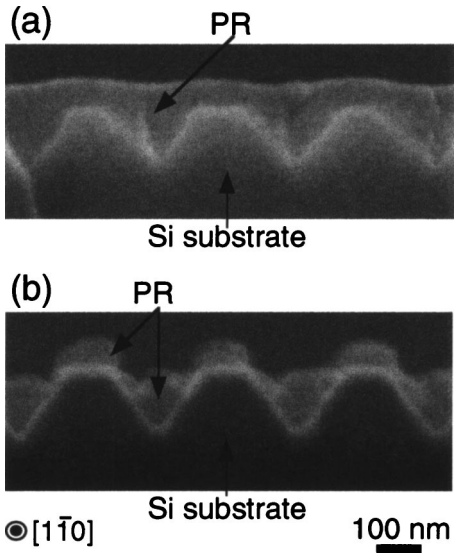


FIG. 2. Side view SEM images of (a) a PR film (a) before and (b) after development (Fig. 1).

the intensity distribution of UV light within a PR film is readily calculated.[10] The major factors that determine the UV intensity are the incident power and polarization, the incident angle, and the reflection from the PR-substrate interface. If IL is applied to a nonplanar surface, many of these factors are affected by the surface topography. The PR persistently remaining within each V-groove in Fig. 3(c) can be the result of a lower UV intensity in this region than on a (001) stripe. If UV incident angle is set to 30° on a (001) plane for a 355 nm period, it becomes 52.1° on a tilted (111) facet in our IL apparatus. Thus, the effective UV dose supplied on (111) plane is lower than that on a (001) plane and the inside of a V-groove is relatively under-exposed. At PR-Si interface around a (001)-{111} intersection, the UV radiation is scattered rather than reflected. This means that UV reflection near this area is very different from that of the remaining region in a single period. Therefore, the disconnected PR lines of Fig. 1 could be due to the nonuniform dose arising from the different reflectivity of the UV beam on each facet and facet intersection in this complex IL geometry. As seen in Sec. IV, this discontinuity enables nanoscale 2D patterning localized on each (001) stripe.

III. PREPARATION OF A Cr ETCH MASK

Figure 3(a) shows a top view SEM image of the PR pattern similar to Fig. 1(a). In Fig. 3(a), the PR on each (001) stripe is close to a linear array of ellipses. These PR features generate openings localized on a (001) stripe in a metal deposition and lift-off process. On the PR pattern of Fig. 3(a), a 10-nm-thick Cr film which will be used as an etch mask in the subsequent AWE was deposited by electron beam evaporation at a deposition rate of 0.2 nm/s.

Figure 3(b) shows a top-view SEM image of a Cr etch mask generated on the periodic V-grooved Si(001) after PR lift-off. As indicated by the arrows, two different openings through which AWE can occur were generated in the Cr film:

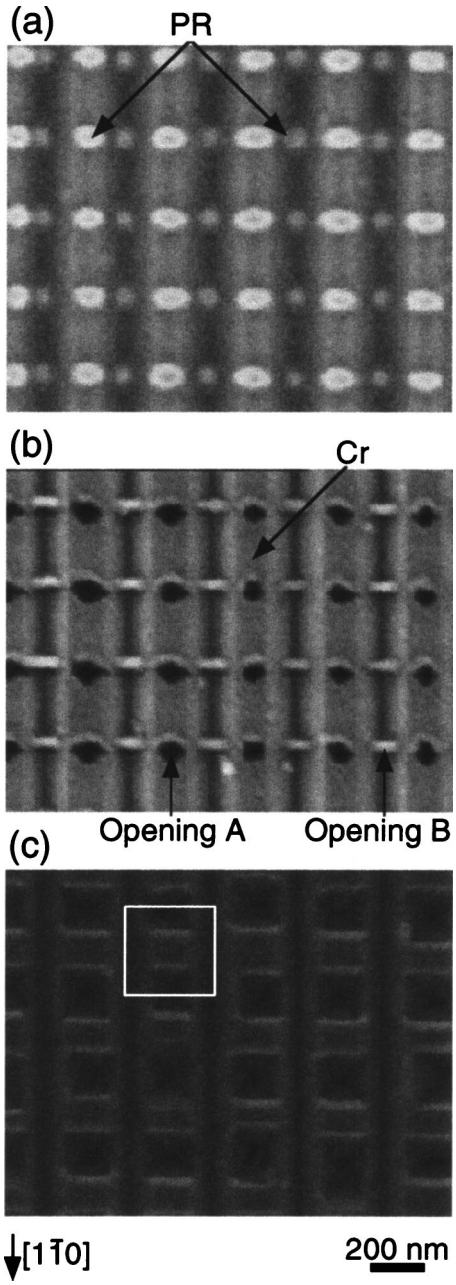


FIG. 3. Top view SEM images of: (a) the PR pattern shown in Fig. 1; (b) a 10-nm-thick Cr etch mask after PR lift-off; and (c) the Si surface after a 50 s AWE. In (b), two different types of openings in the Cr etch mask are indicated: A [on the top (100) facet]; and B [atop the {111} sidewalls]. In (c), the dashed box represents the unit cell of the 2D pattern.

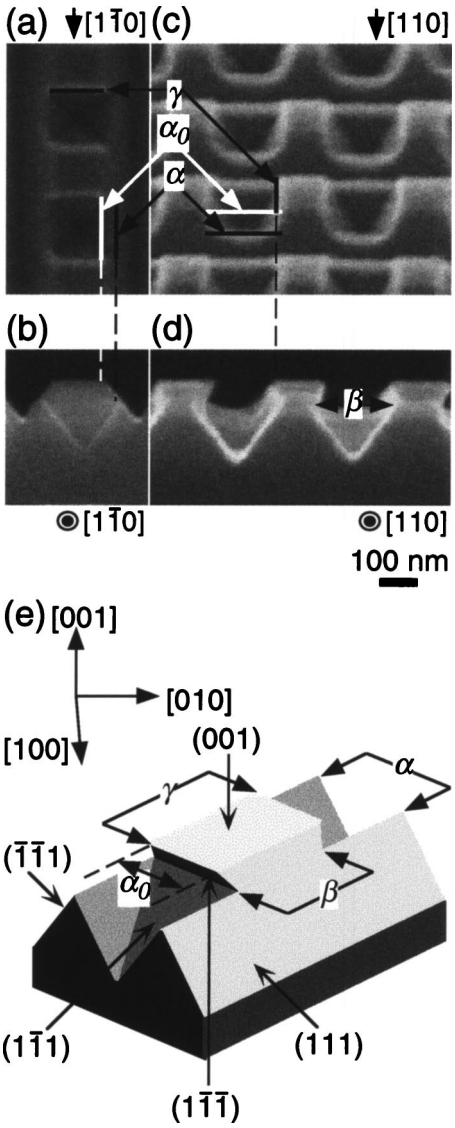


FIG. 4. (a) A top view and (b) a side view SEM image of the 50 s etched sample. (c) A 45°-tilted and (d) a side view SEM image of the same sample. Sample orientations are indicated with various facet intersection lines. See the text for details. (e) Schematic illustration of a single Si mesa corresponding to the dashed box in (c) with the facet identification.

one (labeled A) is on each (001) stripe and the other (labeled B) is at the valley of each V-groove. The opening A has very different size roughly ranging from 70 to 150 nm.

IV. ANISOTROPIC WET ETCHING (AWE)

For AWE, 25 wt.% aqueous KOH solution at 60±5 °C was used. A KOH-based etchant has a high etch selectivity for (001) [anisotropy ratio of (001) to (111)≳400 at 80 °C].[6] The etch rate calibrated on a wide-area Si(001) surface for our conditions was about 7±1 nm/s. For AWE, the sample shown in Fig. 3(b) was dipped in the etchant for 50 s.

Figure 3(c) shows a top view image of the etched Si surface after removing the Cr film. In Fig. 3(b), as mentioned above, there are two classes of openings in the Cr mask where the etchant can access the Si surface. As seen in Fig. 3(c), however, AWE proceeded only through the "A" openings on the (001) stripes. This is because the "B" openings are atop the very low etch-rate (111) V-groove facets. Thus, the disconnected PR pattern of Fig. 1 results in selective etching of the (001) stripe regions.

Generation of facets inside of the etch area is observed in Fig. 3(c). Figure 4 shows magnified images that were taken from four different directions. Figures 4(a) and 4(b) were taken along a [110] direction, whereas Figs. 4(c) and 4(d) were obtained along a [110] direction. Figures 4(b) and 4(d) are both side view, Figs. 4(a) and 4(c) are a top view and a 45°-tilted SEM image of the same sample, respectively. Fig-

ures 4(b)–4(d) clearly reveal the formation of facets inside of the etch area. All the facets inside the etch area generated by AWE are close to {111}. Thus, the 3D shape of the etch area on the (001) stripe in Fig. 4(a) [or Fig. 3(c)] is an inverted pyramid. Also, AWE generates several facet intersections consisting of (111) and (111) facets, (11) and (111) facets, and (001) and (111) facets labeled α, β, and γ, respectively. Here, α which originally corresponded to a (001)-(111) intersection (α_0) is changed to a (111)-(111) intersection after etching. The dashed box of Fig. 3(c) represents a unit period of the 2D pattern. A Si mesa with a rectangular top (001) facet having a size of about $150\pm20\times110\pm10$ nm^2 is surrounded by different {111} facets within each period. Figure 4(e) shows a schematic illustration of a single Si mesa corresponding to the dashed box in Fig. 3(c) along with the facet identification. In the nanoscale range, therefore, AWE is effectively terminated revealing four different internal {111} facets for a 50 s etch. This facet geometry also restricts AWE longitudinally along a (001) stripe and leads to a confined 2D patterning.

Generally, AWE inside the etch area should be terminated once it completely reveals the four different {111} facets. As seen in Fig. 4(b), however, the progression of the interface position α which was originally at the starting (001)-(111) intersection (α_0) to a new position further along the (111) face after AWE implies that the etching does not perfectly stop at a (111)-(111) intersection but removes the {111} facet with a finite etch rate. Then, the etch rate of the (111) facet deduced from the progression of α is about 2 nm/s if α is taken as the distance of $\sim$80 nm from α_0 to its current location in Fig. 4(b) in $\sim$50 s. This is considerably slower than the calibrated etch rate of $\sim$7 nm/s on a (001) plane under the same conditions. Even with this finite etch rate, therefore, a (111) V-groove facet plays the role of an etch barrier in AWE.

The $\sim$2 nm/s etch rate of the {111} face in the KOH solution is much higher than the value predicted from the reported large-area anisotropy ratio.[5] In Fig. 4, AWE forms another unusual structure, the undercutting of the interfaces perpendicular to the original stripes to form a retrograde (111)-(111) interface, denoted by β. The shape of the Si mesa shown in Figs. 4(b) and 4(d) suggests that the formation of β is due to the lower etch rate of the (001)-(111) interface, γ, from which the current intersection evolves [see Fig. 4(e)] than that of α. This means that the (111) facet of a (001)-{111} intersection has an etch rate much lower than $\sim$2 nm/s in our process, as expected. This is also supported by the inactive etching through the ''B'' openings. Further study is necessary to understand the etching mechanism of (111) facet of a {111}-{111} intersection on a Si surface.

Figure 3(c) suggests that the ''A'' openings having different sizes and shapes in the Cr mask were forced into a uniform-size, square-shape etch area on the Si surface in the top view of Fig. 3(c). This improvement is likely also due to AWE on this complex multiple-surface substrate. This also means that the uniformity of the width of each (001) stripe along and across the groove direction which is determined in the first step 1D patterning is more important to the final 2D pattern than that of the additional patterning for the ''A'' openings.

V. SUMMARY AND CONCLUSIONS

A fabrication technique for 2D nanoscale patterning by large-area IL and AWE has been reported. On a Si(001) substrate corrugated by 1D periodic V-grooves, a second 1D patterning perpendicular to the groove direction was performed by IL and AWE. Nanoscale 2D patterning has been demonstrated utilizing {111} facets, the V-groove sidewalls, as etch barriers for confining AWE to the top of the nanoscale-wide (001) stripe. AWE initiated on a (001) stripe extends laterally but slows down after reaching both (001)-(111) facet intersections. Laterally confined AWE induces the same effect on the longitudinal direction along the (001) stripe. Such 2D confinement of AWE leads to a uniform and reproducible nanoscale 2D pattern on Si(001). With this technique, a 355 nm period 2D array of Si mesas with a rectangular top (001) facet of an area $\sim$150 nm$\times$110 nm surrounded by various (111) type facets in a unit period has been achieved. Nanofaceting by IL and AWE might be important for future semiconductor technologies such as nanoscale patterned growth and nanoscale electromechanical systems (NEMS).

ACKNOWLEDGMENTS

Partial support for this work was provided by the ARO/MURI in Deep Subwavelength Optical Nanolithography, DARPA, and the DOE.

[1]S. C. Lee, L. R. Dawson, K. J. Malloy, and S. R. J. Brueck, Appl. Phys. Lett. **79**, 2630 (2001).
[2]S. C. Lee, L. R. Dawson, K. J. Malloy, and S. R. J. Brueck, IEEE J. Sel. Top. Quantum Electron. **JSTQE-8**, 972 (2002).
[3]S. C. Lee, X. Y. Sun, S. D. Hersee, and S. R. J. Brueck, Appl. Phys. Lett. **84**, 2079 (2004).
[4]F. S.-S. Chien, C.-L. Wu, Y.-C. Chou, T. T. Chen, S. Gwo, and W.-F. Wsieh, Appl. Phys. Lett. **75**, 2429 (1999).
[5]C. R. Tellier and S. Durand, Sens. Actuators, A **60**, 168 (1997).
[6]K. E. Bean, IEEE Trans. Electron Devices **ED-25**, 1185 (1978).
[7]K. Itoh, K. Iwameji, and Y. Sasaki, Jpn. J. Appl. Phys., Part 1 **30**, L1605 (1991).
[8]S. Wolf and R. N. Tauber, *Silicon Processing* (Lattice Press, New York, 1986).
[9]A. K. Raub and S. R. J. Brueck, Proc. SPIE **5040**, 667 (2003); A. K. Raub, A. Fraunglass, S. R. J. Brueck, W. Conley, R. Dammel, A. Romano, M. Sato, and W. Heinsberg, *ibid.* (to be published).
[10]L. F. Johnson, G. W. Kammlott, and K. A. Ingersoll, Appl. Opt. **17**, 1165 (1978).

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Crystal Growth 272 (2004) 2–8
JOURNAL OF CRYSTAL GROWTH
www.elsevier.com/locate/jcrysgro

Phase control of GaN on Si by nanoscale faceting in metalorganic vapor-phase epitaxy

S.C. Lee*, X.Y. Sun, S.D. Hersee, S.R.J. Brueck

Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard, SE, Albuquerque, NM 87106, USA

Abstract

Phase stability of GaN on (1 1 1)–(0 0 1) nanoscale faceted Si is investigated. Nanoscale faceting (nanofaceting) is accomplished on a Si(0 0 1) substrate with interferometric lithography and anisotropic wet etching. By relying on the nucleation and associated lateral growth depending on orientation and crystal structure, spatial separation of a cubic-phase region from hexagonal-phase GaN regions which initiate on the facing Si(1 1 1) sidewalls of a 355-nm period V-grooved Si surface is achieved. The appearance of cubic-phase region is associated with a rapid surface planarization with its principal crystal axes parallel to those of the Si(0 0 1) substrate. The boundary between cubic and hexagonal phases is revealed along the Si(1 1 1) sidewalls and extends from each V-groove to the GaN surface for a 600-nm deposition. The phase separation of GaN selectively grown on a nanoscale faceted Si surface and the comparison of its material properties (photoluminescence and Raman scattering for stress measurement) with those of GaN grown on wide-area Si(1 1 1) and Si(0 0 1) imply that nanofaceting can be utilized for phase control of GaN on a Si(0 0 1) substrate.
© 2004 Elsevier B.V. All rights reserved.

PACS: 81.15.Gh

Keywords: A1. Crystal structure; A1. Nanofaceting; A1. Phase transition; A3. Metalorganic vapor-phase epitaxy; B1. Nitrides

1. Introduction

Growth of GaN on a Si substrate offers the exciting potential of integration of GaN devices with well-established Si technology. In spite of this advantage, however, the large lattice and thermal mismatches between these materials remain a critical problem. Several attempts for growth of GaN on a Si(1 1 1) substrate such as wafer patterning to induce lateral growth [1,2] or to confine the growth area [3–5] and various kinds of buffer or intermediate layers [3–6] have yielded improved optical/electrical properties and suggest a possibility of high-quality GaN on a Si substrate.

*Corresponding author.
E-mail address: sclee@chtm.unm.edu (S.C. Lee).

0022-0248/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jcrysgro.2004.08.074

For integration with most Si electronics technology, growth on Si(0 0 1) is much more favorable than on Si(1 1 1). However, deposition of GaN on a (0 0 1) plane results in poly-type rough surface morphology which is not suitable for most device fabrication [5–8]. Some theoretical calculations suggest a relatively small energy difference between hexagonal and cubic phases of GaN compared with other III-nitrides as one of the possible explanations [9,10]. Together with the problem already mentioned, the phase instability of GaN on Si(0 0 1) is another important issue which must be investigated for application to Si technology.

Recent progress in lithography technology and its application to molecular beam epitaxy and metalorganic vapor-phase epitaxy (MOVPE) has made accessible a new regime of growth—nanoscale patterned growth and nano-heteroepitaxy [11,12]. In this work, we address the issue of phase instability with a nanoscale patterning technique. Experimentally, we found that GaN begins to grow with hexagonal phase on the Si{1 1 1} sidewalls of a nanoscale V-grooved (nanogrooved) Si surface, but changes to cubic phase in continued deposition. The cubic-phase GaN is thus spatially separated from the hexagonal-phase region in the resulting crystalline structure with its principal crystal axes parallel to those of the Si(0 0 1) substrate. The morphology of the GaN surface is also significantly improved. These imply that nanofaceting on Si(0 0 1) is a promising phase control technique for single-phase GaN on a limited area of Si(0 0 1). In this work, the phase stability and stress distribution of GaN grown on a nanogrooved Si surface is investigated and compared with those of GaN grown on wide-area Si(1 1 1) and Si(0 0 1).

2. Nanofaceting and MOVPE growth

A 355-nm period one dimensional (1D) nanoscale V-grooved surface is fabricated on a boron-doped 8–15 Ω cm Si(0 0 1) substrate by *i*-line interferometric lithography and aqueous KOH anisotropic wet etching [13]. The pattern direction was chosen so that the anisotropic wet etching revealed {1 1 1} type sidewall orientations in the grooves. Fig. 1 shows cross section scanning electron microscopy (SEM) images of a nanogrooved Si surface. As shown in the bottom part of Fig. 1, each V-groove has well-defined two facing {1 1 1} type sidewalls connected to neighboring grooves by a sub-100 nm-wide (0 0 1) stripe facet. Thus, each V-groove is isolated by the (0 0 1) stripe facets.

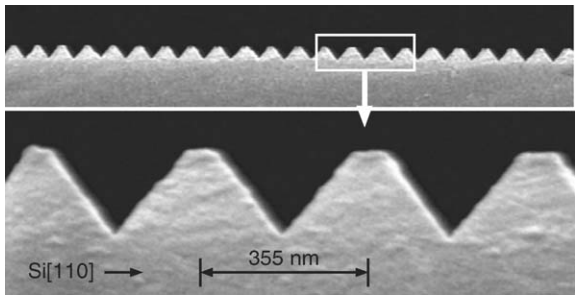


Fig. 1. Cross-sectional SEM images of a 355-nm period nanogrooved surface fabricated on a Si(0 0 1) substrate.

Three different substrates were examined in a single run growth; the nanogrooved Si(0 0 1) shown in Fig. 1, wide-area Si(1 1 1), and wide-area Si(0 0 1). On these substrates, a GaN epilayer was grown by MOVPE with trimethylgallium (TMG) and NH_3. First, the substrates were preheated up to 1100 °C under hydrogen environment and cooled down to 650 °C. At this temperature, a ~1.5-nm thick GaN buffer was grown on each Si substrate. The low-temperature buffer was employed to assist nucleation over the whole substrate surface. To avoid reactions between NH_3 and the Si surface, which could result in Si-nitridation at the beginning of growth, both TMG and NH_3 were simultaneously supplied to the reactor for the buffer growth. On top of this buffer, a 600-nm thick GaN epilayer was grown at 1050 °C. Nominal growth rates were 1.5 nm/min for the buffer and 10 nm/min for the GaN epilayer.

3. Results and discussion

Fig. 2(a) shows a top-view SEM image of the as-grown GaN on the Si nanogrooved substrate. As confirmed in the cross-sectional transmission microscopy (XTEM) below, the surface shown in

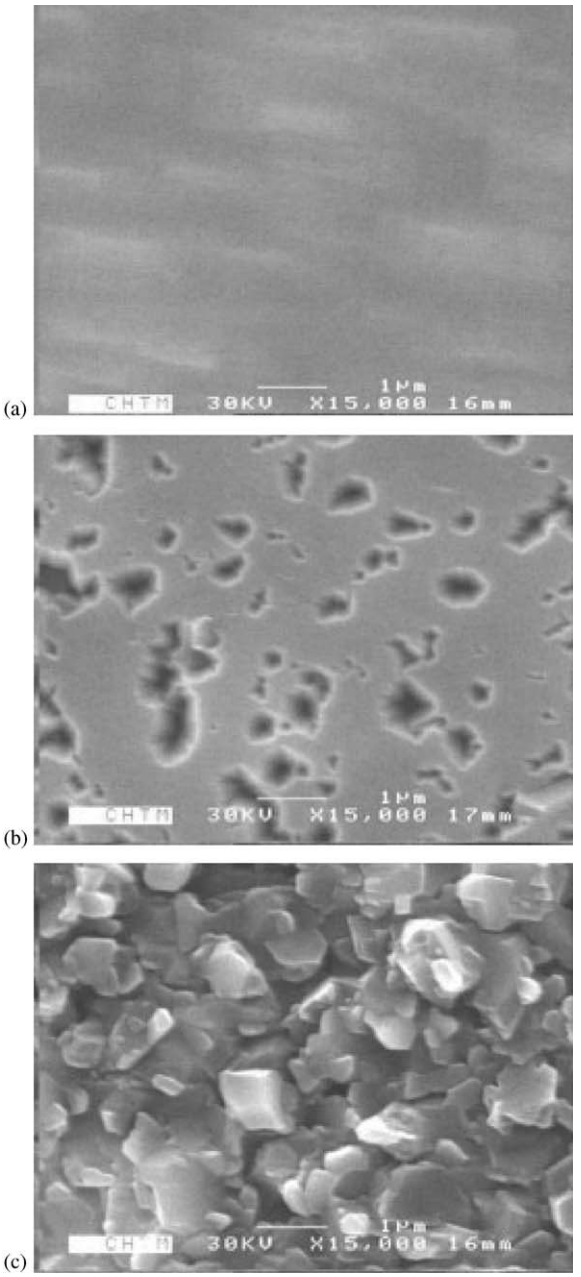


Fig. 2. Top view SEM images of the 600-nm thick GaN on (a) Si nanogrooves, (b) wide-area Si(1 1 1), and (c) wide-area Si(0 0 1) grown under the same conditions.

Fig. 2(a) is smooth and flat. Figs. 2(b) and (c) are top-view SEM images of the GaN deposited on wide-area Si(1 1 1) and Si(0 0 1) under the same growth conditions as that of Fig. 2(a), respectively. Although the GaN on Si(1 1 1) is degraded by a number of μm-size irregular dips, it has a relatively flat surface. On the other hand, the GaN on Si(0 0 1) has a severely degraded morphology which is analogous to polycrystalline film surfaces. This granular or poly-type surface is the typical surface morphology of GaN on wide-area Si(0 0 1) [7,8]. Although the original orientation of the substrate surface was (0 0 1), the GaN surface of Fig. 2(a) is very different from that of Fig. 2(c) because of the impact of the nanofaceting on epitaxial growth [4–8]. The comparison of Fig. 2(a) with Fig. 2(c) implies that nanofaceting significantly improves the surface morphology of GaN on Si(0 0 1).

Fig. 3(a) shows an XTEM image of the as-grown GaN on the Si nanogrooves. This figure confirms that the GaN surface is relatively smooth and flat. As indicated by arrows in Fig. 3(a), an array of voids located on Si(0 0 1) stripe facets between V-grooves is observed at the GaN/Si interface. The inset of Fig. 3(b) reveals that the

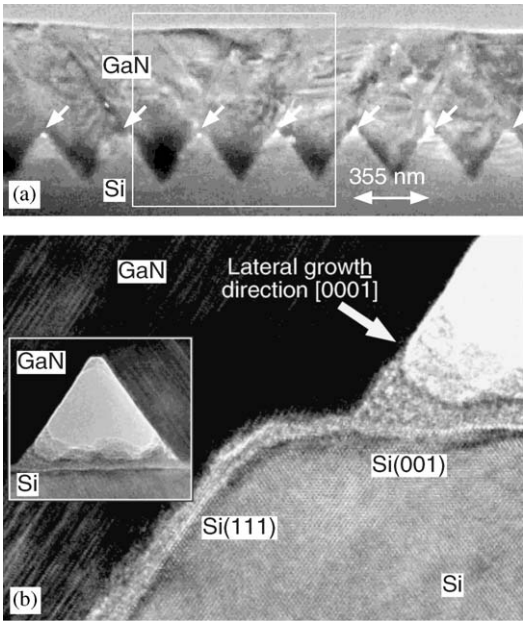


Fig. 3. (a) An XTEM image of the 600-nm thick GaN on Si nanogrooves. The arrows indicate an array of voids on the Si(0 0 1) stripe facets. (b) A high-resolution XTEM image taken near the bottom left corner of a void shown in (a). The inset is a magnification of a single void.

shape of the void (or a nano-channel along the groove direction) is triangular. A similar shape void which is due to selective nucleation and coalescence of GaN epilayers grown on the facing Si(1 1 1) sidewalls of a single groove has been reported previously [14]. Fig. 3(b) which is a high-resolution XTEM image taken near the bottom left corner of the void, reveals that the GaN has hexagonal phase at the initial stage of growth of which the *c*-axis ([0 0 0 1]) is normal to Si{1 1 1} and is highly defected by stacking faults because of large lattice mismatch between GaN and Si. As shown in Fig. 3(b), lateral overgrowth from the neighboring V-grooves toward the Si(0 0 1) stripe facet along GaN[0 0 0 $\bar{1}$] is involved in the formation of the void. Fig. 3, therefore, implies that the planarization of Fig. 2(a) is partly associated with lateral overgrowth and coalescence over the Si(0 0 1) stripe facets.

High-resolution XTEM and selected area electron diffraction (SAED) provide important information about planarization and phase separation. Fig. 4 is a magnified XTEM image of a single V-groove (the white box in Fig. 3(a)). Figs. 5(a)–(f) are high-resolution XTEM images and SAED patterns taken from several areas of Fig. 4. The regions labeled A–F of Fig. 4 correspond to Figs. 5(a)–(f). As seen in these figures, the 600-nm thick GaN exhibits a spatial variation in crystal-

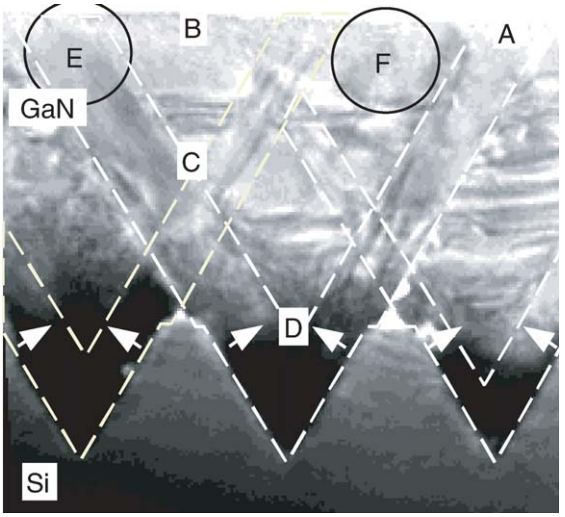


Fig. 4. A magnified XTEM image of the white box in Fig. 3(a).

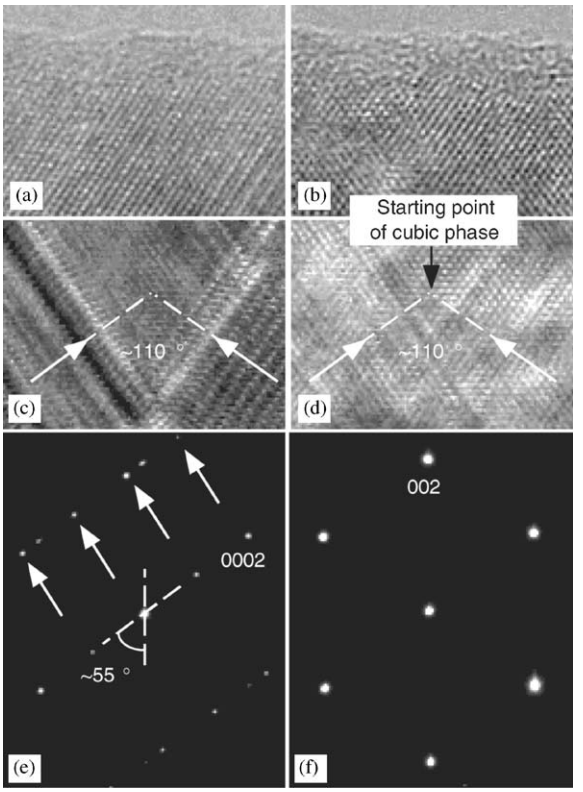


Fig. 5. High-resolution XTEM images [(a)–(d)] and SAED patterns [(e) and (f)] taken from the cross section of the 600-nm thick GaN on Si nanogrooves shown in Fig. 4. The regions labeled in Fig. 4 correspond to (a)–(f) in alphabetical order. In (e), the array of the white arrows indicates the diffraction spots from the hexagonal-phase GaN.

line phases across each period. While regions A and E are hexagonal phase, regions B and F show clear cubic phase. On the other hand, regions C and D have cubic phase over a highly defected region where two hexagonal phase regions are crossing with an angle of $\sim 110^\circ$ in their *c*-axis alignment. This misalignment is due to the $\sim 55^\circ$ tilting of the GaN(0 0 0 1) [or Si(1 1 1)] from the Si(0 0 1) plane.

The XTEM examination reveals the boundaries between these two phases, indicated by the white dashed lines in Fig. 4. They are parallel to Si{1 1 1}. The hexagonal phase extends from each Si V-groove to the surface with forming V-shaped cross section. The white arrows of Fig. 4 denote

the direction of the *c*-axis in each hexagonal-phase GaN. Thus, cubic-phase regions are spatially separated from hexagonal phases along the boundaries in the GaN grown on a nanogrooved Si surface. Region D of Fig. 4 (or Fig. 5(d)) corresponds to the starting point of the phase transition within a single V-groove. From the (0 0 2) spot position of Fig. 5(f), it can be noticed that the orientations of the principal crystal axes of the cubic-phase GaN are the same as those of the Si(0 0 1) substrate. The transition from hexagonal to cubic phase also contributes to the formation of the planar GaN surface of Figs. 2(a) and 3(a) which includes GaN(0 0 1) parallel to Si(0 0 1) at the top surface. In Fig. 5(e), because of the finite aperture size, region E of Fig. 4 (a solid circle) partly covers the adjacent cubic-phase area and Fig. 5(e) consists of the overlap of SAED patterns from cubic and hexagonal phases.

Fig. 6 presents room-temperature photoluminescence (PL) spectra obtained from the GaN epilayers grown on Si nanogrooves and on wide-area Si(1 1 1), shown in Fig. 2. In Fig. 6, two PL peaks at 3.389 and 3.200 eV are observed from the GaN on the nanogrooved Si surface, while only a single PL peak at 3.387 eV is seen from that on wide-area Si(1 1 1). The peaks at 3.389 and 3.387 eV are close to the band edge transition of hexagonal-phase GaN on Si(1 1 1) or of homoepitaxial GaN [15,16]. On the other hand, the peak at 3.200 eV is very close to the PL peak energy of cubic-phase GaN on GaAs(0 0 1) [17,18]. Thus, the GaN on wide-area Si(1 1 1) is mainly hexagonal phase but that on Si nanogrooves has both hexagonal and cubic phase, consistent with the high-resolution XTEM results of Fig. 5.

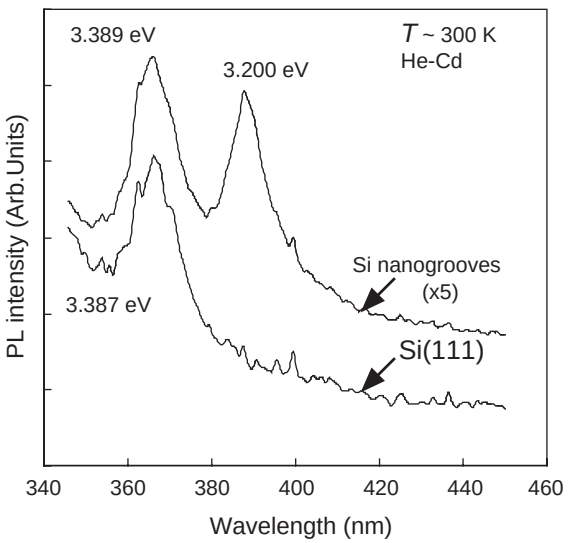


Fig. 6. Room-temperature PL spectra of the GaN grown on Si nanogrooves and on wide-area Si(1 1 1).

The Si nanogrooved surface of Fig. 1 also contains (1 1 1) orientation which allows nucleation and incorporation to proceed selectively on the (1 1 1) facets and leads to growth of hexagonal-phase GaN at the initial stages of growth, as for the wide-area Si(1 1 1) shown in Fig. 2(b). In continued deposition, the GaN on Si nanogrooves undergoes transition to cubic phase. But, the planar wide-area Si(1 1 1) substrate does not result in a transition to the cubic phase. Together with lateral overgrowth and coalescence, the phase transition leads to smooth surface morphology and spatial separation of cubic phase from hexagonal phase in the GaN on Si nanogrooves of which the original surface orientation was identically (0 0 1), like the wide-area Si(0 0 1) used for the GaN of Fig. 2(c). Consequently, the GaN grown on a nanogrooved Si(0 0 1) substrate is significantly different from that grown on Si(1 1 1) and Si(0 0 1). Once the phase transition occurs, its boundary is maintained up to 600-nm deposition. If phase transition had not occurred, the GaN on a nanogrooved Si surface would have been forced to keep periodic misaligned hexagonal phase across the groove direction (the white arrows in Fig. 4) which is energetically very unstable. Thus, the phase transition is likely to reduce the total crystal energy in epitaxial growth. Further study on the phase transition depending on groove width and pattern period is presently under way.

Finally, Fig. 7 presents room-temperature Raman scattering spectra taken from the GaN epilayers grown on Si nanogrooves and on wide-area Si(1 1 1) in backscattering geometry with an incident direction of the laser beam normal to Si(0 0 1) (or GaN(0 0 1)) of Si nanogrooves and to wide-area Si(1 1 1), respectively. As seen in Fig. 7, two peaks near 550 and 565 cm^{-1} are observed from the former while a single peak at 564 cm^{-1}

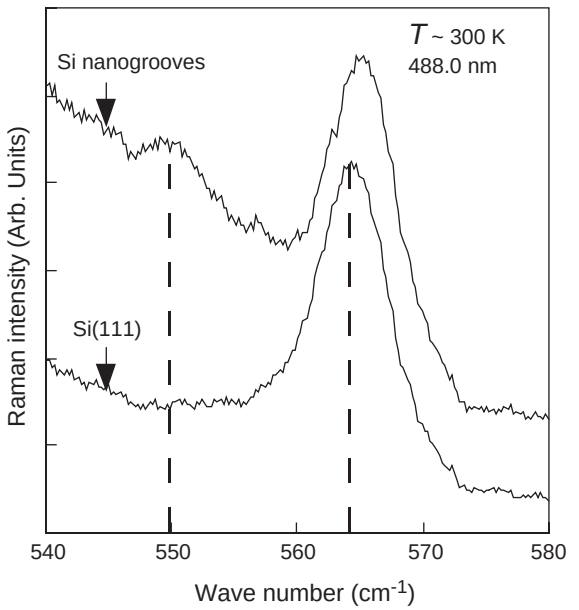


Fig. 7. Room-temperature Raman scattering spectra of the GaN grown on Si nanogrooves and on wide-area Si(1 1 1).

appears in the latter one. The peaks near 564 and 565 cm^{-1} represent the high-frequency E_2 mode of hexagonal-phase GaN and the peak near 550 cm^{-1} is close to the TO mode of cubic-phase GaN on GaAs(0 0 1) [19–22]. Thus, the absence of a TO mode in the GaN on wide-area Si(1 1 1) suggests that the dominant phase of GaN on wide-area Si(1 1 1) is hexagonal, confirming the PL results of Fig. 6. Comparing the frequency shifts of the E_2 mode with that of the corresponding phonon mode of bulk GaN ($\sim$568 cm^{-1}), the E_2 modes of the GaN on Si nanogrooves and on wide-area Si(1 1 1) have about -3 and -4 cm^{-1} shift from that of unstrained GaN, respectively [23,24]. Using a linear approximation to the frequency shift–biaxial stress relationship, these shifts imply $\sim$0.5 and $\sim$0.65 GPa tensile stress [25]. This is consistent with the slight difference of PL peak energies shown in Fig. 6 [16,26]. Thus, the GaN on Si nanogrooves exhibits lowered tensile stress than the growth on wide-area Si(1 1 1). This could be partly explained by the large number of crystal defects near the valley of the V-grooves and at the coalescence regions over the Si(0 0 1) stripe facets which can play a role of additional local relaxation centers for the GaN on the Si nanogrooved substrate.

The TO mode of GaN on GaAs(0 0 1) is at a frequency shift of $\sim$555 cm^{-1}, close to the peak observed near 550 cm^{-1} in Fig. 7 [19,20]. The TO mode from cubic phase is forbidden in the given backscattering geometry and its intensity is thus very weak. The TO mode of the cubic-phase GaN on Si nanogrooves then shows about -5 cm^{-1} frequency shift, implying higher tensile stress than that on GaAs(0 0 1). This contradicts the prediction based on lattice mismatch. However, it should be noted that the cubic-phase GaN on a Si nanogrooved substrate is not directly equivalent to that on wide-area GaAs(0 0 1). Further investigation is required to understand the stress distribution of cubic-phase GaN on Si nanogrooves.

In the GaN on Si nanogrooves of Fig. 4, cubic and hexagonal phases are periodically separated across the groove direction. Since both phases are generated in every groove and comparably extend to the surface, probably none of them would be a dominant phase in the growth direction even though deposition is continued further. For device applications, however, it is possible to obtain pure cubic-phase GaN on a limited area of Si nanogrooves. In Fig. 4, the extension of hexagonal-phase GaN to neighbor grooves after the coalescence over the Si(0 0 1) stripe facets prevents the formation of large-area cubic-phase GaN at the top surface. Such coalescence could be avoided or retarded if the Si(0 0 1) stripe width is increased. By controlling pattern period and duty cycle in grooving process, therefore, growth of pure cubic-phase GaN on a desired area of Si(0 0 1) can be realized.

4. Summary and conclusions

Phase behavior of GaN on nanofaceted Si(0 0 1) has been investigated. On 355-nm period V-grooves fabricated on a Si(0 0 1) substrate, 600-nm deposition of GaN by MOVPE results in spatial separation of cubic- and hexagonal-phase GaN. Orientation-dependent nucleation and associated lateral overgrowth and coalescence lead to

formation of cubic-phase GaN for which the direction of the principal crystal axes are identical to those of the Si(0 0 1) substrate and result in a significant improvement of the surface morphology. Conclusively, nanofaceting is highly predictable in phase control of GaN on Si(0 0 1), and can be applied to growth of cubic-phase GaN on a limited area of Si(0 0 1).

References

[1] P. Kung, D. Walker, M. Hamilton, J. Diaz, M. Razeghi, Appl. Phys. Lett. 74 (1999) 570.
[2] A. Strittmatter, S. Rodt, L. Reissmann, D. Bimberg, H. Schroeder, E. Obermerier, T. Riemann, J. Christen, A. Krost, Appl. Phys. Lett. 78 (2001) 727.
[3] Y. Kawaguchi, Y. Honda, H. Matsushima, M. Yamaguchi, K. Hiramatsu, N. Sawaki, Japan. J. Appl. Phys. 37 (1998) L966.
[4] Y. Honda, Y. Kuroiwa, M. Yamaguchi, N. Sawaki, J. Crystal Growth 242 (2002) 77.
[5] S. Zamir, B. Meyler, J. Salzman, F. Wu, Y. Golan, J. Appl. Phys. 91 (2002) 1191.
[6] T. Yodo, H. Ando, D. Nosei, Y. Harada, M. Tamura, J. Crystal Growth 237–239 (2002) 1104.
[7] T. Lei, M. Fanciulli, R.J. Molnar, T.D. Woustakas, R.J. Graham, J. Scanlon, Appl. Phys. Lett. 59 (1991) 944.
[8] M. Godelwski, E.M. Goldys, M.R. Philips, J.P. Bergman, B. Monemar, R. Langer, A. Barski, MRS Internet J. Nitride Semicond. Res. 3 (1996) 51.
[9] B.J. Min, C.T. Chan, K.H. Ho, Phys. Rev. B 45 (1992) 1159.
[10] C.-Y. Yeh, A.W. Lu, S. Froyen, A. Zunger, Phys. Rev. B 46 (1992) 10086.
[11] S.C. Lee, K.J. Malloy, S.R.J. Brueck, J. Appl. Phys. 90 (2001) 4163.
[12] D. Zubia, S.H. Zaidi, S.R.J. Brueck, S.D. Hersee, Appl. Phys. Lett. 76 (2000) 858.
[13] S.C. Lee, S.R.J. Brueck, J. Vac. Sci. Technol. B 22 (2004) 1949.
[14] S.C. Lee, X.Y. Sun, S.D. Hersee, H. Xu, S.R.J. Brueck, Appl. Phys. Lett. 84 (2004) 2079.
[15] C.Q. Chen, M.E. Gaevski, W.H. Sun, E. Kuokstis, J.P. Zhang, R.S.Q. Fareed, H.M. Wang, J.W. Yang, M.A. Khan, H.-P. Maruska, D.W. Hill, M.M.C. Chou, B. Chai, Appl. Phys. Lett. 81 (2002) 3194.
[16] D.G. Zhao, S.J. Xu, M.H. Xie, S.Y. Tong, H. Yang, Appl. Phys. Lett. 83 (2003) 677.
[17] X.L. Sun, H. Yang, L.X. Zheng, D.P. Xu, J.B. Li, Y.T. Wang, G.H. Li, Z.G. Wang, Appl. Phys. Lett. 74 (1999) 2827.
[18] Z.X. Liu, A.R. Goni, K. Syassen, H. Siegle, C. Thomsen, B. Schoettker, D.J. As, D. Schikora, J. Appl. Phys. 86 (1999) 929.
[19] A. Tabata, R. Enderlein, J.R. Leite, D. Schikora, M. Kloidt, K. Lischka, J. Appl. Phys. 79 (1996) 4137.
[20] H. Siegle, G. Kaczmarczyk, L. Filippidis, A.P. Litvinchuk, A. Hoffmann, C. Thomsen, Phys. Rev. B 55 (1997) 7000.
[21] B.H. Bairamov, O. Guerdal, A. Botchkarev, H. Morkoc, G. Irmer, J. Monecke, Phys. Rev. B 60 (1999) 16741.
[22] S. Tripathy, S.J. Chua, P. Chen, Z.L. Miao, J. Appl. Phys. 92 (2002) 3503.
[23] D.D. Manchon Jr., A.S. Barker Jr., P.J. Dean, R.B. Zatterson, Solid State Commun. 8 (1970) 1227.
[24] L.A. Falkovsky, W. Knap, J.C. Chervin, P. Wisnievski, Phy. Rev. B 57 (1998) 11349.
[25] T. Kozawa, T. Kachi, H. Kano, H. Nagase, N. Koide, K. Manabe, J. Appl. Phys. 77 (1995) 4389.
[26] W. Rieger, T. Metzger, H. Angerer, R. Dimitrov, O. Ambacher, M. Stutzmann, Appl. Phys. Lett. 68 (1996) 970.

Nanoscale Spatial Phase Modulation of GaN on a V-Grooved Si Substrate—Cubic Phase GaN on Si(001) for Monolithic Integration

S. C. Lee, B. Pattada, Stephen D. Hersee, *Fellow, IEEE*, Ying-Bing Jiang, and S. R. J. Brueck, *Fellow, IEEE*

***Abstract*—Nanoscale spatial phase modulation of GaN gorwn on a 355-nm period array of V-grooves fabricated in a Si(001) substrate is reported. Orientation-dependent selective nucleation of GaN in metal–organic vapor phase epitaxy begins from the opposing Si{111} sidewalls and rapidly fills each V-groove. At the initial stages of growth, the GaN deposited on the sidewalls has hexagonal phase with the *c*-axis normal to the Si{111}. As the growth continues, the filling of the V-groove over these misaligned hexagonal phase regions results in a transition to a cubic phase with its principal crystal axes parallel to those of the Si substrate. In a cross-sectional view perpendicular to the grooves, the defected hexagonal phase region and the clean cubic phase region above it form a boundary at the inside of each V-groove which is parallel to the Si{111} sidewalls. The GaN surface is almost planarized for only 75-nm deposition and is parallel to the original (001) plane of the Si substrate. The GaN clearly exhibits nanoscale spatial phase modulation with a periodic separation of hexagonal and cubic crystal structures across the groove direction for 600-nm deposition, implying a possibility of cubic phase GaN on an isolated single V-groove fabricated in a Si(001) substrate for monolithic integration. The structural/optical properties and stress measurements of this phase-modulated GaN grown on a nanoscale faceted Si surface are presented.**

***Index Terms*—GaN on Si, nanofaceting, phase transition.**

I. Introduction

GALLIUM nitride (GaN) and related materials have attracted much attention since the achievement of blue laser diodes on a sapphire substrate [1]. Because of their large bandgap, these materials are also attractive for high-power and high-speed electronic applications. While sapphire has a crystal symmetry that is identical to that of GaN, the large lattice mismatch (16%) and low thermal conductivity ($\sim 0.3\ \mathrm{W/cm \cdot K}$) are disadvantageous [2]. The choice of a substrate for GaN thus remains an important issue.

Lattice mismatch, thermal expansion mismatch, electrical/thermal conductivity, and surface orientation which affects the crystal structure and material properties of GaN grown on it are important factors in the selection of substrates. Growth of III-nitrides on various substrates has been reported. Sapphire is most widely used but SiC, GaN, GaAs, Si, and ZnO have been examined [3]. With the problematic exception of GaN homoepitaxy, SiC has the smallest lattice mismatch (3.5% for 6H SiC) and the highest thermal conductivity (4.9 W/cm·K), but is very expensive and of lower overall quality than traditional semiconductor substrates such as Si [2]. Recently, growth on ZrB_2 which has 0.5% lattice mismatch with thermal conductivity of 0.9–1.1 W/cm·K has been reported [4]. Currently, no substrate for GaN which is cost-effective for high-power operation is available.

Si has potential as a substrate for GaN as a result of the availability of low-cost large-area substrates (to 300-mm diameter) with extremely high quality. Si has a large lattice mismatch with GaN (−17%), but has a thermal conductivity comparable with GaN (1.3–1.5 W/cm K), considerably larger than that of sapphire [2]. However, GaN grown on Si(001) is known to have a polytype phase instability, degrading the crystallinity and resulting in an uncontrolled mixture of cubic and hexagonal phases with poor surface morphology [5]–[7]. Theoretical calculations imply that the polytypism results from the small difference of total crystal energy between the hexagonal and cubic phases [8]. Experimentally, this could be related to the stoichiometry of the front growth surface [9]. For device applications, phase control or phase separation into large-area single-phase GaN on Si(001) is required. Recently, progress in the growth of single phase GaN on Si(001) such as cubic phase GaN with a low-temperature GaN buffer or hexagonal phase GaN with an AlN buffer layer has been reported [5], [10].

In previous work, we have observed orientation-dependent nucleation and the growth of spatially separated hexagonal and cubic phase crystals of GaN on a nanoscale faceted (nanofaceted) Si surface referred to as a flat-bottom groove [11], [12]. Fig. 1 shows a schematic cross-sectional view of this structure which consists of two opposing Si{111} sidewalls isolated by Si(001) top and bottom facets in each period. Coalescence of the two hexagonal phase GaN epilayers separately grown from the opposing Si{111} sidewalls with misaligned *c*-axes results in a structural instability that is relieved by a transition to the cubic phase in the central V-groove region above the bold dashed lines of Fig. 1.

In this work, utilizing the phase transition on a Si grooved surface, we investigate nanoscale spatial phase modulation of GaN on one-dimensional (1-D) periodic V-grooves fabricated in a Si(001) substrate. Growth of GaN on a Si V-groove then may

Manuscript received August 12, 2004; revised November 26, 2004. This work was supported in part by the Defense Advanced Research Projects Agency and in part by the Department of Energy.

S. C. Lee, B. Pattada, S. D. Hersee, and S. R. J. Brueck are with the Center for High Technology Materials,University of New Mexico, Albuquerque, NM 87106 USA (e-mail: sclee@chtm.unm.edu; bkbelliappa@yahoo.com; shersee@chtm.unm.edu; brueck@chtm.unm.edu).

Y.-B. Jiang is with the Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, NM 87103 USA (e-mail: ybjiang@unm.edu).

Digital Object Identifier 10.1109/JQE.2005.843605

0018-9197/$20.00 © 2005 IEEE

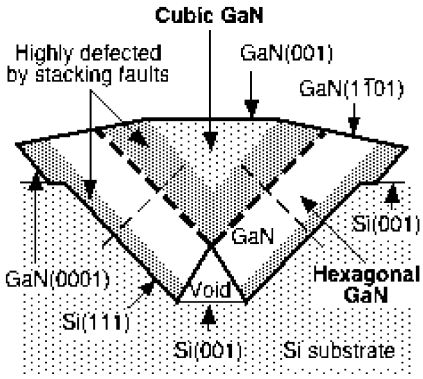


Fig. 1. Schematic illustration of the material phase map for GaN grown on a (001)-(111) nanofaceted Si surface. See [11] for details.

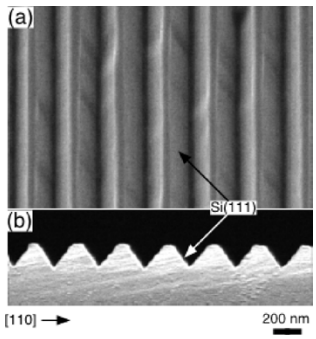


Fig. 2. (a) 45 ° tilted and (b) a cross-sectional SEM image of a 355-nm period 1-D V-grooves fabricated on a Si(001) substrate.

be able to provide more information for polytypism of GaN on Si. Additionally, the completed V-groove is easier to fabricate uniformly over large areas on a Si substrate since it is self-limiting for simple anisotropic etching processes.

Several advantages of cubic phase GaN in device applications already have been suggested. It has been reported that cubic phase GaN can offer a higher electron mobility and the possibility of higher electrical activation in p-doping than hexagonal one, both of which promise important device benefits [9], [13]. Realization of high quality cubic phase GaN is essential to prove these merits. Especially, realization of device-quality cubic phase GaN on Si(001) which is most widely used in semiconductor technologies will strongly impact overall semiconductor technology and manufacturing through monolithic integration of GaN devices with mature Si microelectronics. Also, cleaving for the mirror facets of GaN lasers will be one of the strong advantages for the usage of Si(001) substrates in optoelectronics.

In this work, we examine the phase transition mechanism with a more reliable process—V-grooving—and present in-depth analysis of optical and structural properties of cubic phase GaN grown by metal–organic vapor phase epitaxy (MOVPE) which is very important to the fabrication of micro- and optoelectronic GaN devices on Si(001). With variation of duty cycle and depth in V-grooving, as suggested in this work, spatial phase modulation will lead to the ability to realize cubic-phase GaN on a limited area of a Si(001) substrate. This is a very promising solution for low-cost high-speed high-power GaN functional devices. The GaN on nanofaceted Si(001) presented in this work is thus expected to play a significant role in the monolithic integration of GaN devices on Si(001) which is a very attractive candidate for future semiconductor technologies but currently out of interest for the low-quality epitaxy on a planar substrate.

In this paper, fabrication of the V-grooves on a Si substrate and the epitaxial growth of GaN are described in Section II. Experimental results obtained from GaN grown on this V-grooved Si surface, including orientation-dependent nucleation, crystal structure, lateral growth, formation of the cubic phase, and the optical properties of phase-modulated GaN are presented in Section III. Discussion of these results is described in Section IV with suggestions for future device applications. We summarize all the results in Section V with conclusions.

II. Nanoscale V-Grooving and MOVPE Growth

Using large-area interferometric lithography (IL) and anisotropic wet etching (AWE), a nanoscale 1-D array of V-grooves consisting of (111)-type faceted sidewalls was fabricated on a Si surface. A boron-doped Si(001) substrate with a resistivity of about 8–15 Ω-cm was used. A single longitudinal and transverse mode, frequency-tripled yttrium aluminum garnet (YAG) laser at a 355-nm wavelength was used as an ultraviolet light source in IL. A 355-nm period 1-D stripe photoresist (PR) pattern was generated on a Si(001) substrate and was coated with a 10-nm-thick Cr film. By liftoff process, the PR pattern was transferred into the Cr film, which was used as an etch mask for AWE to generate Si(111)-sidewall V-grooves. The remaining Cr was then stripped, leaving a nanofaceted (001)-(111)-$(\bar{1}\bar{1}1)$-(001) V-groove. The details of IL and AWE processing were presented elsewhere [14], [15]. Fig. 2 shows 45° tilted and cross-sectional scanning-electron microscopy (SEM) images of the patterned Si substrates. In each V-groove, the width of the top (100) facet is 60–80 nm, while the depth of the groove is 180–200 nm.

The nanofaceted substrates were passivated with hydrogen by dipping them in diluted HF immediately before loading into an MOVPE reactor. In the reactor, they were preheated to 1100 °C and cooled down to 650 °C under a hydrogen environment. On the nanofaceted Si surface shown in Fig. 2, two different thickness GaN epilayers were separately deposited using $(CH_3)_3Ga$ and NH_3 source gases. A initial ∼1.5-nm-thick GaN buffer was grown at 650 °C on both samples. This low-temperature-grown buffer provided better nucleation of GaN on the Si surface. On top of this buffer, 75-nm and ∼600-nm-thick GaN epilayers were grown on separate samples at 1050 °C. The thinner growth corresponds to about a half of the depth of the V-groove while the thicker growth is about twice as thick as the groove period. The MOVPE reactor is horizontal type. The V/III ratio for the buffer and the epilayer of both samples were $\sim 4.5 \times 10^3$ and $\sim 10^4$, respectively. Nominal growth rates were 1.5 nm/min for the buffer and 10 nm/min for the GaN epilayers.

As seen in the next section, the actual deposition amount is almost twice that estimated from the nominal growth rate multiplied by the growth time in the 75-nm layers. The growth rates were calibrated with a planar sapphire substrate since growth of GaN on planar wide-area Si(001) or Si(111) did not result in a reasonably flat surface morphology. In MOVPE, all the

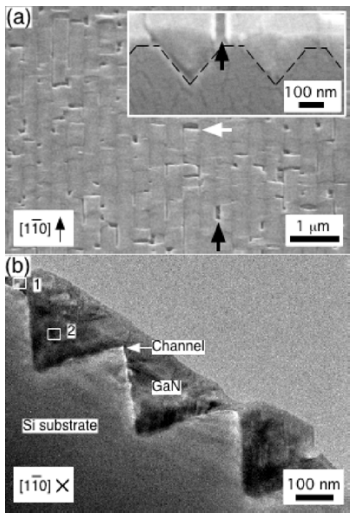


Fig. 3. (a) Top view SEM image of a 75-nm-thick GaN film deposited on the nanofaceted Si surface shown in Fig. 2. The black and a white arrows indicate surface imperfections parallel and normal to the groove direction, respectively. The inset is a magnified 45° tilted SEM image taken at the cleaved edge of the as-grown sample. (b) XTEM image of the GaN shown in (a).

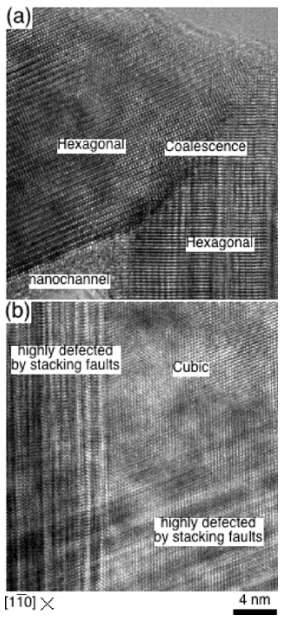


Fig. 4. High-resolution XTEM images of the white boxes labeled (a) 1 and (b) 2 of Fig. 3(b).

gases supplied to a reactor are not fully decomposed for the chemical reactions for epitaxial growth. The actual growth rate thus can be affected by surface conditions. Generally, filling of V-grooves which typically results in surface planarization in epitaxy is a fast process compared with growth on a nearby planar surface under the given growth condition. Also, gas flow dynamics on the nanofaceted Si surface may not be equivalent to that on a planar substrate. Until the V-grooves are completely filled by deposition, therefore, the actual growth rate of GaN on the nanofaceted Si surface can be different from that measured on a planar substrate.

The samples were characterized with SEM, transmission electron microscopy (TEM), atomic force microscopy (AFM), photoluminescence (PL), Raman scattering, and wide-angle X-ray diffractometry (XRD). For PL, a 325-nm wavelength He–Cd laser and a cooled GaAs photomultiplier were used. Raman spectra were taken in backscattering geometry from Si(001) with a 488-nm Ar-ion laser excitation source. X-ray diffraction (θ-2θ scan) was performed at the Cu $K\alpha_1$ wavelength.

III. Results

Fig. 3(a) shows a 45° tilted SEM image of a 75-nm-thick GaN film deposited on the nanofaceted Si surface shown in Fig. 2. The inset of Fig. 3(a) is a magnified 45° tilted SEM image taken at the cleaved edge of the as-grown sample. The dashed line in the inset denotes the GaN/Si interface. Fig. 3(b) is a cross-sectional TEM (XTEM) image of the same sample. As seen in these figures, the original Si V-grooves were filled with GaN and the V-shaped Si surface evolves to a smooth, horizontal GaN surface with only this 75-nm deposition. While the GaN surface above the valley of each groove is almost flat, that above each Si(001) stripe facet shows some fluctuation in Fig. 3(b). These uneven growths atop the Si(001) stripes correspond to small variations parallel the groove direction which are indicated by the black arrows in Fig. 3(a) and its inset. Additionally, cracks crossing the grooves, but bounded within a single groove, are also observed. One of them is highlighted by the white arrow in Fig. 3(a). In Fig. 3(b), a nanoscale triangular cross section channel (nanochannel) or a nanoscale void along the groove direction is observed atop each Si(001) stripe facet. Thus, the GaN also covers the Si(001) stripe facets but with voids atop each stripe.

Fig. 4 shows high resolution XTEM images of selected areas of Fig. 3(b). Fig. 4(a) and (b) correspond to the white boxes labeled 1 and 2 in Fig. 3(b), respectively. Fig. 4(a) shows the area directly above the narrow Si(001) stripe facet. The GaN in this region is hexagonal phase but exhibits a grain boundary along the center line of the channel, resulting from lateral overgrowth from the GaN nucleated on the two adjacent Si {111} facets with oppositely directed c-axes. On the other hand, the GaN in box 2, which is near the inside center of a Si V-groove, is cubic phase as shown in Fig. 4(b). This cubic phase GaN emerges from a large number of stacking faults as the growth proceeds, and shows a phase transition similar to that observed on the flat-bottom groove of Fig. 1 [11].

Fig. 5(a) shows an XTEM image of a 600-nm-thick GaN film grown on a nanofaceted Si surface. Contrast difference clearly reveals the GaN/Si interface. The top GaN surface is flat and smooth. As indicated by white arrows, a nanochannel is again observed on each Si(001) stripe facet.

Fig. 6(a) is an AFM image of the surface morphology of the as-grown sample shown in Fig. 5(a). The original Si grooves are fully covered with GaN by this 600-nm deposition with improved surface morphology as compared with Fig. 3. This result is also significantly flatter and smoother than other reported results on wide-area planar Si(001) surfaces [5]–[7]. Fig. 6(b) is a 45° tilted SEM image of the as-grown sample. The cracks and surface defects parallel to the groove direction which were seen in Fig. 3(a) are not observed on the 600-nm-thick GaN film.

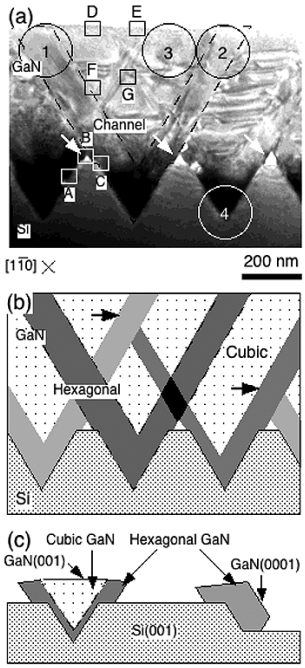


Fig. 5. (a) XTEM image of a 600-nm-thick GaN grown on a nanofaceted Si surface shown in Fig. 2. The white arrows indicate nanochannels on Si(001) stripe facets. (b) Material phase map of the cross section of the as-grown sample shown in (a). (c) Schematic illustration of cubic phase GaN grown on a single V-groove (left) and hexagonal GaN grown on a single Si(111) type facet (right). The width of the V-groove $\sim 1\ \mu\mathrm{m}$ is assumed.

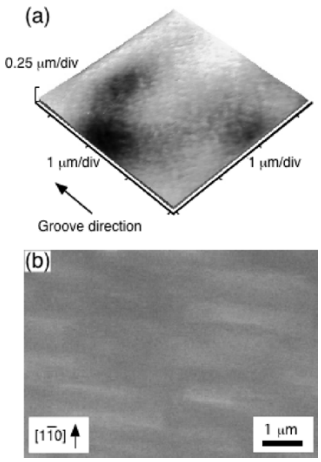


Fig. 6. (a) AFM image of the as-grown 600-nm-thick GaN surface. (b) 45 ° tilted SEM image of the same sample.

Fig. 7 shows high resolution XTEM images taken from the cross section along the GaN–Si interface in Fig. 5(a). Fig. 7(a)–(c) correspond to the white boxes of Fig. 5(a) labeled A to C. As seen in Fig. 7(a), the GaN layer exhibits an amorphous-like phase at the beginning of deposition. After a few nanometers of growth, this amorphous phase evolves to a hexagonal phase crystalline structure, but with a large number of stacking faults. Fig. 7(b) and (c) show high resolution XTEM images taken at the vertexes of a nanochannel on a Si(001) stripe facet. The nanochannel in Fig. 7(b) is large enough and has a sufficiently clear shape to identify the orientation of the sidewall facets. The angle between the sidewall of this channel which is parallel to Si{111} and the Si(001) at the bottom is

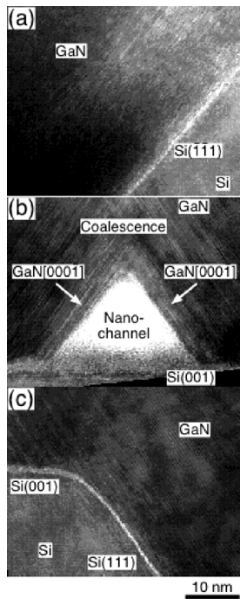


Fig. 7. High-resolution XTEM images of the boxes labeled A (a) through C (c) near the GaN/Si interface in Fig. 5(a).

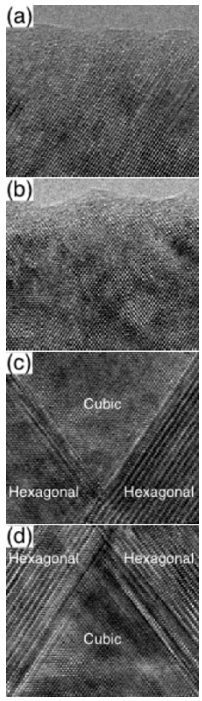


Fig. 8. High-resolution XTEM images of the areas indicated by the boxes labeled (a) D and (b) E near the GaN surface, and (c) F and (d) G near the crossing points of two hexagonal phase GaN regions in Fig. 5(a).

about 55 °. The sidewall orientations of these channels are thus approximately GaN{0001}. The same identification can be applied to the channels shown in Fig. 3(b). As in Fig. 4(a), the two c-axes of the GaN over the nanochannel are misaligned with an angle of $\sim 110\,^{\circ}$ in Fig. 7(b) and (c).

Fig. 8(a) and (b) shows high-resolution XTEM images taken near the GaN top surface regions (the boxes labeled D and E) in Fig. 5(a), respectively. Also, Fig. 8(c) and (d) is a high-resolution XTEM image taken near the crossover regions of two hexagonal phase GaN epilayers which are indicated by boxes F

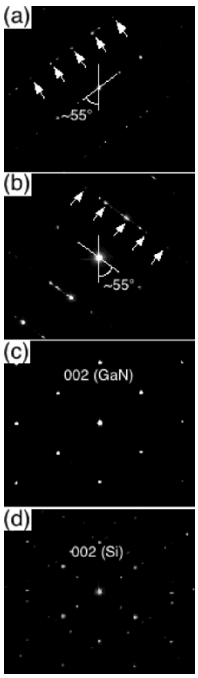


Fig. 9. SAED patterns taken from the regions indicated by the circles in Fig. 5(a). The regions labeled with numeric in Fig. 5(a) (1–4) correspond to (a)–(d). The circles enclosing each letter in Fig. 5(b) represent the aperture size for SAED.

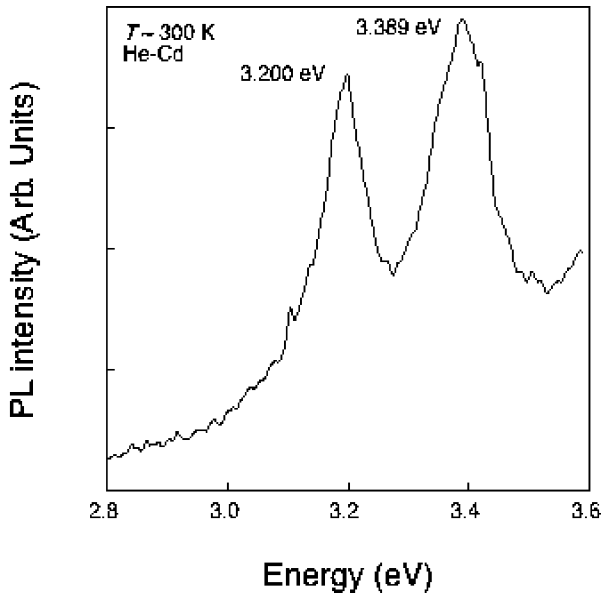


Fig. 10. Room temperature PL of the 600-nm-thick GaN sample shown in Fig. 5(a).

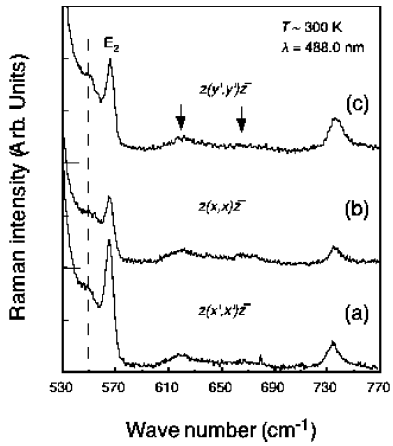


Fig. 11. Raman scattering spectra of the 600-nm-thick GaN sample shown in Fig. 5(a) with both incident and scattered polarizations parallel to (a) $[1\bar{1}0]$ (x', the groove direction), (b) [100] (x), and (c) [110] (y'). The dashed line indicates a weak, broad peak near 550 cm^{-1} in each spectrum.

and G in Fig. 5(a), respectively. The GaN near box D has hexagonal phase but that around box E has cubic phase. On the other hand, the boxes F and G have cubic phase over and under the crossing GaN hexagonal phases, respectively.

From the XTEM on the whole cross section of the 600-nm-thick GaN on the nanoscale V-grooved Si surface of Fig. 5(a), the phase map shown in Fig. 5(b) was obtained. As indicated by the dashed lines in Fig. 5, the hexagonal phase GaN begins to grow from the V-grooved Si surface and extends up to the surface of the epilayer. The cubic phase GaN appears almost at the same point in each groove. It was consistently formed above the defected hexagonal phase regions and was also generated above the crossing points of hexagonal phase GaN. Fig. 5(b) reveals that phase separations similar to those observed in Figs. 1 and 4 periodically occur across the groove direction. That is, the phase of the GaN is modulated across the groove direction with the period same as that of the V-grooves. In Fig. 5, both cubic and hexagonal phase regions have boundaries parallel to Si{111}, extending to the surface without noticeable phase intermixing.

Selected area electron diffraction (SAED) confirms the results of high resolution XTEM of Fig. 8. In Fig. 9, SAED patterns obtained from regions 1 through 4 which are indicated by solid circles in Fig. 5(a), clearly show the crystal structure variation. As seen in Fig. 9(a), region 1 is hexagonal phase. Region 2 of Fig. 9(b) is also hexagonal but its c-axis is 110 ° off from that of region 1. On the other hand, region 3 of Fig. 9(c) has a clear cubic phase. An SAED pattern of Fig. 9(d) corresponding to region 4 which is taken at the GaN–Si interface near the apex of a V-groove includes diffraction spots from the Si substrate together with those from a crossing hexagonal phase region. The circles of Fig. 5(a) correspond to the aperture size for SAED. In Fig. 9(a) and (b), because of the finite aperture size, an electron beam partly covers the adjacent cubic as well as hexagonal phase areas and results in an overlapped SAED pattern from these two different phase regions.

Fig. 10 shows a room-temperature PL spectrum obtained from the 600-nm-thick sample. The PL exhibits two peaks of comparable intensity at 3.389 and 3.200 eV, which are close to the reported PL peak energies of band-edge transitions in hexagonal and cubic phase GaN, respectively [16]–[19]. The line width of each PL peak is about 100 meV, comparable to other reports [13]–[16]

Fig. 11 shows room-temperature Raman spectra from the 600-nm-thick GaN sample taken in a backscattering geometry with the incident laser beam normal to Si(001) [or GaN(001)]. As indicated in the figure, three different polarizations were investigated: electric field parallel to $[1\bar{1}0]$ (the groove direction) [(a)], [100] [(b)], and [110] [(c)]. The polarizations of

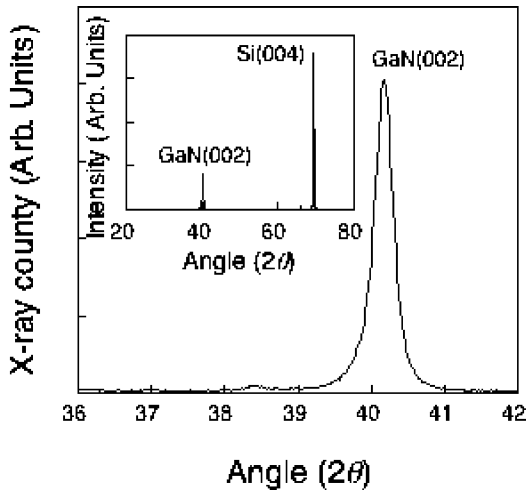


Fig. 12. XRD pattern (θ–2θ scan) near the GaN(002) direction. The inset shows a wide-angle scan of the 600-nm-thick GaN sample shown in Fig. 5(a).

both the incident laser and the scattered beam were parallel in each case. The sample was rotated with respect to the Si[001], the backscattering direction, to vary the geometry. In Fig. 11, the Raman peak intensities are not strongly dependent on polarization direction. There are three Raman peaks in all three spectra of Fig. 11 at ~ 550 cm^{-1}, 565.8 ± 0.5 cm^{-1} and 735.3 ± 1.1 cm^{-1}. The most probable phonon mode close to 550 cm^{-1} is a transverse optical (TO) phonon mode of cubic phase GaN while that close to 565.8 cm^{-1} is a high-frequency E_2 peak of hexagonal phase GaN [20]–[27]. The peak at 735.3 cm^{-1} will be discussed in the next section. The remaining peaks indicated by arrows including a peak at 520.3$\pm$0.1 cm^{-1} (not shown in Fig. 11 but providing background intensity near the E_2 peak) are from the Si substrate [25]–[27].

Fig. 12 is an XRD pattern (θ–2θ scan) from the 600-nm-thick sample showing a GaN(002) peak at $2\theta = 40.18°$ [25], [28]–[30]. The inset is a wide-angle scan spectrum which reveals the Si(004) peak together with GaN(002). The vertical lattice constant of the cubic phase GaN of Fig. 5(a) estimated from the angular position of the GaN(002) peak is 4.485 Å, which is approximately 0.015–0.017 Å smaller than the experimentally measured relaxed vertical lattice constant of cubic phase GaN on GaAs(001) [30]–[32]. No peaks from hexagonal phase GaN such as a GaN(0002) peak were observed since the (0001) is tilted from Si(001) by 54.7°.

IV. Discussion

A. Phase Transition of GaN on a Nanofaceted Si Surface

Fig. 4(b) shows the initiation point of the cubic phase GaN epilayer as it emerges from the two opposing hexagonal phases accompanied by a dense array of stacking faults. The large lattice mismatch between GaN and Si leads to the formation of these stacking faults near the GaN–Si interface. Fig. 9(d) is a SAED pattern taken at the region labeled 4 in Fig. 5(a) which partly includes the Si substrate as a reference. In Fig. 9(d), the diffraction spots from the hexagonal phase GaN are observed together with those from the Si substrate. The comparison of Fig. 9(c) with (d) reveals that the cubic phase GaN has a lattice constant smaller than the Si substrate, as expected, and that the directions of the principal crystal axes of the cubic phase GaN are identical to those of the Si substrate.

The phase map of Fig. 5(b) shows that the boundaries between hexagonal and cubic phases are roughly parallel to the Si{111}sidewalls. The phase boundary observed in Figs. 3(b) and 5(a) is similar to the results obtained from the GaN deposition on a flat-bottom groove under similar growth conditions (Fig. 1). The lower part of the cubic phase GaN region in a single V-groove of Fig. 5(b) corresponds to the area above the bold dashed lines in Fig. 1 [11]. The results of Fig. 5, therefore, suggest that nanofaceting with V-grooves on a Si(001) substrate is sufficient to cause the transition from hexagonal to cubic phase GaN.

Not only before the appearance of the cubic phase, but also at the very initial stage of growth, the crystallinity of the GaN is severely degraded by a large number of stacking faults similar to the highly defected region of Fig. 1. Near the Si V-groove surface, as seen in Fig. 7(a), the crystalline phase of GaN embedded between the cubic phase GaN region and the Si V-groove surface is not clearly identified in XTEM because of the high-density of stacking faults. The generation of a stacking fault in the hexagonal phase region corresponds to the formation of the cubic phase GaN (and *vice versa*) and can be regarded as a phase fluctuation or an aperiodic variation between the two phases. Following this growth region further from the V-groove, to the area shown in high resolution in Fig. 8(a) and SAED in Fig. 9(a) and (b), however, it is clear that this region evolves into a hexagonal crystal phase. Additionally, the SAED pattern shown in Fig. 9(d) confirms the existence of a hexagonal phase at region 4 that includes the GaN–Si interface near the apex of the V-groove. Based on these observations, the defected region near the Si V-grooves was identified as hexagonal phase in Fig. 5(b).

As indicated by the arrows, a V-shaped hexagonal phase region which extends from one V-groove is blocked by another one developed from the adjacent V-groove. This blocking mechanism needs further investigation but could be related with the uniformity of Si V-grooves which affects the filling speed and as a result the initiation point of the phase transition in individual V-grooves.

The cubic phase regions observed in Figs. 3(b) and 5(a) [or Figs. 4(b) and 8(b)–(d)] are almost free from stacking faults and have significantly lower defect density than the hexagonal regions. Also, as seen in Fig. 8(a), the number of defects in the hexagonal regions near the surface is considerably reduced compared with the region near the GaN–Si interface in the 600-nm-thick GaN film. This implies that the overall crystallinity is gradually improved as growth proceeds. Together with surface planarization discussed in Section IV C, the crystallinity improvement of GaN on a nanofaceted Si(001) is very important for device applications.

Energetically, the misaligned hexagonal phase GaNs of Fig. 5 are unstable. The orientation-dependent nucleation and crystal structure of GaN observed in Figs. 3 and 5 indicate that GaN is more stable with hexagonal rather than cubic phase at the initial stage of growth under our deposition conditions. This is supported by a theoretical model for the calculation of total crystal energy [8]. According to this model, the ground state of

GaN is hexagonal. Also, the energy difference between cubic and hexagonal phases is 9.88 meV/atom for GaN. This is relatively small compared with 18.41 meV/atom for AlN which also has a stable hexagonal phase in typical epitaxial growths. In Figs. 5(a) and (b), the initiation of the cubic phase GaN within each a Si V-groove is located at almost the same height from the vertex. This implies that there could be a certain critical deposition amount above which the total crystal energy of GaN on a nanofaceted Si surface is reduced or minimized by changing phases rather than by retaining unstable misaligned hexagonal phases. The relatively small energy difference between the two phases of GaN which could explain more probable polytypism of GaN than of AlN might be one of the physical origins driving the phase transition to lower the total crystal energy.

In the direction normal to Si(111) [or GaN(0001) or GaN(111)] in Fig. 5(a), the lattice mismatch of the (111) plane of the cubic phase GaN with respect to the (0001) plane of the hexagonal phase GaN is -0.0025, which is much smaller than the lattice mismatch of -0.17 between Si(111) and GaN(0001). Although the boundary between hexagonal and cubic phase GaN which corresponds to a high-energy plane defect is not clearly defined, the small lattice mismatch implies an elastically low-cost phase transition since it accompanies quite small variation of strain energy, and thus indirectly supports the transition from hexagonal to cubic phase with minimization of the total crystal energy.

B. Lateral Overgrowth—Formation of Nanochannels

In previous work, we reported the selectivity of GaN nucleation for Si{111} [11], [12]. The orientation selectivity of nucleation results in lateral growth of GaN from the Si{111} sidewalls inside of each V-groove toward the adjacent Si(001) facets. The arrows in Fig. 7(b) indicate the directions of lateral growth conjectured from the TEM images of Figs. 4(a), 7(b)–(c). They are perpendicular to GaN(0001) for our sample structure. Thus the sidewall orientation of the nanochannel on the top Si(001) is determined by the lateral growth along the additional $[000\bar{1}]$. Lateral overgrowth of GaN over SiO_2 patterning has been investigated by several groups [33]–[35]. In contrast to the prior reports, the lateral growth observed in this work is entirely caused by orientation-dependent nucleation, as there is no SiO_2 masking layer present during the growth.

Continued deposition with lateral growth parallel to the $[000\bar{1}]$ over the adjacent Si(001) stripe facet leads to coalescence between the two GaN epilayers separately grown on the Si{111} sidewalls of the neighboring V-grooves, resulting in the sub-100-nm-wide nanochannel on the Si(001) stripe facet between V-grooves which has been observed in Figs. 3(b), 5(a) and 7(b). Since the growth inside of the nanochannel is stopped from the moment of coalescence over a Si(001) stripe facet and the lateral size of the channel can be varied with the pattern period and the width of the Si(001) stripe facet, the lateral overgrowth on a V-grooved surface can be utilized for the fabrication of an array of nanochannels.

C. Planarization

From Fig. 3, the filling of the Si V-grooves by the GaN deposition proceeds rapidly. For the 75-nm deposition, such filling can occur since the depth of the Si V-grooves is 180–200 nm which is roughly twice the deposition thickness. As mentioned previously, the selectivity of GaN nucleation for Si{111} results in negligible deposition on the Si(001) stripe facet located between the V-grooves. Thus, rapid filling could be also partially attributed to selective nucleation on the Si{111} sidewalls in each V-groove.

Continued deposition planarizes the GaN surface, leading to the smooth and flat surface shown in Figs. 5(a) and 6. The short and shallow trenches on the GaN surface over a Si(001) stripe facet along the groove direction in Fig. 3(a) were due to incomplete coalescence for the relatively thin 75-nm deposition and were simply filled in for the 600-nm deposition shown in Fig. 6(b). The direction and spatial confinement of the cracks in Fig. 3(a) suggest that their generation may be related to stress relief. Since the GaN surface of Fig. 3 already consists of cubic and hexagonal phases for 75-nm deposition, the inner part of the cracks is in a cubic phase region and is parallel to Si[110] which is one of the cleavage directions. The confinement of the cracks to a single period of the grooves in Fig. 3(a) could be due to incomplete coalescence. In the coalesced regions, however, its propagation may be blocked by the hexagonal phase GaNs which are located at the edge sides of each groove and thus embedded between cubic phase regions in the top surface. As mentioned above, such nano-cracks are filled by continued growth.

The surface planarization observed in Fig. 5(a) will be very important for the subsequent growth of a device structure on top of the GaN grown on a nanofaceted Si surface. Unlike the GaN shown in Fig. 1, the planarized GaN surface of Fig. 5 does not include GaN$(1\bar{1}01)$. The absence of GaN$(1\bar{1}01)$ facets in the thick GaN layer is due to the coalescence of GaN over a Si(001) stripe facet. In our sample structure, GaN$(3\bar{3}04)$ is the closest orientation which is parallel to GaN(001). Since a (0001) plane of hexagonal phase GaN is parallel to a Si(111) sidewall of the V-groove, there is an angle of 54.7° between GaN(001) and GaN(0001). Then, the angle between GaN$(3\bar{3}04)$ and GaN(001) is only 0.12°. Thus, the crystal structure of the resulting GaN surface of Fig. 5(a) is modulated by cubic [(001)] and hexagonal [$(3\bar{3}04)$] phases with the period of 355 nm perpendicular to the groove direction. As seen in the PL of Fig. 9, this orientational modulation also implies a nanoscale bandgap modulation at the top surface due to the electronic differences between hexagonal and cubic phase GaN.

D. Stress Measurements

The PL spectrum of Fig. 10 supports the TEM result of the coexistence of two different phases of GaN in a single layer. Together with the Raman scattering results shown in Fig. 11 and the XRD studies of Fig. 12, PL provides information on the strain effects for the phase-modulated GaN–Si heterostructure. The PL peak energy at 3.389 eV is very close to bandgap of hexagonal phase GaN but is approximately 40 and 15 meV lower than the reported room-temperature band edge transitions of homoepitaxially grown hexagonal phase GaN and GaN grown on wide-area Si(111), respectively [16], [17]. For the comparison of our experimental results with other reported data, several factors such as substrate materials, buffer layers,

and the thickness of the GaN epilayer which affect PL peak energy by strain effect must be considered.

In the Raman spectra of Fig. 11, the position of the high-frequency E_2 peak from hexagonal phase GaN is at 565.8 cm^{-1} which shows approximately a negative frequency shift of 2 – 3 cm^{-1} from that measured in bulk GaN, including an instrumental error of ± 0.5 cm^{-1} [36], [37]. Several results for hexagonal phase GaN on Si(111) also have reported a 2–3 cm^{-1} negative shift of the E_2 peak, similar to our results [25]–[27]. Assuming a linear strain-Raman shift relationship

$$\Delta\omega = K\sigma_{||} \tag{1}$$

where $\Delta\omega$, K, and $\sigma_{||}$ are the Raman frequency shift for a given phonon mode in cm^{-1}, the linear stress coefficient ($= 6.2$ cm^{-1}/GPa) and the in-plane biaxial stress in GPa, respectively [38], the negative frequency shift of the E_2 peak implies a tensile stress of about 0.35–0.5 GPa in the hexagonal phase GaN, as expected from the difference of lattice constant between GaN and Si.

This finite biaxial tensile stress of the hexagonal phase GaN extracted from the Raman scattering suggests that the red shift of the PL peak energy at 3.389 eV may be partly due to strain effects. From the experimental linear approximation between PL peak energy of hexagonal phase GaN E_{PL} versus biaxial stress measured as

$$\frac{dE_{\mathrm{PL}}}{d\sigma_{||}} \approx 20 \text{ meV/GPa} \tag{2}$$

the red shift of the PL peak energy at 0.35–0.5 GPa is estimated as 7–10 meV [17], [39]. The difference in PL peak energy of Fig. 9 ($\sim$40 meV) from that of homoepitaxially grown GaN and its comparison with the reported PL peak energy of hexagonal phase GaN on wide-area Si(111) ($\sim$15meV) implies that the PL peak at 3.389 eV might be due to the transition of excitons bound to some defects or acceptor impurities such as carbon which is a typical impurity in epitaxial growth. In consideration of the nanofaceting fabrication process, residual contamination by carbon and other impurities with a level higher than for growth of GaN on a unprocessed substrate is not unexpected. Also, the hexagonal phase GaN near the Si V-grooves is degraded by a large number of stacking faults. Thus, the PL peak shift can be qualitatively understood as related to the presence of crystalline defects and impurities.

In case of cubic phase GaN, homoepitaxy is not available and cubic phase GaN on GaAs(001) is the closest to that observed in Fig. 5(a). In contrast to the experimental results for the higher energy PL peak, the peak at 3.200 eV in Fig. 10 is about 10–20 meV lower than the room-temperature PL peak energy of cubic phase GaN on GaAs(001) [18], [19]. In Raman scattering, cubic phase GaN on GaAs(001) exhibits peaks from TO and LO phonon modes which appear near 555 and 740 cm^{-1}, respectively [20]–[23]. The TO phonon mode of cubic phase GaN is forbidden under the given backscattering but may appear by multiple scattering in our sample. As mentioned in the previous section, the weak and broad peak around 550 cm^{-1} is very close to the TO phonon mode of cubic phase GaN on GaAs(001). This is a -5 cm^{-1} shift and implies the same shift for the longitudinal optical (LO) phonon mode at $\sim$ 735 cm^{-1} if it is assumed that both TO and LO phonon modes show the similar frequency shift under a given stress [40], [41]. However, the peak position $\sim$ 735 cm^{-1} also coincides with that of the A_1-LO phonon mode of hexagonal phase GaN on Si(111) [26], [27]. Thus, there might be an accidental overlap of the two phonon modes, one from the cubic phase and the other from the hexagonal phase GaN at $\sim$ 735 cm^{-1} in the Raman spectra of Fig. 10.

From (1), the negative frequency shift of $\sim$ 5 cm^{-1} roughly corresponds to a tensile stress of $\sim$ 1 GPa for cubic phase GaN on GaAs(001). For a more accurate comparison, the bulk elastic properties of cubic GaN must be considered. Nonetheless, it is evident that the negative frequency shift implies tensile stress of cubic-phase GaN on a nanofaceted Si surface which is greater than that on GaAs(001). The vertical lattice constant calculated from the XRD pattern of Fig. 12 and its comparison with that of cubic phase GaN on GaAs(001) also supports the presence of a tensile stress for the cubic phase GaN of Fig. 5(a) higher than that on GaAs(001).

There is a controversy about a type of stress in cubic phase GaN on GaAs(001) [30]. Thus, it should be emphasized that the stress mentioned above is a result of relative comparison. The tensile stress of $\sim$ 1 GPa can explain the red shift of about 10–20 meV in the PL peak at 3.200 eV of Fig. 10 from that of cubic phase GaN on GaAs(001) using (2) as a rough approximation. There is about a 100-nm-thick hexagonal phase GaN layer that forms a buffer layer between the Si substrate and the cubic phase GaN as seen in Fig. 5(a). Thus, the contamination and Si out-diffusion issues are less severe in the cubic phase than the hexagonal phase GaN. Also, the cubic phase GaN has a relatively low density of crystal defects in TEM compared with the hexagonal phase that is defected with stacking faults. Then, the PL peak at 3.200 eV in Fig. 10 is very close to the transition energy of donor bound excitons of cubic phase GaN.

Understanding of strain effect of a cubic phase formed on a nanofaceted Si surface is very important for its device applications. According to Figs. 11 and 12, the cubic phase GaN formed on Si nanoscale V-grooves may be in higher tensile stress than that on wide-area GaAs(001). This, however, contradicts the simplistic prediction based on lattice mismatch (e.g., GaAs has a larger mismatch with GaN than does Si). Although the cubic phase GaN of Fig. 5(a) has the same crystal orientation as the Si substrate, it is not equivalent to cubic phase GaN on wide-area Si(001) or GaAs(001) since there is a misaligned hexagonal phase GaN between them. Also, thermal stress on GaN due to the difference of thermal expansion coefficient between GaAs and Si must be considered [42]. As previously mentioned, however, all physical parameters necessary for the analysis of cubic phase GaN are not available yet. Further research is required to elucidate the overall stress effects for the phase-modulated GaN shown in Fig. 5(a).

E. Application to Monolithic Integration on Si(001)

Fig. 5 is, to our knowledge, the first observation of periodic spatial separation of cubic phase from hexagonal phase GaN ever grown on Si(001) and GaAs(001). The phase map of Fig. 5(b) suggest that the phase modulation or separation of

GaN on a Si(001) substrate can be controlled by the facet geometry on the substrate surface.

Growth of a large-area single phase region of cubic phase GaN on Si(001) would have important device possibilities. While the two different phases are periodically separated at the nanoscale range, as seen in Fig. 5, this may not be useful in practical applications which require a single phase GaN over a large area since both phases are still mixed together on micron and larger scales. Variation of the duty cycle and pattern period of V-grooves provides a technique for controlling the phase distribution. It has been confirmed that the cubic phase is formed in each groove only for 75-nm deposition where coalescence does not completely proceed. By increasing the pattern period to micron range while keeping the width of a V-groove in nanometer range or by fabricating a single V-groove on a desired area, coalescence over the Si(001) stripe facets can be delayed or avoided while phase transition is activated in the V-groove at the initial stage of growth. Then, as schematically shown in Fig. 5(c) which is based on the results of Fig. 3, a GaN stripe-type epilayer grown along a single V-groove dominantly can have a cubic phase top surface with a micron-order width that is enough for the fabrication of devices.

As seen in Fig. 5(c), a single hexagonal phase GaN epilayer is also available on a limited area of Si(001) but is not suitable for device application because its top surface is not parallel to Si(001). For compatibility with conventional device technology, crystal growth techniques must provide planar surface epilayers. A planar top surface of the cubic phase GaN parallel to Si(001) thus highly favorable for device fabrication on nanofaceted Si surface. It has been reported that cubic phase AlGaN (or AlN) and InGaN are also available on (001) type orientation [43]–[45]. On the planar surface of cubic phase GaN on a single V-groove shown in Fig. 5(c), therefore, it is possible to grow various heterostructures composed of GaN and its related compounds in cubic phase, for monolithic integration of micro- and optoelectronic GaN-based devices on Si(001).

In this work, we have investigated spatial phase modulation of GaN on a nanofaceted Si(100) substrate with low-cost but highly reliable conventional process techniques such as large-area i-line IL, AWE, and MOVPE. Further research is required to understand the relation between facet geometry/physical dimension and the relative size of the cubic phase regions and associated strain effects of GaN on a nanofaceted Si surface. Also, the optical/electronic properties of spatially phase-modulated GaN need more investigation for future device applications.

V. Summary and Conclusions

Spatial phase modulation of GaN grown on a nanofaceted Si surface has been investigated. On a 355-nm period, 1-D array of V-grooves fabricated in a Si(001) substrate, selectivity of the nucleation and incorporation of hexagonal phase GaN on Si(111) has been observed in MOVPE growth. Orientation-dependent nucleation results in lateral growth from Si{111} facets toward neighboring top Si(001) facets with the formation of nanochannels between them, and eventually coalescence. For the 75-nm deposition, the entire GaN film is almost planarized to a smooth surface but with a complex, spatially separated phase structure resulting from the periodically misaligned c-axes across the groove direction. This structural instability induces a phase transition from hexagonal to cubic phase on the nanofaceted Si surface, which results in spatial phase separation of GaN. Based on these results, highly predictable phase modulation of GaN has been demonstrated on a V-grooved Si surface with the 600-nm deposition. PL, Raman scattering, and XRD show that both the hexagonal and cubic phase GaN are in tensile stress. Our experimental results therefore suggest that nanofaceting can be used for the formation of cubic phase GaN on limited areas of a Si (001) substrate through phase transition for monolithic integration of GaN devices on Si(001).

References

[1] S. Nakamura, M. Senoh, S. Nagahama, N. Iwasa, T. Yamada, T. Matsushita, H. Kiyoku, and Y. Sugimoto, "InGaN multi-quantum-well-structure laser diodes with cleaved mirror cavity facets," *Jpn J. Appl. Phys.*, vol. 35, pp. L217–L220, Feb. 1996.

[2] A. R. Powell and L. B. Rowland, "SiC materials – Progress, status, and potential loadblocks," *Proc. IEEE*, vol. 90, no. 6, pp. 942–955, Jun. 2002.

[3] T. Matsuoka, "Tenary alloys," in *GaN and Related Materials*, S. J. Pearton, Ed. Amsterdam, The Netherlands: Gordon and Breach Science Publishers, 1997, pp. 53–84.

[4] H. Kinoshita, S. Otani, S. Kamiyama, H. Amano, I. Akasaki, J. Suda, and H. Matsunami, "Zirconium diboride (0001) as an electrically conductive lattice-matched substrate for gallium nitride," *Jpn J. Appl. Phys.*, vol. 40, pp. L1280–L1282, Dec. 2001.

[5] T. Lei, M. Fanciulli, R. J. Molnar, T. D. Woustakas, R. J. Graham, and J. Scanlon, "Epitaxial growth of zincblende and wurtizte GaN thin films on Si(001)," *Appl. Phys. Lett.*, vol. 59, pp. 944–946, Aug. 1991.

[6] T. Yodo, H. Ando, D. Nosei, Y. Harada, and M. Tamura, "Investigation of initial growth process for GaN heteroepitaxial layers grown on Si(0 0 1) and Si(1 1 1) substrates by ECR-assisted MBE," *J. Cryst. Growth*, vol. 237–239, pp. 1104–1109, Apr. 2002.

[7] M. Godelwski, E. M. Goldys, M. R. Philips, J. P. Bergman, B. Monemar, R. Langer, and A. Barski, "Morphology and optical properties of cubic phase GaN epilayers grown on (001) Si," *MRS Internet J. Nitride Semicond. Res.*, vol. 3, 1996.

[8] C.-Y. Yeh, Z. W. Lu, S. Froyen, and A. Zunger, "Zinc-blende-wurtzite polytypism in semiconductors," *Phys. Rev. B.*, vol. 46, pp. 10 086–10 097, Oct. 1992.

[9] J. W. Orton and C. T. Foxon, "Group III nitride semiconductors for short wavelength light-emitting devices," *Rep. Prog. Phys.*, vol. 61, pp. 1–75, Jan. 1998.

[10] J. Wan, R. Venugopal, M. R. Melloch, H. M. Liaw, and W. J. Rummel, "Growth of crack-free GaN thin films on Si(100)," *Appl. Phys. Lett.*, vol. 79, pp. 1459–1461, Sep. 2001.

[11] S. C. Lee, X. Y. Sun, S. D. Hersee, H. Xu, and S. R. J. Brueck, "Spatial phase separation of GaN selectively grown on a nanoscale faceted Si surface," *Appl. Phys. Lett.*, vol. 84, pp. 2079–2081, Mar. 2004.

[12] S. C. Lee, X. Y. Sun, S. D. Hersee, J. Lee, Y.-B. Jiang, H. Xu, and S. R. J. Brueck, "Growth of GaN on a nanoscale periodic faceted Si substrate by metal organic vapor phase epitaxy," in *Proc. 2003 .Int. Symp. Compound Semicond.*, pp. 15–21.

[13] M. A. Khan, Q. Chen, J. Yang, Z. Anwar, M. Blasingame, and M. S. Shur, "Recent advances in III-V nitride electron devices," in *Proc. Tech. Dig. Int. Electron Device Meeting*, vol. 96, Dec. 1996, p. 27.

[14] S. R. J. Brueck, S. H. Zaidi, X. Chen, and Z. Zhang, "Interferometric lithography – From periodic arrays to arbitrary patterns," *Microelec. Engineering*, vol. 41/42, pp. 145–148, Mar. 1998.

[15] S. C. Lee and S. R. J. Brueck, "Nanoscale two-dimensional patterning on Si(001) by large-area interferometric lithography and anisotropic wet etching," *J. Vac. Sci. Technol. B*, vol. 22, pp. 1949–1952, Jul.–Aug. 2004.

[16] C. Q. Chen, M. E. Gaevski, W. H. Sun, E. Kuokstis, J. P. Zhang, R. S. Q. Fareed, H. M. Wang, J. W. Yang, M. A. Khan, H.-P. Maruska, D. W. Hill, M. M. C. Chou, and B. Chai, "GaN homoepitaxy on freestanding (1100) oriented GaN substrates," *Appl. Phys. Lett.*, vol. 81, pp. 3194–3196, Oct. 2002.

[17] D. G. Zhao, S. J. Xu, M. H. Xie, S. Y. Tong, and H. Yang, "Stress and its effect on optical properties of GaN epilayers grown on Si(111), 6H-SiC(0001), and c-plane sapphire," *Appl. Phys. Lett.*, vol. 83, pp. 677–679, Jul. 2003.
[18] X. L. Sun, H. Yang, L. X. Zheng, D. P. Xu, J. B. Li, Y. T. Wang, G. H. Li, and Z. G. Wang, "Stability investigation of cubic GaN films grown by metalorganic chemical vapor deposition on GaAs (001)," *Appl. Phys. Lett.*, vol. 74, pp. 2827–2829, May 1999.
[19] Z. X. Liu, A. R. Goni, K. Syassen, H. Siegle, C. Thomsen, B. Schoettker, D. J. As, and D. Schikora, "Pressure and temperature effects on optical transitions in cubic GaN," *J. Appl. Phys.*, vol. 86, pp. 929–934, Jul. 1999.
[20] A. Tabata, R. Enderlein, J. R. Leite, D. Schikora, M. Kloidt, and K. Lischka, "Comparative Raman studies of cubic and hexagonal GaN epitaxial layers," *J. Appl. Phys.*, vol. 79, pp. 4137–4140, Apr. 1996.
[21] H. Siegle, G. Kaczmarczyk, L. Filippidis, A. P. Litvinchuk, A. Hoffmann, and C. Thomsen, "Zone-boundary phonons in hexagonal and cubic GaN," *Phys. Rev. B*, vol. 55, pp. 7000–7004, Mar. 1997.
[22] A. Tabata, A. P. Lima, J. R. Leite, V. Lemos, D. Schikora, B. Schoettker, U. Koehler, D. J. As, and K. Lischka, "Micro-Raman analysis of cubic GaN layers grown by MBE on (001) GaAs substrate," *Semicond. Sci. Technol.*, vol. 14, pp. 318–322, Apr. 1999.
[23] H. F. Liu, H. Chen, Z. Q. Li, L. Wan, Q. Huang, J. M. Zhou, N. Yang, K. Tao, Y. J. Han, and Y. Luo, "MBE growth and Raman studies of cubic and hexagonal GaN films on (0 0 1)-oriented GaAs substrates," *J. Cryst. Growth*, vol. 218, pp. 191–196, Sep. 2000.
[24] M. S. Liu, S. Prawer, L. A. Bursill, D. J. As, and R. Brenn, "Characterization of the surface irregularities of cubic GaN using micro-Raman spectroscopy," *Appl. Phys. Lett.*, vol. 78, pp. 2658–2860, Apr. 2001.
[25] B. H. Bairamov, O. Guerdal, A. Botchkarev, H. Morkoc, G. Irmer, and J. Monecke, "Direct evidence of tensile strain in wurtzite structure n-GaN layers grown on n-Si(111) using AlN buffer layers," *Phys. Rev. B*, vol. 60, pp. 16 741–16 746, Dec. 1999.
[26] M. Seon, T. Prokofyeva, M. Holtz, S. A. Nikishin, N. N. Faleev, and H. Temkin, "Selective growth of high quality GaN on Si(111) substrates," *Appl. Phys. Lett.*, vol. 76, pp. 1842–1844, Apr. 2000.
[27] S. Tripathy, S. J. Chua, P. Chen, and Z. L. Miao, "Micro-Raman investigation of strain in GaN and $Al_xGa_{1-x}N$/GaN heterostructures grown on Si(111)," *J. Appl. Phys.*, vol. 92, pp. 3503–3510, Oct. 2002.
[28] H. Okumura, S. Misawa, and S. Yoshida, "Epitaxial growth of cubic and hexagonal GaN on GaAs by gas-source molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 59, pp. 1058–1060, Aug. 1991.
[29] T. S. Cheng, L. C. Jenkins, S. E. Hooper, C. T. Foxon, J. W. Orton, and D. E. Lacklison, "Selective growth of zinc-blende, wurtzite, or a mixed phase of gallium nitride by molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 66, pp. 1509–1511, Mar. 1995.
[30] D. P. Xu, Y. T. Wang, H. Yang, S. F. Li, D. G. Zhao, Y. Fu, S. M. Zhang, R. H. Wu, Q. J. Jia, W. L. Zheng, and X. M. Jiang, "Anomalous strains in the cubic-phase GaN films grown on GaAs (001) by metalorganic chemical vapor deposition," *J. Appl. Phys.*, vol. 88, pp. 3762–3764, Sep. 2000.
[31] S. Strite, J. Ruan, Z. Li, A. Salvador, H. Chen, D. J. Smith, W. J. Choyke, and H. Morkoc, "An investigation of the properties of cubic GaN grown on GaAs by plasma-assisted molecular beam epitaxy," *J. Vac. Sci. Technol. B*, vol. 9, pp. 1924–1929, Jul.–Aug. 1991.
[32] K. Kim, W. Lamberecht, and B. Segall, "Electronic structure of GaN with strain and phonon distortions," *Phys. Rev. B*, vol. 50, pp. 1502–1505, Jul. 1994.
[33] O. H. Nam, M. D. Bremer, T. S. Zheleva, and R. F. Davis, "Lateral epitaxy of low defect density GaN layers via organometallic vapor phase epitaxy," *Appl. Phys. Lett.*, vol. 71, pp. 2638–2640, Nov. 1997.
[34] S. Nakamura, "InGaN/GaN/AlGaN-based laser diodes with an estimated lifetime of longer than 10 000," *MRS Bullet.*, vol. 23, p. 27, May 1998.
[35] P. Kung, D. Walker, M. Hamilton, J. Diaz, and M. Razeghi, "Lateral epitaxial overgrowth of GaN films on sapphire and silicon substrates," *Appl. Phys. Lett.*, vol. 74, pp. 570–572, Jan. 1999.
[36] D. D. Manchon Jr., A. S. Barker Jr., P. J. Dean, and R. B. Zatterson, "Optical studies of the phonons and electrons in gallium nitride," *Solid State Commun.*, vol. 8, pp. 1227–1231, Aug. 1970.
[37] L. A. Falkovsky, W. Knap, J. C. Chervin, and P. Wisnievski, "Phonon modes and metal-insulator transition in GaN crystals under pressure," *Phy. Rev. B*, vol. 57, pp. 11 349–11 355, May 1998.
[38] T. Kozawa, T. Kachi, H. Kano, H. Nagase, N. Koide, and K. Manabe, "Thermal stress in GaN epitaxial layers grown on sapphire substrates," *J. Appl. Phys.*, vol. 77, pp. 4389–4392, May 1995.
[39] W. Rieger, T. Metzger, H. Angerer, R. Dimitrov, O. Ambacher, and M. Stutzmann, "Influence of substrate-induce biaxial stress on the optical properties of thin film GaN," *Appl. Phys. Lett.*, vol. 68, pp. 970–972, Feb. 1996.
[40] G. Landa, R. Carles, C. Fontaine, E. Bedel, and A. Munoz-Yaguee, "Optical determination of strains in heterostructures: GaAs on Si as an example," *J. Appl. Phys.*, vol. 66, pp. 196–200, Jul. 1989.
[41] B. Roughani, M. Kallergi, J. Aubel, and S. Sundaram, "Raman scattering by epitaxial GaAs on a Si substrate," *J. Appl. Phys.*, vol. 66, pp. 4946–4949, Nov. 1989.
[42] E. H. Lingunis, N. M. Haegel, and N. H. Karam, "Thermal stress distributions in GaAs on sawtooth-patterned Si substrates: A finite element approach," *Appl. Phys. Lett.*, vol. 61, pp. 2202–2204, Nov. 1992.
[43] A. Nakadaira and H. Tanaka, "Metalorganic vapor-phase epitaxy of cubic $Al_xGa_{1-x}N$ alloy on a GaAs(100) substrate," *Appl. Phys. Lett.*, vol. 70, pp. 2720–2722, May 1997.
[44] W.-T. Lin, L.-C. Meng, G.-J. Chen, and H.-S. Liu, "Epitaxial growth of cubic AlN films on (100) and (111) silicon by pulsed laser ablation," *Appl. Phys. Lett.*, vol. 66, pp. 2066–2068, Apr. 1995.
[45] T. Kitamura, S. H. Cho, Y. Ishida, T. Ide, X. Q. Shen, H. Nakanishi, S. Chichibu, and H. Okumura, "Growth and characterization of cubic InGaN epilayers on 3C-SiC by RF MBE," *J. Cryst. Growth*, vol. 227–228, pp. 471–475, 2001.

S. C. Lee is currently with Center for High Technology Materials at Albuquerque, NM.

His research interests include patterned growth of nanostructures and semiconductor lasers.

B. Pattada was born in India in 1978. He received the B.E. degree in electronics and communication from R.V. College of Engineering, Bangalore, India, in 2001. He is currently pursuing the M.S. degree in electrical engineering at the University of New Mexico, Albuquerque.

Stephen D. Hersee (SM'96–F'03) was born in Hastings, U.K., in 1950. He received the B.Sc. degree (first-class hons.) in 1968 and the Ph. D. degree in 1975, both from Brighton Polytechnic, Brighton, U.K.

He was with the Plessey Co. Ltd., U.K., from 1975 to 1980, Thomson CSF, France, from 1980 to 1986, and General Electric, Syracuse, NY, from 1986 to 1991. He joined the University of New Mexico, Albuquerque, in 1991, where he is currently a Professor in the Electrical and Computer Engineering Department. He teaches courses in semiconductor materials, advanced heterojunction devices, and semiconductor process technology. He is also a member of the Center for High Technology Materials. His research interests include novel semiconductor materials and heteroepitaxy for advanced materials and devices.

Ying-Bing Jiang received the M.S. degree in materials science and engineering from the Northeast Institute of Heavy-Performance Machines, China, in 1995.

He then worked on thin film deposition and materials surface modification by high energy and density plasma at the Chinese Academy of Sciences. He is currently a Research Scientist in the Transmission Electron Microscopy (TEM) Laboratory, University of New Mexico, Albuquerque. His research interests include thin film deposition and materials characterization with analytical TEM.

S. R. J. Brueck (S'63–M'71–SM'98–F'93) was born in New York City in 1944. He received the B.S. degree from Columbia University, New York, in 1965 and the M.S. and Ph.D. degrees from Massachusetts Institute of Technology (MIT), Cambridge, in 1967 and 1971, respectively, all in electrical engineering.

From 1971 to 1985, he was Member of the Technical Staff at MIT's Lincoln Laboratory. In 1985, he moved to the University of New Mexico at Albuquerque, where he is currently a Professor in the Electrical and Computer Engineering Department and the Physics and Astronomy Department and Director of the Center for High Technology Materials. His current research interests include nanoscale lithography, the physics of nanostructures, the nonlinear optics of poled glasses, and semiconductor laser physics.

Radiation from a Dipole Embedded in a Dielectric Slab

S. R. J. Brueck, *Fellow, IEEE*

Invited Paper

***Abstract*—A unified analytical treatment of the radiation from an electric dipole of arbitrary orientation embedded at an arbitrary location within a symmetrically clad dielectric slab is presented. Both the emission into three-dimensional (3-D) radiation modes, corresponding to emission within the critical angle escape cone within the dielectric slab, and into the two-dimensional (2-D) waveguide modes are evaluated from a single calculation. The model is valid for arbitrary dielectric contrast between the slab and the cladding. The mathematical approach uses well-known complex analysis techniques: the 3-D radiation is described by a steepest descents integration around branch cuts while the 2-D waveguide modes correspond to simple poles. The division of the radiated power between the 3-D and 2-D modes is evaluated across the entire range from small dielectric contrast appropriate to diode lasers ($\lesssim$1.1) to the very large dielectric contrast of free-standing semiconductor slabs ($\sim$12–19). Both enhancement and suppression, depending on position, slab width, dielectric contrast, and wavelength, of the total radiated power in comparison with that in an unbounded dielectric medium are found for slab widths on the order of a wavelength with a maximum enhancement of $\sim$30% for these one-dimensional Fabry–Perot structures. For thicker slabs, the total radiation is almost constant and equal to that in the unbounded medium for low dielectric contrast while still exhibiting some modulation as increasing thickness allows additional waveguide modes.**

***Index Terms*—Dielectric waveguides, dipole radiation, microcavity.**

I. Introduction

THE DISTRIBUTION of radiation from a dipole near a planar interface is a classic antenna problem [1]–[3]. As is very well known [4], [5], the addition of a second dielectric interface, e.g., a dipole embedded in a dielectric slab, adds significant complexity to the problem, largely because the radiation is now distributed between three-dimensional (3-D) radiation modes as in the single interface case and two-dimensional (2-D) waveguide modes supported by the dielectric slab.

Despite this long history, the problem is increasingly important today with developments such as sophisticated diode laser structures—both edge emitters and vertical microcavity lasers [6], enhanced LEDs [7], and 2-D photonic bandgap lasers [8]. Two general classes of treatment have appeared in the literature. The first class [9]–[12] provides an analytic treatment while the second provides a complete numerical solution of Maxwell's equations in an arbitrary geometry using finite element or other numerical techniques [13], [14]. The analytic results have relied largely on a somewhat *ad hoc* combination of results from 3-D radiation and 2-D waveguide mode calculations. Ho *et al.* in a very complete chapter [15] obtained results very similar to those presented here; they do not present closed form analytic results for all of the radiation components nor allow the dipole to be arbitrarily located within the dielectric slab. The numerical calculations, while applicable to arbitrary structures, are difficult to generalize to structures not presented in detail without recoding the problem and do not serve to build up physical intuition as quickly as is possible with analytic results.

The purpose of this paper is to present simple, closed-form, analytic expressions for the emission into both 3-D continuum radiation and 2-D waveguide modes for an electric dipole of arbitrary location and orientation within a symmetrically clad dielectric slab. These expressions are derived from a common calculation based on mathematical apparatus extended from the seminal work of Sommerfeld [1]. In this formulation, the radiation from a dipole is expanded into plane waves and the angularly-dependent Fresnel boundary conditions for reflection/refraction at a dielectric interface are applied to each component plane wave. The various emission modes appear as singularities in the inverse transform in the complex plane: the 3-D modes as branch cuts that are evaluated by a steepest descents method; the 2-D modes as poles that are evaluated by the calculus of residues. The advantage of these analytic solutions is that they are readily evaluated for arbitrary dielectric constants ranging from the very large dielectric discontinuities of microcavity lasers to the much more modest discontinuities of, e.g., double-heterostructure edge-emitting lasers. The model is straightforwardly extensible to more complex layered structures (e.g., asymmetric slabs, vertical-cavity lasers, periodic grating structures), although this extension is not pursued in this initial report. Examples of the results for a variety of dielectric slabs are discussed with particular attention to the distribution of radiated energy between the 3-D and 2-D modes and to the comparison of the total radiated power with that of a dipole in an unbounded dielectric material.

For a dipole in a dielectric slab, both macroscopic electromagnetic effects arising from the modified density of states of

Manuscript received October 9, 2000. This work was supported in part by the Multidisciplinary University Research Program of the Department of Defense and by the Air Force Office of Scientific Research.

The author is with the Center for High Technology Materials and the Departments of Electrical and Computer Engineering and Physics and Astronomy, University of New Mexico, Albuquerque, NM 87106 USA.

Publisher Item Identifier S 1077-260X(00)11607-7.

1077–260X/00$10.00 © 2000 IEEE

the electromagnetic field and microscopic Clausius–Mossotti local-field effects need to be considered. The former are the subject of this paper, the latter are consistently normalized out of the results by comparing the radiation patterns and intensities for the slab to those in an unbounded medium of the same dielectric constant.

As is well known, for very large dielectric discontinuities [for example, a Si ($\varepsilon_1/\varepsilon_0 \equiv \kappa_1 \sim 18$) or GaAs ($\kappa_1 \sim 12$) slab surrounded by air ($\kappa_2 \sim 1$)] and extremely thin slabs ($\sim\lambda/\kappa_1^{1/2}$) where λ is the free-space wavelength and ε_1 the slab dielectric constant) the total radiated energy can be either larger or smaller than that in bulk depending on the precise relationship of slab width L, λ, and ε_1 and the location of the dipole within the slab. When considered in a cavity quantum electrodynamic context, this corresponds to cavity enhancement or suppression of the spontaneous radiative lifetime [5], [9]–[13]. We show that the maximum enhancement for a simple slab geometry saturates as a function of dielectric contrast at about 15% for dipoles oriented perpendicular and 30% for dipoles oriented parallel to the slab with the optimal transverse location being the center of the slab for slab widths of order $\lambda/\kappa_1^{1/2}$. Our motivation for attacking this problem was unexplained, strongly enhanced optical–phonon Raman scattering results for nanoscale Si gratings [16]; therefore, many of the high dielectric constant numerical examples are given for parameters appropriate to these visible experiments. The application of this theory to the Raman experiments will be discussed in a subsequent publication.

II. Hertz Vector Formalism

Following Sommerfeld [1], we formulate the problem in terms of the Hertz vector ($\vec{\Pi}$) from which the electric and magnetic fields follow:

$$\vec{E} = \nabla(\nabla \cdot \vec{\Pi}) + k_0^2 \kappa_i \vec{\Pi} \tag{1}$$

$$\vec{H} = -i\frac{k_0 \kappa_i}{\eta_0} \nabla \times \vec{\Pi} = \frac{-i}{\eta_0 k_0} \nabla \times \vec{E} \tag{2}$$

where

$$k_0^2 = \omega^2 \mu_0 \varepsilon_0$$
$$\kappa_i \equiv n_i^2 = \varepsilon_i/\varepsilon_0$$

and

$$\eta_0 = \sqrt{\mu_0/\varepsilon_0}.$$

The impulse response for the Hertz vector in an unbounded medium for a dipole located at the origin and oriented along the z axis is given by

$$\overrightarrow{\Pi} = \phi_z \hat{e}_z = \frac{p}{4\pi\varepsilon_0 \kappa_i} \frac{e^{ik_0 n_i R}}{R} \hat{e}_z \tag{3}$$

where p is the dipole moment. Applying (1) and (2) gives the familiar dipole radiation fields

$$E_\theta = \frac{k_0^2 p}{4\pi\varepsilon_0} \sin\theta \frac{e^{ik_0 n_i R}}{R} = \frac{k_0^3 p}{4\pi\varepsilon_0} \sin\theta \frac{e^{in_i \tilde{R}}}{\tilde{R}}$$

$$H_\varphi = \frac{k_0^2 n_i p}{4\pi\eta_0\varepsilon_0} \sin\theta \frac{e^{ik_0 n_i R}}{R} = \frac{k_0^3 n_i p}{4\pi\eta_0\varepsilon_0} \sin\theta \frac{e^{in_i \tilde{R}}}{\tilde{R}}$$

$$S_R = \frac{k_0^4 n_i |p|^2}{32\pi^2 \eta_0 \varepsilon_0^2 R^2} \sin^2\theta$$

$$P = \int S_R R^2 \, d\Omega = \frac{k_0^4 \mathrm{Re}(n_i)|p|^2}{12\pi\eta_0\varepsilon_0^2} \tag{4}$$

where

S_R	radial Poynting vector;
P	total radiated power from the dipole;
$\tilde{R} = k_0 R$	dimensionless distance scaled to the free-space wavelength.

To contain the complexity of the equations, only the radiation fields ($\propto 1/R$) are presented. All near-field effects are, nonetheless, included in the analysis.

It is useful to investigate the 3-D Fourier transform of the impulse response, given by

$$\phi(\vec{R}) = \frac{k_0 e^{in_i \tilde{R}}}{\kappa_i \tilde{R}}$$

$$\Phi(\vec{K}) = \int \phi(\vec{R}) e^{i\vec{K}\cdot\vec{R}} d^3R = \frac{4\pi}{\kappa_i k_0^2 (K^2 - \kappa_i)} \tag{5}$$

where $\vec{K}$ is the transform coordinate and $K = |\vec{K}|$. This expression can be put into a more suitable form for evaluating the boundary conditions by carrying out the inverse transform in the z direction (perpendicular to the plane surfaces) by the calculus of residues and expressing the remaining integral of the transform in cylindrical coordinates [2]

$$\phi(\vec{R}) = \frac{1}{\kappa_i} \int_0^\infty \frac{k_0}{\gamma_i} e^{-\gamma_i |\tilde{z}|} J_0(\lambda\tilde{\rho}) \lambda \, d\lambda$$
$$= \frac{1}{2\kappa_i} \int_{-\infty}^\infty \frac{k_0}{\gamma_i} e^{-\gamma_i |\tilde{z}|} H_0^1(\lambda\tilde{\rho}) \lambda \, d\lambda \tag{6}$$

where λ is the radial transform coordinate, $\tilde{\rho} = k_0\rho$ and $\tilde{z} = k_0 z$ are the dimensionless radial and vertical spatial coordinates, and

$$\gamma_i = \sqrt{\lambda^2 - \kappa_i} \underset{\lambda \to 0}{\longrightarrow} -in_i \tag{7}$$

and the sign of the square root is chosen to ensure $\mathrm{Re}(\gamma_i) > 0$ for convergence of (6). J_0 is the zeroth-order Bessel function and H_0^1 is the Hankel function of the first kind of zeroth order. The second form of this expression will be useful for evaluation of the complex integrations since the integration extends across the entire λ axis.

III. Vertical Dipole

Initially, the analysis will be restricted to a vertical dipole because the mathematics is slightly simpler since a one-component Hertz vector suffices whereas a two-component Hertz vector is required for a horizontal dipole. The results for a horizontal dipole will be presented below. An arbitrary orientation can always be reduced to a superposition of the appropriate vertical and horizontal components. The geometry for a vertical dipole (relative to the horizontal slab) is shown in Fig. 1. A dipole is located at a vertical distance a above the origin within an infinitely long slab of width L and dielectric constant ε_1.

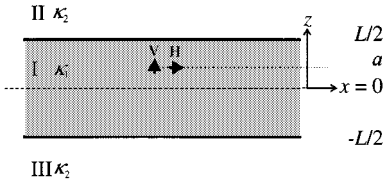


Fig. 1. Geometry of the problem: a dipole oriented either vertically or horizontally is located at a height a from the center of a slab of width L and relative dielectric constant κ_1 clad on both sides with semi-infinite materials of relative dielectric constant κ_2.

The slab is adjacent to two semi-infinite half planes of dielectric constant ε_2. In this situation, the Hertz vector has only a z component [2], [17]

$$\vec{\Pi}^i = \phi_z^i \hat{e}_z$$

with the boundary conditions

$$\begin{aligned} \kappa_1 \phi_z^1 &= \kappa_2 \phi_z^2 \\ \frac{\partial \phi_z^1}{\partial z} &= \frac{\partial \phi_z^2}{\partial z}. \end{aligned} \tag{8}$$

In the semi-infinite cladding media above and below the slab, the trial solutions are homogeneous while inside the slab there are both inhomogeneous and homogenous contributions given by

$$\begin{aligned} \phi_z^{II}(\vec{R}) &= \int_0^\infty k_0 \lambda \, d\lambda J_0(\lambda\tilde{\rho}) e^{-\gamma_2(\tilde{z}-\tilde{L}/2)} \frac{j}{\gamma_2}, \\ &\qquad \tilde{z} > \tilde{L}/2 \\ \phi_z^{I}(\vec{R}) &= \int_0^\infty k_0 \lambda \, d\lambda J_0(\lambda\tilde{\rho}) \\ &\quad \times \left[\frac{1}{\kappa_1 \gamma_1} e^{-\gamma_1 |\tilde{z}-\tilde{a}|} + \frac{g}{\gamma_1} e^{-\gamma_1 \tilde{z}} + \frac{h}{\gamma_1} e^{+\gamma_1 \tilde{z}} \right], \\ &\qquad -\tilde{L}/2 < \tilde{z} < \tilde{L}/2 \\ \phi_z^{III}(\vec{R}) &= \int_0^\infty k_0 \lambda \, d\lambda J_0(\lambda\tilde{\rho}) e^{\gamma_2(\tilde{z}+\tilde{L}/2)} \frac{f}{\gamma_2}, \\ &\qquad \tilde{z} < -\tilde{L}/2. \end{aligned} \tag{9}$$

Applying the boundary conditions (8) at the top and bottom edges of the slab in the usual way results in (10) as shown at the bottom of the page where

$$R_p = \frac{1-\alpha}{1+\alpha}, \qquad \text{with } \alpha = \frac{\kappa_1 \gamma_2}{\kappa_2 \gamma_1}. \tag{11}$$

These expressions have very simple physical interpretations. R_p is the magnetic field reflectivity for a TM-polarized plane wave incident from region I onto region II. The prefactor of $(1-R_p)$ accounts for the transmission into region II. The two terms in the numerator arise from 1) emission toward the top interface and 2) emission toward the bottom interface that is reflected into the upward direction at the bottom interface. The denominator accounts for multiple reflections in this planar cavity. It is straightforward to show that the poles of the integrand (zeros of the denominator) correspond to the (even and odd) bound TM-waveguide modes. The coupling to these modes will be evaluated by the residues at the poles when the integration is carried out. There are branch points in the λ plane as a result of the multivalued γ_i's. The integration around the branch cuts corresponds to the 3-D radiation modes. As noted above, a significant virtue of this formulation is that both the 2-D waveguide modes and the 3-D radiation fields are obtained consistently from a single calculation. Before considering the evaluation of the integrals in (10), the equivalent results for a horizontal dipole will be presented.

IV. Horizontal Dipole

A two-component Hertz vector is required [1], [2] to describe the fields of a horizontal dipole:

$$\vec{\Pi}^i = \phi_x^i \hat{e}_x + \phi_z^i \hat{e}_z$$

with the boundary conditions

$$\begin{aligned} \kappa_1 \phi_x^1 &= \kappa_2 \phi_x^2 \\ \kappa_1 \frac{\partial \phi_x^1}{\partial z} &= \kappa_2 \frac{\partial \phi_x^2}{\partial z} \\ \kappa_1 \phi_z^1 &= \kappa_2 \phi_z^2 \\ \frac{\partial \phi_x^1}{\partial x} + \frac{\partial \phi_z^1}{\partial z} &= \frac{\partial \phi_x^2}{\partial x} + \frac{\partial \phi_z^2}{\partial z}. \end{aligned} \tag{12}$$

The x and z components are coupled by the last boundary condition, so the approach is to solve first for the x components using the first two boundary conditions and then use that solution to evaluate the z components. The inhomogeneous driving term appears only in the x component inside the slab. The result for the x components is shown in (13) at the bottom of the next page where

$$R_s = \frac{1-\beta}{1+\beta}, \qquad \text{with } \beta = \frac{\gamma_2}{\gamma_1}. \tag{14}$$

These expressions are very similar to the vertical dipole case. R_p is replaced with R_s throughout where R_s is the electric field reflectivity for a TE-polarized plane wave incident on the dielectric interface from medium I. ε_2 now appears in the expressions for the Hertz vector outside of the slab rather than ε_1 as was the

$$\begin{aligned} \phi_z^{II}(\vec{R}) &= \int_0^\infty k_0 \lambda \, d\lambda J_0(\lambda\tilde{\rho}) \frac{(1-R_p) e^{-\gamma_1 \tilde{L}/2} (e^{-\gamma_1 \tilde{a}} + R_p e^{-\gamma_1 \tilde{L}} e^{+\gamma_1 \tilde{a}})}{\kappa_1 \gamma_2 (1 - R_p^2 e^{-2\gamma_1 \tilde{L}})} e^{-\gamma_2(\tilde{z}-\tilde{L}/2)} \\ \phi_z^{I}(\vec{R}) &= \int_0^\infty k_0 \lambda \, d\lambda J_0(\lambda\tilde{\rho}) \left[\frac{e^{-\gamma_1 |\tilde{z}-\tilde{a}|}}{\kappa_1 \gamma_1} + \frac{2R_p e^{-\gamma_1 \tilde{L}} \{\cosh[\gamma_1(\tilde{z}+\tilde{a})] + R_p e^{-\gamma_1 \tilde{L}} \cosh[\gamma_1(\tilde{z}-\tilde{a})]\}}{\kappa_1 \gamma_1 (1 - R_p^2 e^{-2\gamma_1 \tilde{L}})} \right] \\ \phi_z^{III}(\vec{R}) &= \phi_z^{II}(\vec{R}), \qquad \text{with } \tilde{a} \to -\tilde{a} \end{aligned} \tag{10}$$

case for the vertical dipole. The poles in these integrands now correspond to the TE waveguide modes. These results can now be used to evaluate the z components of the Hertz vectors as shown in (15) at the bottom of the page.

These expressions now involve both the TE and TM poles. Fortunately, as is shown below, much of the algebraic complexity of these expressions is simplified when the residues are evaluated.

V. Evaluation of the Inverse Transform

The formidable apparatus of complex analysis can now be used on these integral solutions. The manipulations are similar for both vertical and horizontal dipole cases and are presented in great detail for a single interface in [2]. Only a very abbreviated presentation for the vertical dipole is given. To avoid undue repetition, the results for the horizontal dipole are presented without detailed derivation.

The first step is to extend the path of integration along the entire Re(λ) axis by shifting to the Hankel function form [cf. (6)] of the integrals. The singularities of the integrand in the λ plane are illustrated in Fig. 2. As noted above, there is a pole of the integrand corresponding to the lowest order bound mode (TM for a vertical dipole). For simplicity of illustration, it is assumed that the slab is sufficiently thin ($L < \lambda/|n_1|$) so that only the zeroth-order mode is supported. The analytic results are presented for an arbitrary slab thickness. As a result of the multivalued square root functions, there are branch points at $\lambda = \pm n_1$ and $\pm n_2$. The branch cuts are chosen to extend parallel to the Im(λ) axis to $\pm\infty$. There is also a branch point at $\lambda = 0$ corresponding to the divergence of the Hankel function at the origin; the corresponding branch cut extends to $-\infty$ along the negative Im(λ) axis and plays no role in the evaluation of the radiation fields.

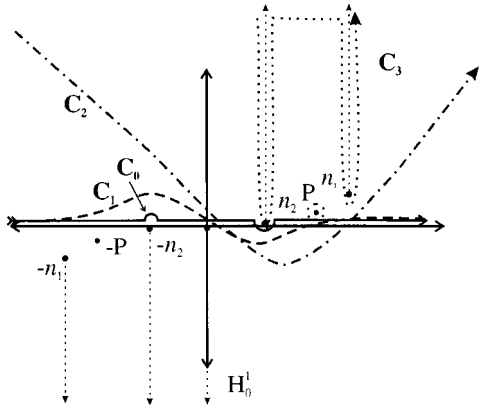


Fig. 2. Singularities and paths of integration in the complex λ plane for the inverse transform. The branch points correspond to the square roots in n_1 and n_2 and to the properties of the Hankel function. The poles correspond to the waveguide modes. The integration path labeled C_0 is the original inverse transform path, the curves labeled $C_1 - C_3$ are the steepest descents paths for evaluation of the inverse transform for external angles θ of 0° (perpendicular to the slab), 45°, and 90° (along the slab).

A. Radiation Modes for Vertical Dipole

Also shown in Fig. 2 are several integration paths that are convenient for the evaluation of the inverse transform. The solid path, C_0, is the undeformed path that corresponds to the derivation of the inverse transform. It passes below the branch points at $\lambda = n_1$ and n_2, above the branch points at $\lambda = -n_1, -n_2$ as well as above the branch point at the origin. A small imaginary part to n_2 has been assumed in drawing this figure; this imaginary part is set to zero in the final result. A steepest descents evaluation, valid for large distances from the source, provides an analytic approach to integration along the branch cuts. This approach allows evaluation of the radiation fields of the dipole a distances greater than several wavelengths from the

$$\phi_x^{II}(\vec{R}) = \int_0^\infty k_0\lambda\, d\lambda J_0(\lambda\tilde{\rho})\frac{(1-R_s)e^{-\gamma_1\tilde{L}/2}(e^{-\gamma_1\tilde{a}} + R_s e^{-\gamma_1\tilde{L}}e^{+\gamma_1\tilde{a}})}{\kappa_2\gamma_2(1-R_s^2e^{-2\gamma_1\tilde{L}})}\, e^{-\gamma_2(\tilde{z}-\tilde{L}/2)}$$

$$\phi_x^{I}(\vec{R}) = \int_0^\infty k_0\lambda\, d\lambda J_0(\lambda\tilde{\rho})\left[\frac{e^{-\gamma_1|\tilde{z}-\tilde{a}|}}{\kappa_1\gamma_1} + \frac{2R_s e^{-\gamma_1\tilde{L}}\{\cosh[\gamma_1(\tilde{z}+\tilde{a})] + R_s e^{-\gamma_1\tilde{L}}\cosh[\gamma_1(\tilde{z}-\tilde{a})]\}}{\kappa_1\gamma_1(1-R_s^2e^{-2\gamma_1\tilde{L}})}\right]$$

$$\phi_x^{III}(\vec{R}) = \phi_x^{II}(\vec{R}), \qquad \text{with } \tilde{a} \to -\tilde{a} \tag{13}$$

$$\phi_z^{II} = \int_0^\infty \lambda\, d\lambda J_0(\lambda\tilde{\rho})\frac{ik_0\xi(1-R_s)(\kappa_1-\kappa_2)e^{-\gamma_1\tilde{L}/2}e^{-\gamma_2(\tilde{z}-\tilde{L}/2)}}{\kappa_2\gamma_2(1-R_s^2e^{-2\gamma_1\tilde{L}})(1-R_p^2e^{-2\gamma_1\tilde{L}})(\kappa_2\gamma_1+\kappa_1\gamma_2)}$$
$$\times\left\{e^{\gamma_1\tilde{a}}\left[1+(R_p-R_s-R_sR_p)e^{-2\gamma_1\tilde{L}}\right] + e^{-\gamma_1\tilde{a}}e^{-\gamma_1\tilde{L}}\left[-1+R_s-R_p+R_sR_pe^{-2\gamma_1\tilde{L}}\right]\right\}$$

$$\phi_z^{I} = \int_0^\infty \lambda\, d\lambda J_0(\lambda\tilde{\rho})\frac{2ik_0\xi(1-R_s)(\kappa_1-\kappa_2)e^{-\gamma_1\tilde{L}}}{\kappa_1\gamma_2(1-R_s^2e^{-2\gamma_1\tilde{L}})(1-R_p^2e^{-2\gamma_1\tilde{L}})(\kappa_2\gamma_1+\kappa_1\gamma_2)}$$
$$\times\left[\left(1-R_pR_se^{-2\gamma_1\tilde{L}}\right)\sinh[\gamma_1(\tilde{z}+\tilde{a})] + (R_s-R_p)e^{-\gamma_1\tilde{L}}\sinh[\gamma_1(\tilde{z}-\tilde{a})]\right]$$

$$\phi_z^{III} = \phi_z^{II}, \qquad \text{with } \tilde{a} \to -\tilde{a} \tag{15}$$

source; however, the formulation of the problem correctly includes all near-field effects at the slab boundaries which can be much less than a wavelength from the dipole. The appropriate integration path for this evaluation depends on the external polar angle, θ, for which the integration is performed. These paths are shown for three particular values of θ in the figure and for evaluation of the integral around the n_2 branch cut, corresponding to the radiation fields in medium II. The path for evaluation of the radiation fields normal to the slab, $\theta = 0$, is given by the (dashed) path C_1. The path for an intermediate angle $\theta = \pi/4$ is denoted (dash–dot) C_2. Finally, the path for evaluation of the fields very close to the slab $\theta = \pi/2$ is denoted (dotted) C_3. In each case, the slopes of the paths asymptotically approach $\mathrm{Im}(\lambda)/\mathrm{Re}(\lambda) = \tan\theta$ for large λ.

For $\theta = 0$, deformation of the path of integration from C_0 to C_1 does not pass any singularities in the complex plane, so the steepest descents integration gives the entire response. As θ increases, the contributions from crossing the branch point at n_1 and the pole P must be included in the response. The integration around the n_1 branch cut contributes only to near fields in medium II and can be neglected since we are interested only in the radiation fields many wavelengths away from the slab. The roles of the integrals around the branch cuts are reversed for evaluation of the radiation fields in medium I. The residue at the pole gives the radiation coupled into the bound waveguide mode (including the evanescent fields in the cladding regions).

To make this more concrete, evaluate the radiation fields for arbitrary angles in region II. The starting point is (10), modified by the transition to the Hankel function as described above:

$$\phi_z^{II}(\vec{R}) = \int_{-\infty}^{\infty} k_0\lambda\, d\lambda H_0^1(\lambda\tilde{\rho}) \times \frac{(1-R_p)e^{-\gamma_1\tilde{L}/2}(e^{-\gamma_1\tilde{a}} + R_p e^{-\gamma_1\tilde{L}}e^{+\gamma_1\tilde{a}})}{2\kappa_1\gamma_2(1-R_p^2e^{-2\gamma_1\tilde{L}})} \times e^{-\gamma_2(\tilde{z}-\tilde{L}/2)}. \tag{16}$$

Substitute an integral expression for the Hankel function [2], viz.

$$H_0^1(x) = \frac{4e^{ix}}{\pi}\int_0^{\infty}\frac{e^{-y^2/2}}{\sqrt{4ix-y^2}}\,dy \tag{17}$$

and reverse the order of integration to give

$$\phi_z^{II}(\vec{R}) = \frac{2k_0}{\pi}\int_0^{\infty} dy e^{-y^2/2}\int_{-\infty}^{\infty}\frac{k_0\lambda}{\sqrt{4i\lambda\tilde{\rho}-y^2}}\,d\lambda \times \frac{(1-R_p)e^{(\gamma_2-\gamma_1)\tilde{L}/2}(e^{-\gamma_1\tilde{a}} + R_p e^{-\gamma_1\tilde{L}}e^{+\gamma_1\tilde{a}})}{2\kappa_1\gamma_2(1-R_p^2e^{-2\gamma_1\tilde{L}})} \times e^{i\lambda\tilde{\rho}-\gamma_2\tilde{z}}. \tag{18}$$

The next step is to transform the last exponential into a suitable expression for a steepest descents evaluation (i.e., to find a path through a saddle point with a constant imaginary argument to eliminate rapid oscillations of the exponential). Transform to spherical coordinates ($\tilde{\rho} = \tilde{R}\sin\theta$; $\tilde{z} = \tilde{R}\cos\theta$) and make the conformal mapping:

$$\begin{aligned} \lambda &= n_2\sin\tau \\ d\lambda &= n_2\cos\tau\, d\tau \\ \gamma_2 &= \sqrt{\lambda^2-\kappa_2} \to -in_2\cos\tau \\ \gamma_1 &= \sqrt{\lambda^2-\kappa_1} \to -in_1\sqrt{1-m^2\sin^2\tau} \end{aligned} \tag{19}$$

with

$$m \equiv n_2/n_1. \tag{20}$$

Since we are only interested in the lowest order radiation terms (scaling as $1/R$), it is permissible to neglect the y^2 in the radical and carry out the simple Gaussian integral over y. Including these transformations and approximations gives (21) as shown at the bottom of the page where the path of integration is initially over the transformed $\mathrm{Re}(\lambda)$ axis. The deformation of this integration path to the steepest descents path will be carried out after the next step. Now define

$$w = \tau - \theta \tag{22}$$

to express the exponential part of the integrand as a simple cosine function

$$\phi_z^{II}(\vec{R}) = \frac{2ik_0n_2}{\kappa_1}\sqrt{\frac{2}{\pi}}\int_C \times \frac{\sin(w+\theta)}{\sqrt{4in_2\tilde{R}\sin\theta\sin(w+\theta)}}\,dw \times F(\theta, w+\theta)e^{in_2\tilde{R}\cos w}. \tag{23}$$

The saddle point for the integration is clearly at $w = 0$ and the steepest descents path of integration is chosen to keep $\mathrm{Re}(\cos w) = 1$ and expressing $w = u + iv$, gives for the integration path $\cosh(u)\cos(v) = 1$. Tracing this equation back through the transform gives the contours already discussed in the λ plane.

$$\phi_z^{II}(\vec{R}) = \frac{2ik_0n_2}{\kappa_1}\sqrt{\frac{2}{\pi}}\int_C \frac{\sin\tau\, d\tau}{\sqrt{4in_2\tilde{R}\sin\theta\sin\tau}}F(\theta,\tau)e^{in_2\tilde{R}(\sin\theta\sin\tau+\cos\theta\cos\tau)}$$

$$F(\theta,\tau) = \frac{(1-R_p)\left[e^{in_1(\tilde{L}/2-\tilde{a})\sqrt{1-m^2\sin^2\theta}} + R_p e^{in_1(3\tilde{L}/2+\tilde{a})\sqrt{1-m^2\sin^2\theta}}\right]}{1-R_p^2e^{2in_1\tilde{L}\sqrt{1-m^2\sin^2\theta}}}\, e^{i(n_1\sqrt{1-m^2\sin^2\theta}-n_2\cos\theta)\tilde{L}/2} \tag{21}$$

One final transformation is needed to complete the integration, *viz*.

$$\cos w = 1 - \frac{x^2}{2in_2\tilde{R}} \tag{24}$$

then

$$\phi_z^{II}(\vec{R}) = \frac{2ik_0n_2}{\kappa_1}\, e^{in_2\tilde{R}}\sqrt{\frac{2}{\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx \times \frac{\sin(w+\theta)}{\sqrt{4in_2\tilde{R}\,\sin\theta\,\sin(w+\theta)}}\,\frac{dw}{dx}\,F(\theta,\, w+\theta) \tag{25}$$

which displays the usual steepest descents form with w now a function of x. The dominant contributions to the integral occur for small x and the remaining terms need only be expanded in a power series around $x = 0$. Since we are interested in the radiation terms only, it is sufficient to evaluate all of the terms in the integrand except for the Gaussian exponential at $x = 0$, giving the final result

$$\phi_z^{II}(\vec{R}) \xrightarrow[3D]{} \frac{k_0 e^{in_2\tilde{R}}}{\kappa_1\tilde{R}}\, e^{i(n_1\cos\vartheta - n_2\cos\theta)\tilde{L}/2} \times \frac{(1-R_{p0})[e^{-in_1\tilde{a}\cos\vartheta} + R_{p0}e^{in_1(\tilde{L}+\tilde{a})\cos\vartheta}]}{1-R_{p0}^2e^{2in_1\tilde{L}\cos\vartheta}} \tag{26}$$

where

$$\cos\vartheta \equiv \sqrt{1-m^2\sin^2\theta}$$
$$R_{p0} = \frac{m\cos\vartheta - \cos\theta}{m\cos\vartheta + \cos\theta}. \tag{27}$$

After all of this complex manipulation, the final result has a very intuitive form. Remember that θ is the external angle in the cladding material above the dielectric slab. Then from Snell's law, ϑ is the internal angle within the slab. R_{p0} is the angularly dependent magnetic field reflectivity and since by definition $m < 1$, $R_{p0} < 0$ at $\theta = 0$. The first factor in the result is the unbounded medium I contribution to the Hertz vector (5), the second factor is an almost pure phase shift reflecting the division of the propagation between medium I and medium II, and the third factor describes the angularly dependent transmission into medium II in terms of the familiar TM-polarized Fresnel reflectivity modified by multiple reflections in the slab geometry. This term has a series of maxima at $2n_1\tilde{L}\cos\vartheta = 2l\pi$ (l an integer) as a result of the Fabry–Perot modes of the slab geometry. The two terms in the numerator come from the upward and the downward-and-reflected emitted waves in the 2-D Fourier transform of the dipole radiation, at $a = 0$ dipole in the middle of the slab, these two terms combine to eliminate every other slab resonance. As a result of the sign of R_{p0}, the even resonances are absent in the response.

In order to proceed further, it is convenient to shift from the Hertz vector to the corresponding electric and magnetic fields using (1) and (2). Since again we are only interested in the far fields, this is a relatively simple operation with the result

$$E_\theta \propto k_0^3m^2\sin\theta F_z^V(\theta,\,\tilde{L},\,\tilde{a})\frac{e^{in_2\tilde{R}}}{\tilde{R}}$$
$$H_\varphi \propto \frac{k_0^3n_2}{\eta_0}\,m^2\sin\theta F_z^V(\theta,\,\tilde{L},\,\tilde{a})\frac{e^{in_2\tilde{R}}}{\tilde{R}} \tag{28}$$

with

$$F_z^V(\theta,\,\tilde{L},\,\tilde{a}) = e^{i(n_1\cos\vartheta - n_2\cos\theta)\tilde{L}/2} \times \frac{(1-R_{p0})\left[e^{-in_1\tilde{a}\cos\vartheta} + R_{p0}e^{in_1(\tilde{L}+\tilde{a})\cos\vartheta}\right]}{1-R_{p0}^2e^{2in_1\tilde{L}\cos\vartheta}}. \tag{29}$$

For a single interface,

$$F_z^V(\theta,\,\tilde{L},\,\tilde{a}) \xrightarrow[\text{single interface}]{} e^{in_1(\tilde{L}/2-\tilde{a})\cos\vartheta}e^{-in_2\tilde{L}\cos\theta/2}(1-R_{p0}). \tag{30}$$

Finally integrating across the entire upper hemisphere and normalizing to the power radiated by the dipole in an unbounded medium of dielectric constant ε_1 gives

$$\hat{P}_V^{II} = \frac{3|m|^5}{4}\int_0^{\pi/2}|F_z^V(\theta,\,\tilde{L},\,\tilde{a})|^2\sin^3\theta\,d\theta. \tag{31}$$

Here the "hat" is a reminder that this result is normalized to the unbounded medium radiated power. The expression for the power radiated into the lower half-space is identical with $a \to -a$. Note the very strong dependence of the radiated power on the dielectric discontinuity $\propto m^5$; for a GaAs slab in air, $m \sim 1/3.6$ and $m^5 \sim 4\times10^{-3}$. As is well known, very little of the radiated power escapes from a high-dielectric constant slab. For a single interface, this radiation distribution has been derived [18] previously with a very different approach.

As is shown below, for a horizontal dipole, a larger fraction of the dipole power is radiated for the same dielectric discontinuity. Detailed evaluation of the radiated power is deferred to a later section after the equivalent result for the horizontal dipole is presented. The question we next address is the power radiated into the waveguide slab modes, since so little power escapes from the slab, we expect that much of the radiation power is contained in these slab modes.

B. 2-D Waveguide Modes for Vertical Dipole

As has been noted already, the waveguide modes correspond to the residues at the poles of the denominator that correspond to the slab waveguide modes. These correspond to $R_pe^{-\gamma_1\tilde{L}} = \pm1$ with the plus sign for even modes and the minus sign for the odd modes. For a sufficiently thin slab ($L < \lambda/n_1$), there is only a single even mode. We start the evaluation for this case and generalize later to arbitrary L. For distances of several wavelengths from the source, it is appropriate to use the asymptotic limit of the Hankel function

$$H_0^1(z) \xrightarrow[z\,\text{large}]{} -i\sqrt{\frac{2}{\pi z}}\,e^{iz}. \tag{32}$$

Applying these results to the Hertz vector contribution inside the slab, and neglecting the inhomogeneous solution that does not have the pole associated with the cavity resonances, gives

$$\phi_z^I(\vec{R}) = \frac{ik_0}{\kappa_1}\sqrt{\frac{8}{\pi}}\int_{-\infty}^{\infty}\frac{\lambda\,d\lambda}{\sqrt{\lambda\tilde{\rho}}}\,e^{i\lambda\tilde{\rho}} \times\left[\frac{\cosh(\gamma_1\tilde{z})\cosh(\gamma_1\tilde{a})}{\gamma_1(1-R_pe^{-\gamma_1\tilde{L}})(1+R_pe^{-\gamma_1\tilde{L}})}\right]. \quad (33)$$

The next step is to expand the denominator around the pole at $\lambda = k_p$ where k_p is the solution of the transcendental dispersion relationship for the mode. After some algebra, the result is

$$1-R_pe^{-\gamma_1\tilde{L}} \approx k_p(\lambda-k_p)\frac{\tilde{L}}{\gamma_{1p}} \times\left[1+\frac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}[(\kappa_1+\kappa_2)k_p^2-\kappa_1\kappa_2]}\right] \quad (34)$$

where γ_{ip} denotes γ_i evaluated at $\lambda = k_p$. Using this expression, the equivalent integral to (16) for medium I can be evaluated as $2\pi i\times$ (residue at the pole), giving the simple result

$$\phi_z^I(\vec{R}) = k_0\sqrt{\frac{8\pi}{k_p\tilde{\rho}}}\,e^{ik_p\tilde{\rho}} \times\frac{\cosh(\gamma_{1p}\tilde{z})\cosh(\gamma_{1p}\tilde{a})}{\kappa_1\tilde{L}\left[1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}[(\kappa_1+\kappa_2)k_p^2-\kappa_1\kappa_2]}\right]}. \quad (35)$$

Detailed discussion will be deferred until later; at this point, it is appropriate to note a few simple features of this result. First, this represents a cylindrical wave propagating along the slab and decaying as $1/\sqrt{\tilde{\rho}}$, in contrast to the 3-D result from the branch cut integration, (26), that gave a spherical wave propagating in the radial direction decaying as $1/\tilde{R}$. The $\cosh(\gamma_{1p}\tilde{z})$ represents the mode variation across the slab, while the $\cosh(\gamma_{1p}\tilde{a})$ is the overlap of the dipole with the mode profile which is strongest in the middle of the slab. For small $\tilde{L}$, $\phi_z^I \to 0$ since $k_p \to n_2$ and, therefore, $\gamma_2 \to 0$; while, for large $\tilde{L}$, ϕ_z^I scales as $1/\tilde{L}$.

As was the case for the 3-D modes, the next step is to convert from the Hertz vector to the corresponding electric and magnetic fields

$$\vec{E}^I = \frac{k_0^3k_p^2}{\kappa_1}\sqrt{\frac{8\pi}{k_p\tilde{\rho}}}\,e^{ik_p\tilde{\rho}}\left[\frac{i\gamma_{1p}}{k_p}\sinh(\gamma_{1p}\tilde{z})\hat{e}_\rho+\cosh(\gamma_{1p}\tilde{z})\hat{e}_z\right] \times\frac{\cosh(\gamma_{1p}\tilde{a})}{\tilde{L}\left[1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}[(\kappa_1+\kappa_2)k_p^2-\kappa_1\kappa_2]}\right]}$$

$$\vec{H}^I = -\frac{k_0^3}{\eta_0}\sqrt{\frac{8\pi k_p}{\tilde{\rho}}}\,e^{ik_p\tilde{\rho}}\cosh(\gamma_{1p}\tilde{z})\hat{e}_\varphi \times\frac{\cosh(\gamma_{1p}\tilde{a})}{\tilde{L}\left[1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}[(\kappa_1+\kappa_2)k_p^2-\kappa_1\kappa_2]}\right]}. \quad (36)$$

The evaluation of the power radiated into the mode requires an integration across the mode profile that scales as $\cosh(\gamma_{1p}\tilde{z})$ inside the slab and as $\cosh(\gamma_{1p}\tilde{L}/2)e^{\mp\gamma_{2p}(\tilde{z}\mp\tilde{L}/2)}$ in the cladding regions above ($-$ signs) and below ($+$ signs) the slab. This integration gives

$$\begin{aligned}O_{\rm TM}^e &= \mathrm{Re}\left(\frac{1}{\kappa_1}\right)\int_{-L/2}^{L/2}|\cosh(\gamma_{1p}k_0z)|^2\,dz + \frac{2|\cosh(\gamma_{1p}k_0L/2)|^2}{\kappa_2}\int_0^\infty e^{i(\gamma_{2p}-\gamma_{2p}^*)k_{0z}}\,dz \\ &= \frac{1}{2}\mathrm{Re}\left(\frac{1}{\kappa_1}\right)\left(\frac{\sinh(\gamma_{1p}'k_0L)}{\gamma_{1p}'}+\frac{\sin(\gamma_{1p}''k_0L)}{\gamma_{1p}''}\right) + \frac{|\cosh(\gamma_{1p}k_0L/2)|^2}{\kappa_2\gamma_{2p}''} \end{aligned} \quad (37)$$

where the real and imaginary parts of the γ_i's are denoted by γ_i' and γ_i'', respectively. The final result for the power radiated into the TM mode by a vertical dipole, integrated across all angles and normalized to the unbounded medium radiated power, is

$$\hat{P}_{\rm TM}^V = \frac{3\pi\,|\cosh(\gamma_{1p}\tilde{a})|^2\,|k_p|^2O_{\rm TM}^e}{\mathrm{Re}(n_1)\tilde{L}^2\left|1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}\left[(\kappa_1+\kappa_2)k_p^2-\kappa_1\kappa_2\right]}\right|^2}. \quad (38)$$

Generalizing this result to an arbitrarily slab thickness involves summation over all allowed even and odd modes, viz.

$$\hat{P}_{\rm TM}^V = \sum_{even}\frac{3\pi\,|\cosh(\gamma_{1p}\tilde{a})|^2\,|k_p|^2O_{\rm TM}^e}{\mathrm{Re}(n_1)\tilde{L}^2\left|1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}\left[(\kappa_1+\kappa_2)(k_p^n)^2-\kappa_1\kappa_2\right]}\right|^2} + \sum_{odd}\frac{3\pi\,|\sinh(\gamma_{1p}\tilde{a})|^2\,|k_p|^2O_{\rm TM}^o}{\mathrm{Re}(n_1)\tilde{L}^2\left|1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2p}\tilde{L}\left[(\kappa_1+\kappa_2)(k_p^n)^2-\kappa_1\kappa_2\right]}\right|^2} \quad (39)$$

where

$$O_{\rm TM}^o = \frac{1}{2}\mathrm{Re}\left(\frac{1}{\kappa_1}\right)\left(\frac{\sinh(\gamma_{1p}'k_0L)}{\gamma_{1p}'}-\frac{\sin(\gamma_{1p}''k_0L)}{\gamma_{1p}''}\right) + \frac{|\sinh(\gamma_{1p}k_0L/2)|^2}{\kappa_2\gamma_{2p}''} \quad (40)$$

is defined analogously with $O_{\rm TM}^e$.

C. 3-D Radiation Modes for Horizontal Dipole

The calculation proceeds in much the same way as that for the vertical dipole, with the added complexity of a two-component Hertz vector. For the ϕ_x component, the result is available from that of the vertical dipole with the substitutions noted above for

the Hertz vector

$$\phi_x^{\mathrm{II}}\left(\vec{R}\right) \underset{3D}{\longrightarrow} \frac{k_0 e^{in_2\tilde{R}}}{\kappa_2\tilde{R}}\, e^{i(n_1\cos\vartheta - n_2\cos\theta)\tilde{L}/2} \times \frac{(1-R_{s0})\left[e^{-in_1\tilde{a}\cos\vartheta} + R_{s0}e^{in_1(\tilde{L}+\tilde{a})\cos\vartheta}\right]}{1-R_{s0}^2\, e^{2in_1\tilde{L}\cos\vartheta}} \equiv F_x^H\left(\theta,\tilde{L},\tilde{a}\right)\frac{e^{in_2\tilde{R}}}{\kappa_2\tilde{R}} \tag{41}$$

where

$$\cos\vartheta \equiv \sqrt{1-m^2\sin^2\theta}; \quad R_{s0} = \frac{\cos\vartheta - m\cos\theta}{\cos\vartheta + m\cos\theta}. \tag{42}$$

Here R_{s0} is the TE-polarized electric field Fresnel reflectivity across the boundary from region I into region II. The interpretation of this result is very similar to that discussed above; we will show below that, while the interference of the emission components arising from the up and down directions again cancels half of the cavity resonances in the denominator, because R_{s0} is positive for small angles, while R_{p0} is negative, the alternate set of resonances is eliminated.

As expected from the Hertz vector, the z component is more algebraically complex involving both the TE- and TM-Fresnel reflectivities as shown in (43) at the bottom of the page. As above, the next step is to obtain the electric and magnetic radiation fields in the cladding half-plane from the Hertz vector components.

$$\vec{E} = \frac{k_0^2 e^{in_1\tilde{R}}}{\tilde{R}} \times \left[F_x^H\sin\varphi\hat{e}_\varphi + (F_x^H\cos\theta - F_z^H\sin^2\theta)\cos\varphi\hat{e}_\theta\right]$$

$$\vec{H} = \frac{k_0^2 k_2 e^{in_1\tilde{R}}}{\eta_0\tilde{R}} \times \left[(F_x^H\cos\theta - F_z^H\sin^2\theta)\cos\varphi\hat{e}_\varphi - F_x^H\sin\varphi\hat{e}_\theta\right]$$

$$\hat{P}_{rad}^H = \frac{3|m|}{8}\int_0^{\pi/2} \times \left[|F_x^H|^2 + |F_x^H\cos\theta - F_z^H\sin^2\theta|^2\right]\sin\theta\, d\theta. \tag{44}$$

In the single interface limit, these results reduce to

$$F_x^H\left(\theta,\tilde{L},\tilde{a}\right) \xrightarrow[\text{single interface}]{} \times e^{in_1(\tilde{L}/2-\tilde{a})\cos\vartheta}e^{-in_2\tilde{L}\cos\theta/2}(1-R_{s0})$$

$$F_z^H\left(\theta,\tilde{L},\tilde{a}\right) \xrightarrow[\text{single interface}]{} \times e^{in_1(\tilde{L}/2-\tilde{a})\cos\vartheta}e^{-in_2\tilde{L}\cos\theta/2}\frac{n_2(1-R_{s0})(\kappa_1-\kappa_2)}{\kappa_2 n_1\cos\vartheta + \kappa_1 n_2\cos\theta}. \tag{45}$$

Detailed discussion of this result and the comparison between the horizontal and vertical dipoles in various dielectric structures will be deferred until after the waveguide mode radiation results for the horizontal dipole are presented.

D. 2-D Waveguide Modes for Horizontal Dipole

The horizontal dipole can couple to TE (s) waveguide modes through the $\hat{e}_\varphi$ component of the mode electric field and to TM (p) waveguide modes through the $\hat{e}_\rho$ component. Both sets of poles occur in the ϕ_z^H response. The procedure for evaluating the radiated power is similar to that outlined above and only the final results are given here:

$$\hat{P}_{\mathrm{TE}}^H = \sum_{even}\frac{3\pi}{2\mathrm{Re}(n_1)}\frac{|\cosh(\gamma_{1s}\tilde{a})|^2}{\tilde{L}^2\left|1+\frac{2}{\gamma_{2s}\tilde{L}}\right|^2}O_{\mathrm{TE}}^e + \sum_{odd}\frac{3\pi}{2\mathrm{Re}(n_1)}\frac{|\sinh(\gamma_{1s}\tilde{a})|^2}{\tilde{L}^2\left|1+\frac{2}{\gamma_{2s}\tilde{L}}\right|^2}O_{\mathrm{TE}}^o$$

$$O_{\mathrm{TE}}^e = \frac{\sinh(\gamma'_{1s}\tilde{L})}{\gamma'_{1s}} + \frac{\sin(\gamma''_{1s}\tilde{L})}{\gamma''_{1s}} + 2\frac{\left|\cosh(\gamma_{1s}\tilde{L}/2)\right|^2}{\gamma'_{2s}}$$

$$O_{\mathrm{TE}}^o = \frac{\sinh(\gamma'_{1s}\tilde{L})}{\gamma'_{1s}} - \frac{\sin(\gamma'_{1s}\tilde{L})}{\gamma''_{1s}} + 2\frac{\left|\sinh(\gamma_{1s}\tilde{L}/2)\right|^2}{\gamma'_{2s}} \tag{46}$$

$$\phi_z^{\mathrm{II}}\left(\vec{R}\right) \underset{3D}{\longrightarrow} in_2\sin\theta\cos\varphi\,\frac{e^{in_2\tilde{R}}}{\kappa_2\tilde{R}}\,F_z^H\left(\theta,\tilde{L},\tilde{a}\right)$$

$$F_z^H\left(\theta,\tilde{L},\tilde{a}\right) = \frac{e^{i(n_1\cos\vartheta - n_2\cos\theta)\tilde{L}/2}(1-R_{s0})(\kappa_1-\kappa_2)A\left(\theta,\tilde{L},\tilde{a}\right)}{\left(1-R_{p0}^2e^{2in_1\tilde{L}\cos\vartheta}\right)\left(1-R_{s0}^2e^{2in_1\tilde{L}\cos\vartheta}\right)(\kappa_2 n_1\cos\vartheta + \kappa_1 n_2\cos\theta)}$$

$$A\left(\theta,\tilde{L},\tilde{a}\right) = e^{-in_1\tilde{a}\cos\vartheta}\left[1+(-R_{s0}+R_{p0}-R_{s0}R_{p0})e^{2in_1\tilde{L}\cos\vartheta}\right] + e^{in_1(\tilde{L}+\tilde{a})\cos\vartheta}\left[-1+R_{s0}-R_{p0}+R_{s0}R_{p0}e^{2in_1\tilde{L}\cos\vartheta}\right] \tag{43}$$

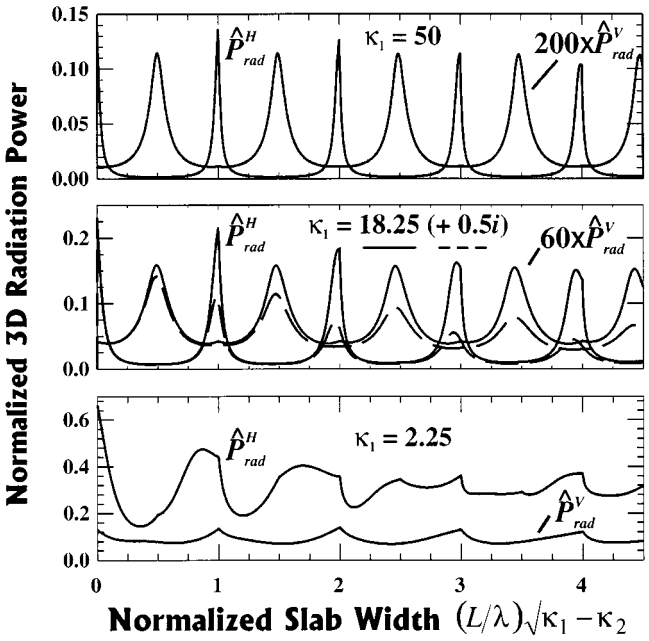


Fig. 3. Total power (above and below the slab) into the 3-D radiation modes (normalized to an unbounded dielectric medium) for horizontal and vertical dipoles as a function of normalized slab width for three dielectric constants. For large dielectric contrasts, the radiation clearly shows alternating Fabry–Perot slab resonances. Radiation is strongly suppressed for the vertical dipole in comparison with the horizontal dipole. For the lowest dielectric contrast, the resonances are less pronounced and correspond to the onset of waveguide modes.

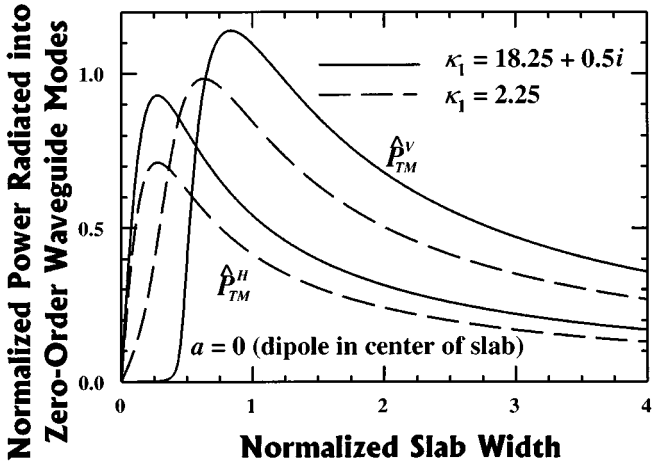


Fig. 4. Power into the lowest, zeroth-order, even waveguide mode (normalized to an unbounded dielectric medium) for dielectric constants corresponding to Si and SiO_2. The peak in the radiated power occurs as the effective mode index shifts from the index of the cladding to that of the slab. The maximum radiated power for a vertical dipole for the large Si dielectric constant is greater than that in an unbounded medium corresponding to an enhanced radiative efficiency.

and

$$\hat{P}^H_{\rm TM} = \sum_{even} \frac{3\pi\,|\sinh(\gamma_{1\rm p}\tilde{a})|^2\,|\gamma_{1\rm p}|^2}{2{\rm Re}(n_1)\tilde{L}^2\left|1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2\rm p}\tilde{L}\left[(\kappa_1+\kappa_2)k_{\rm p}^2-\kappa_1\kappa_2\right]}\right|^2} O^e_{\rm TM} + \sum_{odd} \frac{3\pi\,|\cosh(\gamma_{1\rm p}\tilde{a})|^2\,|\gamma_{1\rm p}|^2}{2{\rm Re}(n_1)\tilde{L}^2\left|1+\dfrac{2\kappa_1\kappa_2}{\gamma_{2\rm p}\tilde{L}\left[(\kappa_1+\kappa_2)k_{\rm p}^2-\kappa_1\kappa_2\right]}\right|^2} O^o_{\rm TM} \tag{47}$$

where the definitions of the mode integrals $O^{e,o}_{\rm TM}$ are given in (37) and (40). The coupling for a dipole in the center of the slab ($\tilde{a} = 0$) is to the even TE and the odd TM modes.

E. 3-D Radiation Coupling

The total 3-D radiation (integrated over both cladding regions above and below the slab) is plotted in Fig. 3 as a function of the normalized slab width for vertical and horizontal dipoles in the center of the slab for three different slab dielectric constant values corresponding to SiO_2 [$\kappa_1 = 2.25$], Si [at a visible wavelength $\kappa_1 = 18.25\ (+0.5i)$], and a fictitious high dielectric constant material [$\kappa_1 = 50$] to demonstrate the limiting behavior of the radiation. The cladding dielectric constant is taken as unity in all cases. A varying scale factor has been applied to the (smaller) vertical dipole results so that both the vertical and horizontal dipole 3-D radiation powers are visible on the same scale. The low efficiency of coupling into the radiation modes for the vertical dipole is expected; because of the small critical angle at high dielectric contrast, only radiation emitted very near vertical, where the dipolar radiation vanishes, can exit the slab.

For the two larger slab dielectric constants, the periodic behavior resulting from the Fabry–Perot denominators in (10) and (13) is clearly evident. Because of the differing signs of R_{p0} and R_{s0}, the resonances alternate with peaks in the horizontal dipole radiation at even numbers of half-wavelengths and in the vertical dipole radiation at odd half-wavelength slab thicknesses. The horizontal dipole resonances are narrower than the corresponding resonances for the vertical dipole. The effects of absorption are indicated as dotted lines in the middle panel for the Si case. For thicker slabs, the resonances broaden out as a result of the angular integrations. It is straightforward to show from the analytic results that the radiated power away from the resonances scales as $|m|^5$ for the vertical dipole and as $|m|^3$ for the horizontal dipole, while at the peaks the scaling is $|m|^3$ (vertical) and $|m|$ (horizontal). For the lower dielectric constant, the resonances are less pronounced with the cusps of the curves scaling in thickness as $\sqrt{\kappa_1 - \kappa_2}/\lambda$ rather than the limiting result of n_1/λ for high dielectric contrast. These features in the 3-D radiation pattern all correspond to the onset of 2-D waveguide modes with increasing slab thickness as is discussed below.

F. 2-D Waveguide Mode Coupling

The coupling into the lowest order bound mode (TE for the horizontal dipole, TM for the vertical dipole) for a dipole in the center of the slab is shown in Fig. 4 as a function of the normalized slab thickness for two slab dielectric constants, again corresponding to SiO_2 and to Si at visible wavelengths. As before, these results are normalized to the unbounded dielectric radiated power. For very thin slabs, $\ll\lambda/n_1$, the radiative coupling vanishes in both orientations because the modes are very spread out ($\gamma_2 \to 0$) and, therefore, their intensity at the slab is very low.

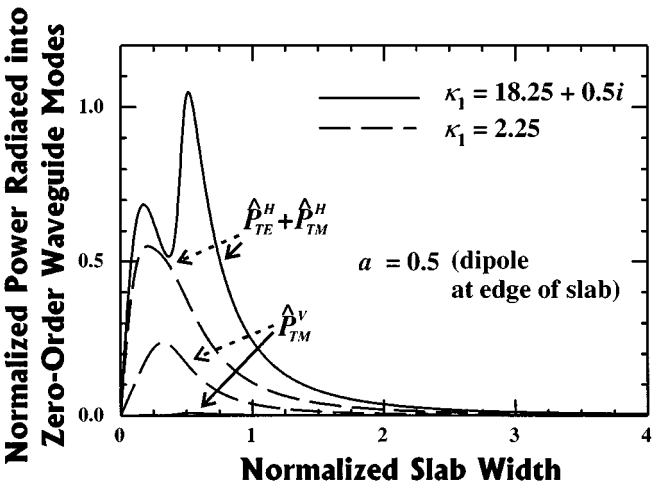


Fig. 5. Same as Fig. 4 except the dipole is at the edge of the slab. There are now two peaks for the horizontal dipole corresponding to both the zero-order TE and TM modes. These are clearly distinct for the higher dielectric constant, but are merged for the lower κ_1. The radiation for the vertical dipole at high κ_1 is almost completely suppressed in this configuration as a result of the very small value of the modal electric field at the edge of the dielectric.

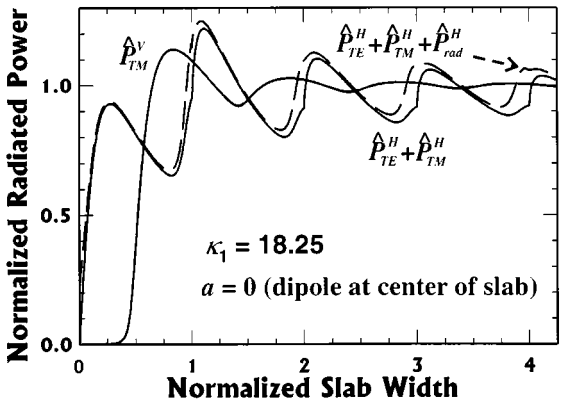


Fig. 6. The distribution of radiated power into all modes for a dipole located in the center of a Si slab. For a vertical dipole, the only contribution is from the even TM modes. Except for very thin slabs, the radiation is roughly constant and equal to that in an unbounded dielectric medium. For a horizontal dipole, there are contributions from both the even TE and odd TM modes. At the peak of the radiative coupling, where coupling to TE_0, TE_2 and TM_1 all contribute, the radiation peaks at about 1.3 times that in bulk. The addition of the 3-D radiation results in the dotted curve for the horizontal dipole, filling in radiative energy just at the onset of each mode. The 3-D radiation contributions are too small to be visible on this scale for the vertical dipole.

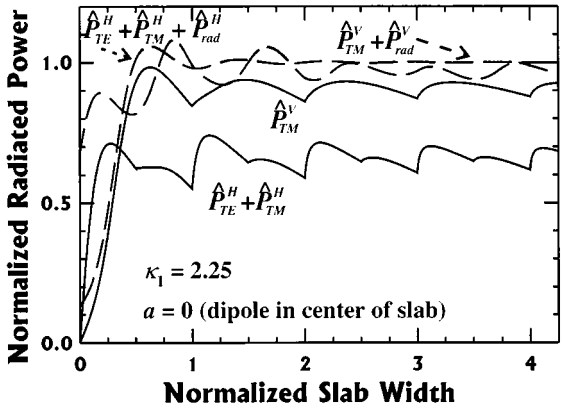


Fig. 7. The distribution of radiated power into all modes for a dipole located in the center of an SiO_2 slab. The larger contribution of the 3-D radiation to the total power is evident for both dipole orientations.

For very wide slabs, again the mode is spread out and of low intensity, from the analytic results the radiation power into a specific mode decays as $1/L^2$. The peak radiative coupling in each case occurs for slab thicknesses where the geometrical waveguiding effects are strongest ($k_{p,s}$ midway between n_2 and n_1). For the Si dielectric constant, the normalized vertical mode radiated power is larger than unity, implying a larger radiative cross section for the dipole than in an unbounded medium. This is by now a familiar cavity electrodynamic result [5]–[7], [9]–[15]. For arbitrarily large dielectric constants, this emission enhancement saturates at about 15% above the unbounded medium result. When the contributions of higher order modes are added, the horizontal dipole emission becomes slightly larger than that of the vertical dipole and saturates at about 30% above the unbounded medium result. Of course, larger enhancements of the emission efficiency are achieved with more dimensions of confinement, e.g., wires and boxes. For this symmetric position in the center of the slab, there is no coupling to the lowest order (even) TM mode [$\propto |\sinh(\gamma_1 a)|^2$]. If the dipole is moved away from the center, this coupling is allowed. An example is shown in Fig. 5 for the dipole just at the edge of the slab for the same dielectric constant values. For the Si dielectric constant, the two peaks corresponding to the TE and TM modes are clearly evident, at the lower SiO_2 dielectric constant they merge into a single, broader peak. The radiative coupling for the vertical dipole is so low as to be barely visible for the Si dielectric constant as a result of the small amplitude of the z-directed TM mode field at the edge of the slab for large dielectric contrast.

As the slab thickness is increased, radiation is possible into other modes and the total radiated power is a sum over all available channels including both 2-D waveguide- and 3-D radiation modes. Both the 2-D waveguide radiation and the total radiated power as a function of slab thickness is shown in Fig. 6 for the Si dielectric constant and in Fig. 7 for the SiO_2 dielectric constant. In both cases, the dipole is taken to be in the center of the slab. The additional modes smooth out the radiation power as a function of slab width although the onset of each mode is clearly visible for both dielectric constants. For the higher dielectric constant (Fig. 6), the addition of the radiation into both the $l = 1$ odd TM mode (shoulder) and the $l = 2$ even TE mode (peak) increases the maximum radiated power for the horizontal dipole slightly above that of the vertical dipole. There are two curves presented in Fig. 6 for the horizontal dipole: the lower curve is the radiated power into the waveguide modes, and the higher curve (dotted), which is filled in just before the onset of each mode, includes the contributions of the power into the 3-D radiation modes out the sides of the slab. This 3-D radiation contribution is too small to be visible for the vertical dipole. As noted above in the discussion of the 3-D radiation, the peaks in the 3-D radiation occur at the valleys of the 2-D modal emission and tend to lead to a more uniform emission as a function of slab width. Mathematically, both of these effects are described by the multiple-reflection, Fabry–Perot denominators [cf. (10), (13), and (15)] that play a role in both the 3-D and 2-D calculations. At the lower dielectric constant (Fig. 7)

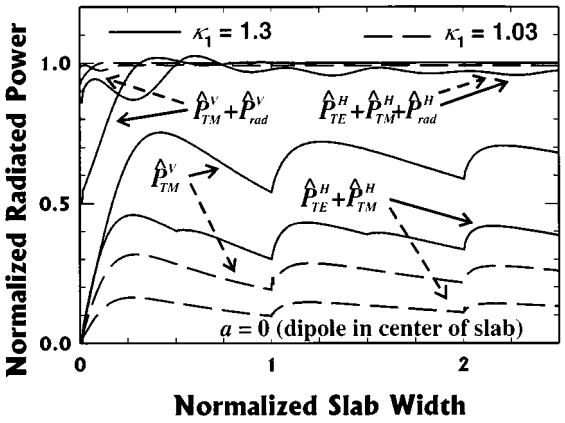


Fig. 8. The distribution of radiated power into all modes for κ_1's of 1.3 and 1.03 for a dipole located in the center of the slab. As the dielectric constants are decreased, an increasing fraction of the power goes into the 3-D radiation modes.

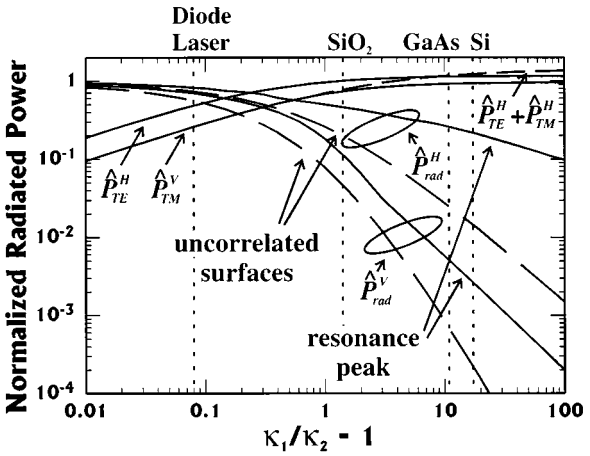


Fig. 9. Summary of the dielectric contrast dependence of the power radiated into both the 3-D radiation and 2-D waveguide modes. The slab width is allowed to vary as the dielectric constants are varied. The peak values of the 3-D radiation power at the first peak ($\sim\lambda n_2/2n_1$ for the vertical dipole and $\sim\lambda n_2/n_1$ for the horizontal dipole at high contrast) are displayed. The 2-D waveguide mode power is the peak of the zeroth-order mode (TM_0 for a vertical and TE_0 for a horizontal dipole). The maximum radiation for the horizontal dipole at high contrast shifts to the sum of the powers into TE_0, TE_2 and TM_1 as indicated by the dashed line giving a maximum enhancement in the radiated power of $\sim$30% over a dipole in an unbounded medium.

the contributions of the 3-D radiation are significant for both dipole orientations. The peak total radiation (sum of 2-D waveguide- and 3-D radiation-modes) is shifted away from the 2-D waveguide peak and is higher for the horizontal than for the vertical dipole. The contributions of the $l = 1$ odd TM mode for the horizontal dipole are clearly evident at half-integral normalized slab widths. These trends are continued for lower dielectric constants as shown in Fig. 8 for dielectric constants of 1.3 and 1.03, spanning the typical dielectric contrasts for heterostructure diode lasers. The lower the dielectric contrast, the larger the fraction of power radiated into the 3-D modes. For the dielectric constant of 1.3, the total radiated power for a vertical dipole is almost uniform at unity, i.e., equal to that in an unbounded medium, except for very thin slabs. There is somewhat more structure for a horizontal dipole and the total power radiated both peaks at a higher level and saturates at the slightly lower level than for the vertical dipole. For the 1.03 dielectric constant case, all of the structure is essentially eliminated. These curves were calculated for a unity cladding dielectric constant, identical curves are obtained for the actual semiconductor laser case of much larger cladding dielectric constants, but the same percentage difference between the slab and cladding dielectric constants.

Useful insight is gained by plotting the radiated power and its division between 3-D radiation and 2-D waveguide modes as a function of the dielectric discontinuity. Fig. 9 shows this dependence for lossless dielectrics. In each case, the peak radiated power is plotted for both horizontal and vertically oriented dipoles located at the center of a slab. The slab width is a free parameter that is allowed to vary for different dielectric discontinuities. The curves labeled vertical (horizontal) dipole—TM (TE) mode correspond to the largest values in the plots of 2-D radiated power versus slab width as shown in Figs. 5–8. As noted above, these values saturate for large dielectric discontinuities. The dashed curve labeled $\hat{P}^H_{\rm TE}+\hat{P}^H_{\rm TM}$ corresponds to the peak emission power for the horizontal dipole at large dielectric contrast when the sum of powers at for slab widths of about a normalized wavelength is larger than the zeroth order mode peak. The two dashed curves labeled uncorrelated surfaces are the power emitted into the 3-D radiation modes out the two sides of the slab neglecting any Fabry–Perot cavity effects, e.g., just twice the single surface result. As required, in the limit of small dielectric discontinuities, these powers approach unity. In the opposite limit of large dielectric discontinuity, these curves scale as m^5 for the vertical dipole and m^3 for the horizontal dipole. Finally, the two solid curves labeled resonance peak are the total emitted radiation at the first Fabry–Perot peak of the emission out both sides of the slab and include the effects of the slab resonances that are important in vertical cavity lasers. For low discontinuity, these approach unity, for high discontinuities they scale as m^3 (vertical) and m (horizontal). The slight bend in the curve for the 3-D radiation of the vertical dipole at $(\kappa_1/\kappa_2 - 1) \sim 3$ corresponds to the shift of the peak from even integral reduced wavelengths at small dielectric contrast to odd integral half wavelengths at high dielectric contrast (cf. Fig. 3). The dielectric discontinuities corresponding to important classes of semiconductor and glass dielectric constants are marked on the figure. These curves were calculated for $\kappa_2 = 1$ (e.g., air or vacuum). As noted above, for the case of a diode laser with semiconductor claddings equivalent results are obtained as long as the x axis is normalized as $\sqrt{\kappa_1 - \kappa_2}L/\lambda$; all of these results are universal curves as a function of relative dielectric discontinuity.

VI. Summary

An analysis of the radiation emitted from a dipole of arbitrary location and orientation in a symmetrically clad dielectric slab has been presented. The results for arbitrary dielectric contrasts are consistent with previous work [15] and are presented in terms of simple, closed-form equations for both the 3-D free-space emission into plane waves and the 2-D emission into waveguide modes of the slab. These results arise from a unified analysis based on the early work of Sommerfeld [1] for a dipole near a dielectric interface. It is interesting to find

that work that was motivated by the discovery of long-distance radiotelegraphy has fresh relevance to current microcavity and diode laser research.

The analytic results are evaluated for a wide range of dielectric contrasts ranging from those of free-standing semiconductor films to those of double-heterostructure diode lasers. Over this wide range, the distribution of power into the 3-D and 2-D modes varies widely. For large dielectric contrast, almost all of the radiated power is into the waveguide modes with only a small amount of power into the radiation modes. For a vertically oriented dipole at the center of the slab, the power into the radiated modes scales as m^5 for a single interface and as m^3 at the peak of the Fabry–Perot resonances. This vertical dipole emission is weak because of the small critical angle escape cone and the null in emission along the direction of the dipole. The peak emission, for a slab thickness close to the wavelength, saturates at 15% above that of a dipole in an unbounded dielectric medium. The emission into the 3-D radiation modes is much larger for a horizontally oriented dipole in a high index slab since the dipole emission is strongest in the direction of the slab normal, the power into the radiated modes scales as m^3 for a single interface and as m at the peak of the Fabry–Perot resonances. For this case, the peak power into the lowest order mode saturates at about 5% below the unbounded medium result; however, the maximum power into all waveguide modes occurs for slab widths of about a reduced wavelength where power is radiated into all three of the zeroth- and second-order (even) TE modes and the first-order (odd) TM mode. At this peak, the radiated power saturates at a level $\sim$30% larger than that in the unbounded medium. Even for very small dielectric discontinuities, an enhancement of the total (2-D and 3-D) horizontal dipole radiated power persists for slab thicknesses of about a reduced wavelength.

These results have application to a wide variety of topics of current interest including:

- to crystal membrane microresonantors using dielectric confinement in one direction and photonic crystal confinement within the plane of the membrane;
- to enhanced LEDs where the object is to couple more of the emitted radiation into the useful 3-D modes and less into the trapped waveguide modes;
- to traditional edge-emitting lasers where the radiation into the waveguide modes can support lasing while the power into the 3-D modes represents lost excitation;
- to VCSELs where just the opposite comments apply;
- and to nanostructure enhanced Raman scattering.

For the Raman experiment, three distinct analyzes are required: 1) the distribution of the excitation for a plane wave incident on a deep, periodic grating structure; 2) the emission into waveguide and radiation modes (this analysis); and 3) the coupling of the waveguide modes into free-space modes at the top of the nanostructured grating. Extensions of this work to more complex dielectric structures such as VCSELs and periodic gratings, and to more confined structures such as photonic wires and dots will be discussed in ongoing work.

ACKNOWLEDGMENT

The authors would like to thank Dr. P. G. Eliseev and Dr. P. Pulaski for extensive, very useful discussions.

REFERENCES

[1] A. Sommerfeld, "Über die Ausbreitung der Wellen in der drahtlosen Telegraphie," *Ann. Physik*, vol. 28, pp. 665–737, 1909.
[2] A. Baños, *Dipole Radiation in the Presence of a Conducting Half-Space*. Oxford: Pergammon Press, 1966.
[3] W. Lukosz, "Light emission by magnetic and electric dipoles close to a plane interface—III: Radiation patterns of dipoles with arbitrary orientation," *J. Opt. Soc. Amer.*, vol. 69, pp. 1495–1503, 1979.
[4] D. Marcuse, *Theory of Dielectric Optical Waveguides*. San Diego: Academic Press, 1991.
[5] R. R. Chance, A. Prock, and R. Silbey, "Molecular fluorescence and energy transfer near interfaces," in *Adv. in Chem. Phys. XXXVII*, I. Prigogine and S. A. Rice, Eds. New York: Wiley, 1978.
[6] D. G. Deppe, C. Lei, C. C. Lin, and D. L. Huffaker, "Spontaneous emission from planar microstructures," *J. Mod. Opt.*, vol. 41, p. 325, 1994.
[7] R. Windisch, P. Heremans, A. Knobloch, P. Keisel, G. H. Dohler, B. Dutta, and G. Borghs, "Light-emitting diodes with 31% external quantum efficiency by outcoupling of lateral waveguide modes," *Appl. Phys. Lett.*, vol. 74, pp. 2256–2258, 1999.
[8] A. Scherer, O. Painter, B. D'Urso, R. K. Lee, and A. Yariv, "In GaAsP photonic band gap crystal membrane microresonator," *J. Vac. Sci. Technol.*, vol. B16, pp. 3906–3910, 1998.
[9] G. Björk, S. Machida, Y. Yamamoto, and K. Igeta, "Modification of spontaneous emission rate in planar dielectric microcavity structures," *Phys. Rev.*, vol. A44, pp. 669–681, 1991.
[10] W. Zakowicz and A. Bledowski, "Spontaneous emission in the presence of a dielectric slab," *Phys. Rev.*, vol. A52, pp. 1640–1650, 1995.
[11] H. Rigneault and S. Monneret, "Modal analysis of spontaneous emission in a planar microcavity," *Phys. Rev.*, vol. A54, pp. 2356–2368, 1996.
[12] H. Benisty, H. De Neve, and C. Weisbuch, "Impact of planar microcavity effects on light extraction—Part I: Basic concepts and analytical trends," *IEEE J. Quantum Electron.*, vol. 34, pp. 1612–1631, 1998.
[13] ——, "Impact of planar microcavity effects on light extraction—Part II: Selected exact simulations and the role of photon recycling," *IEEE J. Quantum Electron.*, vol. 34, pp. 1632–1643, 1998.
[14] Y. Xu, R. K. Lee, and A. Yariv, "Finite-difference time-domain analysis of spontaneous emission in a microdisk cavity," *Phys. Rev.*, vol. A61, 2000.
[15] S.-T. Ho, D. Yen Chu, J.-P. Zhang, S. Wu, and M.-K. Chin, "Dielectric photonic wells and wires and spontaneous emission coupling efficiency of microdisk and photonic-wire semiconductor lasers," in *Optical Processes in Microcavities*, R. K. Chang and A. J. Campillo, Eds. Singapore: World Scientific, 1996, pp. 339–387.
[16] S. H. Zaidi, A.-S. Chu, and S. R. J. Brueck, "Optical properties of nanoscale, one-dimensional silicon grating structures," *J. Appl. Phys.*, vol. 80, pp. 6997–7008, 1996.
[17] J. A. Stratton, *Electromagnetic Theory*, 1st ed. New York: McGraw Hill, 1941, pp. 573–575.
[18] W. Lukosz and R. E. Kunz, "Light emission by electric and magnetic dipoles close to a plane interface—II: Radiation patterns of perpendicular dipoles," *J. Opt. Soc. Amer.*, vol. 67, pp. 1607–1615, 1977.



S. R. J. Brueck (S'63–M'71–SM'89–F'93) was born in New York City in 1944. He received the B.S. degree from Columbia University, New York, NY, in 1965 and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, MA, in 1967 and 1971, respectively, all in electrical engineering.

From 1971 to 1985, he was a Staff Member at MIT Lincoln Laboratory. In 1985, he moved to the University of New Mexico, Albuquerque, where he is currently a Professor in the Electrical and Computer Engineering and Physics and Astronomy Departments and Director of the Center for High Technology Materials. His current research interests include nanoscale lithography, the physics of nanostructures, and nonlinear optics of poled glasses.

Metallic inductive and capacitive grids: Theory and Experiment

B. K. Minhas, W. Fan, K. Agi, S. R .J. Brueck and K. J. Malloy
Center for High Technology Materials, 1313 Goddard SE,
Albuquerque, New Mexico 87106

We present experimental validation of theoretical modeling of both capacitive (dot) and inductive (hole) metallic crossed gratings in the mid-infrared (2-5 μm). The gratings are fabricated using interferometric lithography and modeled using rigorous coupled wave analysis. The transmittance spectra of these gratings display the coupling of incident light into surface plasma waves, and our experimental and numerical investigations suggest that surface plasma wave coupling also describes the behavior of capacitive grids.

Submitted to Journal of the Optical Society of America B

I. Introduction

Crossed gratings, of which metallic grids are special cases, find application across the electromagnetic spectrum. They are used in the microwave regime as filters and for making efficient antennas[1]. In the far infrared, they find applications as beam splitters and mirrors[2]. Their use as filters[3,4] and solar selective surfaces[5] in the visible and near infrared has also been demonstrated. Recently there has been interest in the behavior of sub-wavelength aperture metallic crossed gratings[6] (inductive or hole grids in particular) in the visible with potential applications for novel optoelectronic devices. In that work, the important features of the transmission spectra of sub-wavelength inductive grids were characterized on the basis of a wavevector conservation model. The authors, adopting a 'kinematic' approach, showed that the salient spectral features of the experimentally measured transmittance correspond to the coupling of incident radiation to surface plasma waves (SPW)[6]. Although this method explains the wavelength position of the experimental features, a rigorous diffraction model is necessary to quantitatively describe all experimental features and as a tool for crossed grating design. The importance of this approach is underscored by the recent work of Heinzel et al.[7], Martin-Moreno et al.[8] and Popov et al.,[9] where the behavior of metallic, crossed gratings was analyzed using vector diffraction models. Reference [7] presented experimental results and numerical modeling (based on rigorous coupled wave analysis (RCWA)[10]) for periodically structured metal surfaces with potential applications as filters in thermophotovoltaic systems. Martin-Moreno et al.[8] and Popov et al.[9] specifically examined the behavior of metallic, subwavelength hole arrays. Reference [8] treated the electromagnetic fields in the metallic inductive grids using surface impedance boundary conditions, while Ref. [9] employed the Fourier modal method[11]. Both these papers analyzed the experimental work of Ebbesen et al.,[6] and attempted to explain the mechanism of unusually high transmittance through metallic, subwavelength aperture inductive grids. However, to the best of our knowledge, rigorous numerical modeling of finite conductivity capacitive grids along with experimental verification has not yet been presented (McPhedran et al.[12] and Botten et al.[13] have numerically modeled perfectly conducting capacitive grids using a rigorous modal approach).

In this paper we present a rigorous diffraction model validated by experimental data for both inductive (hole) and capacitive (dot) sub-wavelength, metallic, crossed gratings under normal plane wave illumination. We present the effects of varying grating parameters on the mid-IR transmittance spectrum. The paper is organized as follows. First, we briefly explain the fabrication and experimental characterization of crossed gratings. This is followed by a description of the numerical model. Then, we discuss the coupling of incident radiation into SPWs for inductive grids and show that similar coupling also occurs for capacitive grids. Coupling in SPWs for both kinds of gratings are described using the rigorous diffraction model. Finally, the experimental verification of the diffraction model is demonstrated by modeling gratings with varying parameters.

II. Fabrication and experimental characterization

The gratings used to validate the modeling were fabricated using interferometric lithography and lift-off pattern transfer. A double polished silicon sample is used as a substrate material as it is transparent in the 2–5 μm wavelength regime. After coating the substrate with photoresist, interferometric lithography[14] is used to expose the pattern. Developing the photoresist and metallization by e-beam evaporation of gold follows. The final step is an acetone jet lift-off leaving the desired pattern.

Experimental characterization of these structures is done using Fourier transform infrared spectroscopy (FTIR) with an unpolarized light source. All of the results presented in this paper are normal incidence transmittance spectra normalized to the system response without any sample.

III. Modeling

Our numerical model takes into account the following important aspects of diffraction from crossed gratings.

III.a: Finite conductivity of the gratings

Crossed gratings are modeled in the microwave regime by taking the metallic scatterers to have infinite conductivity[15-17]. While this approximation works well at microwave frequencies where metals have very high conductivity, its use for the visible and ultraviolet spectral region is clearly inappropriate. In the infrared spectral region, the suitability of the infinite conductivity approximation is uncertain. We avoid these questions by using a rigorous diffraction model, which takes into account the finite conductivity of the gratings. This approach comes at a high computational cost.

Since the late 1970s, significant work has been done in modeling crossed gratings with finite conductivity. Derrick et al.[18,19] and Harris et al.[20] proposed a model based on a coordinate transformation, Vincent[21] used a finite difference method, while Bruno et al.[22] employed a variation of boundaries method. More recently, Kettunen et al.,[23] presented numerical results for metallic inductive grids based on the rigorous coupled wave analysis (RCWA). Their modeling was done principally for the case of rectangular scatterers and included the dielectric function expansion changes suggested by Li[11] to improve the convergence of the algorithm. Here, we present modeling based on RCWA for finite conductivity rectangular as well as circular metallic scatterers, arranged in a square lattice on a silicon substrate.

b: Finite substrate effects

The results presented in this paper are for the normal incidence transmittance spectra of metallic crossed gratings placed on a silicon substrate. The spectral range of interest is from 2-5 μm, where silicon is practically transparent. Clearly, in this regime assuming the silicon substrate to be infinite (as is commonly done in grating diffraction modeling) is not justified due to multiple reflections taking place within the high-index substrate. The thickness of the substrate is much larger than the coherence length of the incident light for the chosen experimental conditions, therefore the appropriate way to model the substrate is to do incoherent addition of the diffracted orders inside the substrate.

Previous work based on this approach has been done for one-dimensional gratings by Chateau et al.[24] and Li[25]. Here we extend the energy scattering matrix approach used by Li[25] to model diffraction from a crossed gratings made on a thick transparent substrate.

In order to further study the usefulness of incorporating finite substrate effects, we model the transmittance spectra of capacitive grids for the 2-5 μm wavelength range using both the infinite substrate model (ISM) and the finite substrate model (FSM). Our numerical simulations show that while the FSM is necessary to quantitatively compare with the experimental measurements, the ISM captures the essential features of the transmittance spectra.

In this paper, because of computational constraints, we use the simpler ISM to numerically study the behavior of crossed gratings and use the FSM whenever we compare the numerical modeling with the experimental measurements.

c: Convergence of the numerical model

Our numerical model is based on RCWA, and we study the convergence behavior of this algorithm for the ISM and FSM.

For the infinite substrate case (ISM), we take the silicon substrate to be infinitely thick, and for normal plane wave illumination, take advantage of the symmetry of the configuration[26] to reduce the numerical size of the problem. Specifically for normal incidence, the equivalence of the positive and negative Fourier coefficients allows reduction of the final eigenvalue problem from $2(2N+1)^2$ to $2(N+1)^2$, where N is the number of positive or negative spatial harmonics for both the dielectric function and the field expansion. This reduction allows the algorithm to be implemented on a personal computer for reasonably high values of N.

Validating the performance of any numerical algorithm is an essential part of establishing the utility of the algorithm. Unlike the situation for one-dimensional gratings where the convergence properties of RCWA for metallic gratings have been well established[27,28], little is known of the behavior of two-dimensional RCWA algorithms. Initially, we examined the numerically calculated diffracted power in the transmitted zero order of capacitive grids as a function of the mode number N up to the maximum number of modes as allowed by our computational resources. Numerical evaluations show that the variation in diffracted power as a function of mode number N is at the most ±5% after N =15. In the results shown in this paper we have used N = 15 for numerical modeling and expect the convergence error to behave similarly for all structures. However, the final proof of the validity of the two-dimensional RCWA lies in the experimental verification presented later in this paper.

For the FSM, we include the finite thickness of the silicon substrate. An efficient way to implement this is to use a scattering matrix algorithm[29]. Here we use the full scattering matrix recursion[30] to account for all propagating orders incident on the grating from the substrate side. Depending on the grating parameters, the diffraction orders propagating in the substrate are obliquely incident on the grating from the substrate side. The grating configuration in this case is no longer symmetric and unlike the ISM, we cannot reduce the numerical size of the eigenvalue problem. Therefore for the FSM we restrict the mode number to N =10, and expect the convergence error to be within ±7%.

IV. Surface plasma waves and crossed gratings

It has been experimentally[6] demonstrated that inductive grids, or metal films perforated with subwavelength hole arrays, have a higher transmittance than that predicted by single aperture theory[31]. Transmittance efficiencies showing a 1000-fold increase[31] are observed in subwavelength inductive grids when compared with the results of single aperture theory. This phenomenon is attributed to the resonant coupling of the incident light to the SPW. The wavelength at which SPW coupling occurs is described by wavevector conservation, which is essentially a 'kinematic' model. This model, which gives the in-plane momentum necessary to

couple to SPWs, explains the wavelength dependent features of the experimental data. However, a detailed analysis of the magnitude and variation of the coupling requires a rigorous diffraction model. We apply both the 'kinematic' model and RCWA to study metallic crossed gratings.

To illustrate, consider first the surface plasma wave equation that gives the momentum conservation condition necessary to couple normally incident radiation to an infinite metal sheet[32]

$$k_{\parallel} = \frac{2p}{l}\,\mathrm{Re}\left(\sqrt{\frac{e_m e_d}{e_m + e_d}}\right) \tag{1}$$

Here $k_{\parallel}$ is the projection of the incident wave vector parallel to the dielectric-metal interface, λ is wavelength of incident light and ε_m and ε_d are the complex permittivities of the metal and the incident medium respectively. A metal film periodically perforated with holes can be considered as a perturbation of smooth metal film, and Eq. (1) may be used to determine the coupling to surface plasma waves for inductive grids. For the special case of a square lattice and normally incident radiation, $k_{\parallel}$ is given as $\frac{2p}{\Lambda}\sqrt{i^2 + j^2}$, where Λ is the pitch of the grating and i, j are integers (at least one of which is nonzero). Substituting into Eq. (1), we get [31]

$$\frac{l_{i,j}}{\Lambda} = \frac{1}{\sqrt{\left(i^2 + j^2\right)}}\,\mathrm{Re}\left(\sqrt{\frac{e_m e_d}{\left(e_m + e_d\right)}}\right). \tag{2}$$

The dielectric medium in this equation is either the superstrate (assumed to be air) or the substrate (silicon). The air-metal interface and the substrate-metal interface will each provide a set of (i, j) for Eq. (2) corresponding to air-metal SPW coupling and substrate-metal SPW coupling[33].

A phenomenon closely associated with the appearance of SPW coupling in metallic gratings is known as Wood's anomaly[34]. This is the condition where the wavevector of a given grating order becomes tangent to the grating plane, and is marked by rapid variations in the intensity of the remaining orders. The wavelength condition for Wood's anomalies for normally incident illumination on a crossed grating is given as [31]

$$\frac{l_{i,j}}{\Lambda} = \sqrt{\frac{e_d}{\left(i^2 + j^2\right)}}. \tag{3}$$

Here ε_d is the permitivity of the superstrate or substrate, and accordingly, the air and silicon interfaces each give a set of Wood's anomalies. From Eqs. (2) and (3), it may be noted that for metallic gratings where $|\varepsilon_m| >> \varepsilon_d$, the wavelength conditions for SPW coupling and Wood's anomalies are very close.

Now, let us consider the case of a patterned gold film on a silicon substrate with a pitch of 1.2 μm. Equation (2) predicts coupling to the (0, ±1) or (±1, 0) substrate-metal SPW at λ = 4.16 μm. The results of the RCWA model for inductive grids are presented in . 1, which shows the transmittance and absorptance of a grid with a thickness of 0.1 μm and a square air aperture of 0.36 μm^2. The absorptance A is simply given by

$$A = 1 - \sum_d R_d - \sum_d T_d\,, \tag{4}$$

where R_d and T_d are the diffraction efficiencies for order 'd' in the incident and transmitted media respectively. The modeling presented in this figure takes the substrate to be infinite, includes the dispersion of dielectric function of gold[35] and takes the index of silicon to be 3.44 for 2-5 μm wavelength range. Also, the transmittance is defined to be the sum of diffraction efficiencies in all the transmitted orders. Figure 1 also shows the absorptance of a homogeneous 0.1 μm thick gold film on a silicon substrate; the difference between the absorptance spectra of the inductive grid and a gold film is indicative of the power carried by SPWs. From Fig. 1 it may be seen that the SPW wavelength as calculated from Eq. (2) matches the rigorous diffraction model very well although the exact wavelengths of SPW coupling are offset slightly from those given by Eq. (2), (e.g., for the (0, 1) substrate-metal coupling the transmittance reaches a maximum at λ = 4.28 μm instead of λ = 4.16 μm). Note that we take the transmittance maximum as the SPW coupling wavelength[33]. A possible reason for this wavelength difference as obtained from wavevector conservation (Eq. (2)) and RCWA is the fact that Eq. (2) is a perturbation on the SPW dispersion relation for homogenous metal films and therefore can only describe the SPW coupling wavelength approximately. Also shown in Fig. 1 are examples of the

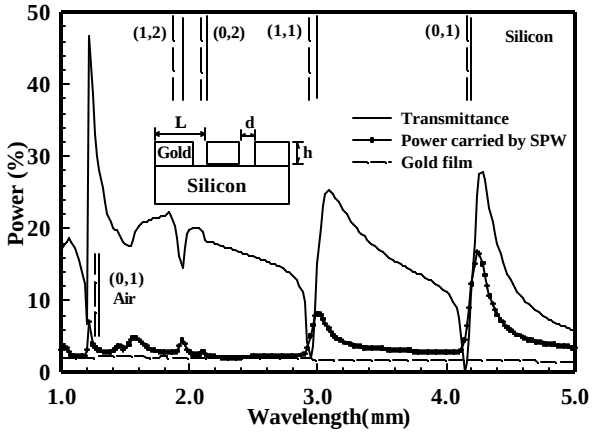


Figure 1: Transmittance and SPW coupling for an inductive grid with square air holes in gold on a silicon substrate. Referring to the inset, Λ = 1.2 μm, d = 0.6 μm, and h = 0.1 μm. Wavelength resolution for the numerical modeling is 0.02 μm and index of silicon is taken as 3.44, independent of wavelength. The locations of SPW coupling for air-metal and silicon-metal are obtained from Eq. (2) and are labeled by solid vertical line segments. The locations of Wood's anomalies are obtained from Eq. (3) and are labeled as dashed vertical line segments. Also shown in the figure is the absorptance of a 0.1 μm thick gold film on a silicon substrate.

air-metal SPW (denoted by solid vertical lines), substrate-metal SPW (shown by solid vertical lines) and Wood's anomalies for the incident and transmitted media (denoted by dashed vertical lines). For the case of inductive grids, we generally observe well-defined transmittance minima and maxima corresponding to the Wood's anomalies and SPW coupling, respectively[33].

We now study the applicability of the 'kinematic' model to capacitive grids. Clearly some additional evidence is necessary to apply Eq. (2), as unlike the inductive grids where a continuous metal film periodically perforated with holes allows for extended conduction, the capacitive grids are metal islands having no conduction between them. It remains to be shown that SPWs are established in capacitive grids.

One of the important features of SPWs is the energy carried by these surface waves near the coupling wavelengths. This is evident for inductive grids in Fig. 1, where we observe maxima in the absorptance spectrum near the location of SPW coupling. Fig. 1 also shows that the energy carried by the SPW occurs in a narrow spectral band. Furthermore, the features associated with Wood's anomalies and SPWs coupling affect the transmittance spectrum differently; Wood's anomalies occur as minima in transmittance while SPW coupling occur as maxima in transmittance. Also note that the wavelengths of the Wood's anomalies (as defined by Eq. (3)) and of the SPW coupling (as defined by Eq. (2)) are almost superimposed. We now examine the absorptance spectrum of capacitive grids. Figure 2 shows the transmittance and absorptance for a capacitive grid having a pitch of Λ = 1.2 μm, a grating thickness of 0.1 μm with square gold

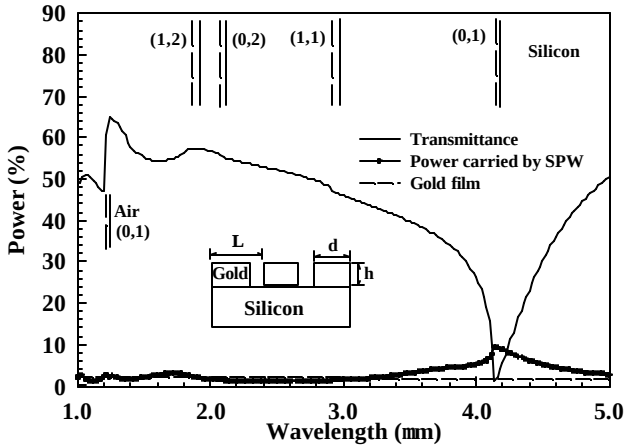


Figure2: Transmittance and SPW coupling, for a capacitive grid with square gold scatterers, on a silicon substrate. Referring to the inset, Λ = 1.2 μm, d = 0.6 μm, and h = 0.1 μm. The wavelength resolution of numerical modeling is 0.02 μm and index of silicon is taken as 3.44 for all wavelengths. The locations of SPW coupling for air-metal and silicon-metal are obtained from Eq. (2) and are labeled by solid vertical line segments. The locations of Wood's anomalies are obtained from Eq. (3) and are labeled as dashed vertical line segments. Apart from the (0, 1) substrate-metal SPW, coupling to higher order SPWs is very weak for this particular capacitive grid and is only visible as discontinuity in the slope of transmittance spectrum. Also shown is the absorptance of 0.1 μm thick gold film on a silicon substrate

scatterers of area 0.36 μm^2. From the absorptance curve in Fig. 2, we observe SPW-like features near the (0, 1) substrate-metal coupling; also in contrast to the previous inductive grid, the energy carried by these plasma-like waves is spread over a wider spectral band, which may be indicative of high damping of these plasma-like waves. The transmittance spectrum of this capacitive grid, unlike that of the inductive grid, does not show well-defined minima and maxima near the locations of the Wood's anomalies and of SPWs coupling, respectively. Instead, the features associated with Wood's anomalies and SPWs influence the transmittance spectrum in a similar manner; they are characterized either by a minimum in the transmittance spectrum (as shown in Fig. 2 for (0,1) substrate-metal coupling) or by points of inflection in the transmittance spectrum (as shown in Fig. 2 for higher-order substrate-metal coupling). The absorptance spectrum of the capacitive grid shows a maximum that suggests its behavior may be described in terms of SPW coupling.

V. Capacitive and inductive grids

From Figs. 1 and 2, we observe that surface plasma wave features occur for both capacitive and inductive grids. We experimentally verify our modeling by fabricating inductive and capacitive grids of the same pitch and grating thickness. The transmittance spectra for these structures are shown in Figs. 3 and 4 along with the results of numerical computations. Here in the numerical model we are taking into account the finite thickness of the silicon substrate (FSM). The FSM has a larger convergence error than the ISM; furthermore the size of the circular scatterers is determined by scanning electron microscopy (SEM), which also introduces some measurement error. In order to account for these factors, we carry out a mean square fit between the experimental and numerical data. While the numerical model assumes illumination with a normally incident, linearly polarized plane wave, experimental characterization of inductive and capacitive grids is done using an unpolarized source in the FTIR. The transmittance or reflectance spectra of square lattices under normal plane wave illumination are polarization independent[2].

Figures 3 and 4 show that the experiment and numerical model are in reasonably good agreement. Furthermore, the inductive and capacitive grids display surface plasma wave features at approximately the same wavelengths, but the strength and shape of surface plasma wave coupling is very different for each of them. For inductive grids the amount of light transmitted away from the SPW resonance is small and changes in transmittance spectrum are sharp and narrow at the coupling wavelengths, whereas in capacitive grids, we note a gradual change in transmittance near the wavelengths of SPW coupling. One of the reasons is a dependence on the grid fill factor for capacitive and inductive grids, FF_C and FF_I, defined as

$$FF_C = \frac{p}{4}\left(\frac{d}{\Lambda}\right)^2$$

for circular scatterers in capacitive grids and as

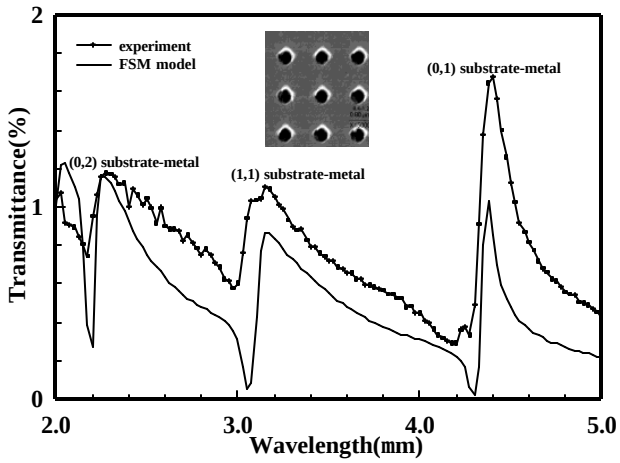


Figure 3: A comparison between experimental and rigorous modeling results for transmission through an inductive grid of circular air holes in gold on a silicon substrate. Here $\Lambda = 1.24$ μm, h = 0.1 μm and the air hole diameter obtained from both SEM measurement and mean square fitting of the numerical model to the data is 0.40 μm. The wavelength resolution of the numerical modeling is 0.025 μm.

$$FF_I = 1 - \frac{p}{4}\left(\frac{d}{\Lambda}\right)^2$$

for circular air holes in an inductive grid. Here d is the diameter of circular scatterers for capacitive grids or of the air hole for inductive grids. The structures shown in Figs. 3 and 4 have a fill factor of 91.8% for the inductive grid and 29.5% for the capacitive grid. Also, for this particular inductive grid (shown in Fig. 3), the absolute efficiency of transmittance (obtained by dividing the fraction of light transmitted to fraction of the surface area occupied by holes[6]) at positions of maxima is ~0.25. This is much less than that

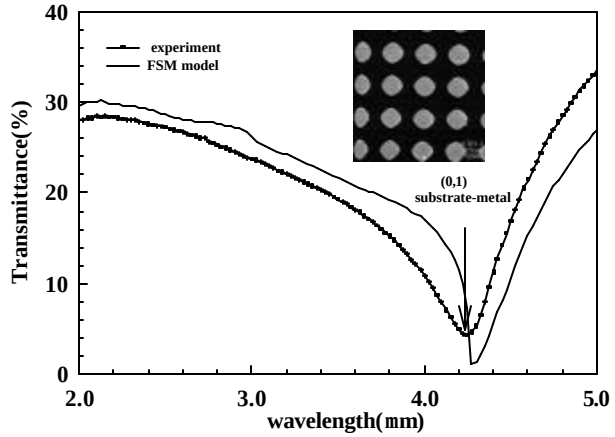


Figure 4: Experimental and rigorous modeling results for transmittance through a capacitive grid of circular gold scatterers on a silicon substrate. Here $\Lambda = 1.24$ μm, h = 0.1 μm and diameter of the gold scatterers obtained from SEM measurement is 0.76 μm and from mean square fitting of the numerical model to experimental data is 0.75 μm. The wavelength resolution for numerical model is 0.025μm.

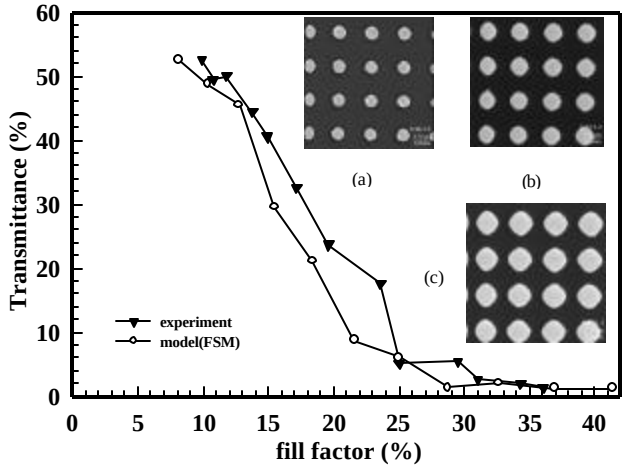


Figure 5: The experimental and numerically calculated dependence of transmittance through a capacitive grid with fill factor. The transmittance is shown for the case of (0, 1) substrate-metal SPW coupling at $\lambda = 4.3$ μm. Referring to the inset in Fig. 2, the fill factor for a capacitive grid is given as $(\pi/4)(d/\Lambda)^2$. The grid consisted of circular gold scatterers on a silicon substrate, had a pitch of 1.24 μm and a thickness of 0.1μm. Note the 'noise' in numerically calculated transmittance due to convergence error in the FSM. The insets in the figure also show SEM images of capacitive grids with fill factors of a) 10.8%, b) 19.6% and c) 34.3%.

observed by Ebbesen et. al.[6], for the case of subwavelength hole arrays in silver films made on a quartz substrate. This apparent discrepancy is a candidate for further study, although one of the reasons may be the high-index silicon substrate used in this study. For the capacitive grid (shown in Fig. 4), it may be noted the wavelength location of SPW coupling is offset between the experimental data and numerical modeling; this difference is attributed to convergence error in the FSM.

VI. Fill factor study

We experimentally and numerically investigated the dependence of the (0, 1) substrate–metal SPW coupling as a function of the fill factor for capacitive grids. All samples were 0.1 μm thick circular gold scatters on a silicon substrate and with a pitch of 1.24 μm. Figure 5 shows the dependence of the experimental and numerical transmittance at $\lambda = 4.3$ μm on fill factor. For the modeling, the fill factor is varied by changing the diameter of circular gold scatterers from 0.4 μm to 0.9 μm in steps of 0.05 μm. Numerical modeling for capacitive grids of different fill factors is done using the FSM and a mean square fit was not performed for the modeled and experimental measurements. SEM measurements are done to determine the size of the circular gold scatterers and the transmittance spectra are obtained using FTIR. We note from Fig. 5 that the convergence error in the FSM appears as discontinuous variations in the slope of the numerically calculated transmittance. Nevertheless, the model and experiment display the same trend, specifically, transmittance through the grid at the wavelength of the (0, 1) substrate-metal SPW coupling ($\lambda = 4.3$ μm in this case) decreases monotonically with fill factor.

VII. Conclusion

We have experimentally validated numerical results based on RCWA for metallic inductive and capacitive grids in the mid-IR. We find that the features in the transmittance spectrum of both inductive and capacitive grids are defined by coupling to surface plasma waves and the RCWA algorithm adequately describes both the position and magnitude of these features. A systematic study of the behavior of capacitive grids as a function of fill factor showed that transmittance decreases monotonically with fill factor at the wavelength of SPW coupling.

VIII. Acknowledgements

This research was supported by DARPA and by ARO/MURI. One of the authors B. Minhas will like to thank Dr. Lifeng Li of Tsinghua University (China) and Dr. Pasi Vahimaa of University of Joensuu (Finland) for help with the debugging stage of the diffraction code.

XI. References

[1]T. K. Wu, "Frequency Selective Surface and Grid Array," in *Wiley series in Microwave and Optical engineering* (John Wiley & Sons, New York, 1995).

[2]R. Ulrich, "Far-Infrared properties of metallic mesh and its complementary structure," Infrared Physics 7, 37-55 (1967).

[3]Song Peng and G. Michael Morris, "Experimental demonstration of resonant anomalies in diffraction from two-dimensional gratings," Optics Letters 21 (8), 549-551 (1996).

[4]Song Peng and G. Michael Morris, "Resonant scattering from two-dimensional gratings," Journal of the Optical Society of America A 13 (5), 993-1005 (1996).

[5]C. M. Horwitz, "A new solar selective surface," Optics Communications 11 (2), 210-212 (1974).

[6]T. W. Ebbesen and H. J. Lezec and H. F. Ghaemi and T. Thio and P. A. Wolff, "Extraordinary optical transmission through sub-wavelength hole arrays," Nature 391, 667-669 (1998).

[7]A. Heinzel and V. Boerner and A. Gombert and B. Blasi and V. Wittwer and J. Luther, "Radiation filters and emitters for the NIR based on periodically structured metal surfaces," Journal of Modern Optics 47 (13), 2399-2419 (2000).

[8]L. Martin-Moreno and F. J. Garcia-Vidal and H. J. Lezec and K. M. Pellerin and T. Thio and J. B. Pendry and T. W. Ebbesen, "Theory of Extraordinary Optical Transmission through Subwavelength Hole Arrays," Physical Review Letters 86 (6), 1114-1117 (2001).

[9]E. Popov and M. Neviere and S. Enoch and R. Reinisch, "Theory of light transmission through subwavelength periodic hole arrays," Physical Review B 62 (23), 16100-16108 (2000).

[10]M. G. Moharam and Eric B. Grann and Drew A. Pommet, "Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings," Journal of the Optical Society of America A 12 (5), 1068-1076 (1995).

[11]Lifeng Li, "New formulation of the Fourier modal method for crossed surface-relief gratings," Journal of the Optical Society of America A 14 (10), 2758-2767 (1997).

[12]R. C. McPhedran and D. H. Dawes and L. C. Botten and N. A. Nicorovici, "On-axis diffraction by perfectly conducting capacitive grids," Journal of Electromagnetic Waves and Applications 10 (8), 1085-1111 (1996).

[13]L. C. Botten and R. C. McPhedran and N. A. Nicorovici and A. B. Movchan, "Off-axis diffraction by perfectly conducting capacitive grids: modal formulation and verification," Journal of Electromagnetic Waves and Applications 12 (7), 847-882 (1998).

[14]Xiaolan Chen and Saleem H. Zaidi and S. R. J. Brueck, "Interferometric lithography of sub-micrometer sparse hole arrays for field-emission display applications," Journal of Vacuum Science and Technology B 14 (5), 3339-3349 (1996).

[15]Chao-Chun Chen, "Transmission through a conducting screen perforated periodically with apertures," IEEE Transactions on Microwave Theory and Techniques 18 (9), 627-632 (1970).

[16]Raj Mittra and Chi H. Chan and Tom Cwik, "Techniques for Analyzing Frequency Selective Surfaces - A Review," Proceedings of the IEEE 76 (12), 1593-1615 (1988).

[17]R. C. McPhedran and D. Maystre, "On the Theory and Solar Application of Inductive Grids," Applied Physics 14 (1), 1-20 (1977).

[18]G. H. Derrick and R. C. McPhedran and D. Maystre and M. Neviere, "Crossed Gratings: A Theory and its Applications," Applied Physics 18, 39-52 (1979).

[19]R. Petit, "Electromagnetic theory of gratings," in *Topics in current physics* (Springer-Verlag, Berlin; New York, 1980), Vol. 22.

[20]J. B. Harris and T. W. Preist and J. R. Sambles and R. N. Thorpe and R. A. Watts, "Optical response of bigratings," Journal of the Optical Society of America A 13 (10), 2041-2049 (1996).

[21]P. Vincent, "A finite-difference method for dielectric and conducting crossed gratings," Optics Communications 26 (3), 293-296 (1978).

[22]Oscar P. Bruno and Fernando Reitich, "Numerical solution of diffraction problems: a method of variation of boundaries . III. Doubly periodic gratings," Journal of the Optical Society of America A 10 (12), 2551-2562 (1993).

[23]Ville Kettunen and Markku Kuittinen and Jari Turunen and Pasi Vahimaa, "Spectral filtering with finitely conducting inductive grids," Journal of the Optical Society of America A 15 (10), 2783-2785 (1998).

[24]N. Chateau and J. P. Hugonin and B. Guldimann and P. Chavel, "Two-wave diffraction of quasi-monochromatic light by a volume grating deposited on a thick transparent plate," Optics Communications 103 (5-6), 444-452 (1993).

[25]Lifeng Li, "Calculation of diffraction efficiencies of a grating made on a thick transparent plate," Optics Communications 160 (1-3), 15-21 (1999).

[26]Philippe Lalanne and Dominique Lemercier-Lalanne, "On the effective medium theory of subwavelength periodic structures," Journal of Modern Optics 43 (10), 2063-2085 (1996).

[27]Lifeng Li and Charles W. Haggans, "Convergence of the coupled-wave method for metallic lamellar diffraction gratings," Journal of the Optical Society of America A 10 (6), 1184-1189 (1993).

[28]P. Lalanne and G. Michael Morris, "Highly improved convergence of the coupled-wave method for TM polarization," Journal of the Optical Society of America A 13 (4), 779-784 (1996).

[29]Nicolas Chateau and Jean-Paul Hugonin, "Algorithm for the rigorous coupled-wave analysis of grating diffraction," Journal of the Optical Society of America A 11 (4), 1321-1331 (1994).

[30]Lifeng Li, "Formulation and comparison of two recursive matrix algorithms for modeling layered diffraction gratings," Journal of the Optical Society of America A 13 (5), 1024-1035 (1996).

[31]Tae Jin Kim and Tineke Thio and T. W. Ebbesen and D.E. Grupp and H. J. Lezec, "Control of optical transmission through

metals perforated with subwavelength hole arrays," Optics Letters 24 (4), 256-258 (1999).

[32]H. Raether, *Surface Plasmons on smooth and rough surfaces and on gratings* (Springer -Verlag, Berlin; New York, 1988).

[33]Tineke Thio and H. F. Ghaemi and H. J. Lezec and P. A. Wolff and T. W. Ebbesen, "Surface-plasmon-enhanced transmission through hole arrays in Cr films," Journal of the Optical Society of America B 16 (10), 1743-1748 (1999).

[34]R. W. Wood, "On a Remarkable Case of Uneven Distribution of Light in a Diffraction Grating Spectrum," Philosophical Magazine 4, 396-402 (1902).

[35]Edward D. Palik, "Handbook of optical constants of solids," in *Academic Press handbook series* (Academic Press, Orlando, 1985).

Optical and Electrical Properties of Nanostructured Metal–Silicon–Metal Photodetectors

A. K. Sharma, *Fellow, IEEE,* S. H. Zaidi, P. C. Logofătu, and S. R. J. Brueck, *Fellow, IEEE*

***Abstract*—We report an experimental evaluation of the performance of silicon (Si) photodetectors incorporating one-dimensional (1-D) arrays of rectangular and triangular-shaped nanoscale structures within their active regions. A significant ($\sim 2\times$) enhancement in photoresponse is achieved in these devices across the 400- to 900-nm spectral region due to the modification of optical absorption properties that results from structuring the Si surface on physical optics scales smaller than the wavelength, which both reduces the reflectivity and concentrates the optical field closer to the surface. Both patterned (triangular and rectangular lineshape) and planar Ni-Si back-to-back Schottky barrier metal-semiconductor-metal photodetectors on n-type ($\sim 5 \times 10^{14}$ cm^{-3}) bulk Si were studied. 1-D $\sim$50-250-nm linewidth, $\sim$1000-nm depth, grating structures were fabricated by a combination of interferometric lithography and dry etching. The nanoscale grating structures significantly modify the absorption, reflectance, and transmission characteristics of the semiconductor: air interface. These changes result in improved electrical response leading to increased external quantum efficiency (from $\sim 44\%$ for planar to $\sim 81\%$ for structured devices at $\lambda = 700$ nm). In addition, a faster time constant ($\sim$ 1700 ps for planar to $\sim$ 600 ps for structured at $\lambda = 900$ nm) is achieved by increasing the absorption near the surface where the carriers can be rapidly collected. Experimental quantum efficiency and photocurrents results are compared with a theoretical photocurrent model based on rigorous coupled-wave analysis of nanostructured gratings.**

***Index Terms*—Interferometric lithography, metal–semiconductor–metal (MSM), nanostructures, photodetectors, rigourous coupled-wave analysis (RCWA), silicon.**

I. Introduction

METAL-semiconductor-metal (MSM) photodetectors are attractive for many applications because of their high sensitivity-bandwidth product. [1] MSM photodetectors are generally planar devices with Schottky barriers on either side of an exposed semiconductor absorption region. The electrodes are often interdigitated to increase the active region area while optimizing the electric fields in the carrier collection region. In the ultraviolet and blue spectral regions, Si is a very attractive material [2] for MSMs because of its ready integration with Si electronics and comparable carrier saturation velocities with III-V semiconductors. However, at longer wavelengths, the relatively weak absorption, resulting from the Si indirect bandstructure, results in loss of both sensitivity and speed as carriers generated in field-free regions below the depletion region are subject to recombination losses as they slowly diffuse into the collection region of the MSM detector.

There is considerable interest in photodetectors for near-IR wavelengths ($\lambda = 0.77 - 0.85$ μm) for data communications applications due, in part, to the advantages of GaAs-based diode lasers. Both Si and GaAs are possible candidates for fabrication of optical receivers [3]. GaAs is attractive due to its short absorption length ($\sim$ 1.0 μm at $\lambda = 850$ nm), which makes it possible to have large bandwidth combined with good responsivity [4], but integration with Si electronics remains a significant issue. Silicon offers the potential of lower fabrication cost and direct integration with complementary metal-oxide-semiconductor (CMOS) technology; the difficulty is in the long absorption length (~ 15 μm at $\lambda = 850$ nm) that leads to difficult tradeoffs between efficiency and speed of MSM detectors. Si optoelectronic circuits with good sensitivity, but poor response times have been reported. [5] Si MSM lateral interdigitated photodetectors combine high speed with CMOS process compatibility. These detectors operate at very high frequencies at ultraviolet [2] and visible [6] wavelengths.

Some attempts have been reported aimed at improving the Si MSM detector quantum efficiency at visible and near IR wavelengths by etching deep vertical [7] and U-shaped [8] trench electrodes. These devices exhibit improvement in the visible wavelength range in comparison to planar contacts. However in order for these schemes to be effective for the near IR region the trench depths would need to be several tens of micrometers making the fabrication process complicated and expensive, and dramatically increasing the device capacitance. Liu *et al.* [9] reported improved response speed by fabrication of MSM photodetectors on silicon-on-insulator (SOI) substrates. The key feature in speed enhancement is the buried oxide layer that limits the active Si thickness. By reducing the Si film thickness to $\sim$ 100 nm, a photodetector bandwidth of 140 GHz at a 780-nm wavelength was achieved, however, at the cost of a very low quantum efficiency $\sim 1\%$. Lee *et al.* [10] have proposed a MSM detector configuration on 5-μm-thick Si membrane using a geometrical-texturing scheme for enhanced light trapping on the back surface. Light trapping in the thin membrane results in minimal reduction in responsivity, while reducing carrier transit times; however, the fabrication is complex and incompatible with standard Si device processing. Other attempts to improve the absorption by using hydrogenated amorphous Si (α-Si:H)

Manuscript received April 17, 2002; revised August 13, 2002. This work was supported by AFOSR and AFRL Space Vehicles Directorate.

A. K. Sharma is with the Air Force Research Laboratory, Space Vehicles Directorate, Kirtland AFB, NM 87111 USA and also with the Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106 USA.

S. H. Zaidi is with Gratings Inc., Albuquerque, NM 87109 USA.

P. C. Logofătu was with the National Institute of Laser, Plasma and Radiation Physics, Bucharest, Romania. He is now with the Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106 USA.

S. R. J. Brueck is with the Department of Electrical and Computer Engineering and the Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM 87106 USA.

Digital Object Identifier 10.1109/JQE.2002.805112

0018-9197/02$17.00 © 2002 IEEE

have focused on modifying the long-range structural symmetry of crystalline Si thereby relaxing the k-vector selection rule for optical transitions [11]. However, device efficiency is reduced as a result of the very short carrier lifetimes ($\sim$ 1 ps) in amorphous Si, and again integration with standard Si processing is problematic.

In this paper, we introduce a new Si MSM device concept that integrates a 1000-nm-deep subwavelength grating on the top of the active region. As we demonstrate below, the grating both serves as an effective anti-reflection coating, reducing the Si reflectivity over a broad wavelength range and at the same time enhances both the efficiency and the speed of the device by coupling energy into higher diffraction orders that are absorbed closer to the Si surface. These gains are achieved while retaining the capability of integration with standard Si CMOS process flows. In this paper, we report on one-dimensional (1-D) grating structures, which leads to polarization anisotropy in the detector response. For applications that require polarization-independent response, a 2-D grating structure can be used.

In Section II, we present the rigorous coupled-wave analysis (RCWA) used in the analysis of the absorption, reflection and transmission properties of 1-D Si nanostructures. The fabrication details of the nanometer period, 1-D Si gratings and the MSM photodetectors are presented in Sections III and IV. Sections V and VI present the experimental dc dark-current and spectral photoresponse results respectively and introduce a detailed model to explain the mechanism behind the enhanced photoresponse signal. In Section VII, we present the experimental time response results and provide a qualitative interpretation of the results based on the model given in Section VI. Section VIII concludes with an analysis and discussion.

II. Reflection, Transmission, and Absorption Properties of 1-D Nanostructure Gratings

The electromagnetic wave interaction with nanoscale grating structures is a complex function of both optical (wavelength, polarization, angle of incidence) and structural (period, profile, line:space ratio, depth) parameters [12]. Analysis of these effects starts from a Floquet expansion of the scattered and transmitted fields in terms of the grating wavevector

$$k_m = k_0\left(\sin(\theta) + m\frac{\lambda}{d}\right); \quad k_0 = \frac{2\pi}{\lambda} \tag{1}$$

where θ is the angle of incidence, d is the grating period, λ is the wavelength, and m is an integer. A well-developed approach is the RCWA [13], where the inhomogeneous dielectric constant $\varepsilon(x)$ in the grating region is expanded into a Fourier series that couples the equations for various values of m. A RCWA-based code was used to model the reflectance and transmission properties of the 1-D structured gratings presented in this paper. The absorption (A_g) in the nanostructured grating region can be calculated by energy conservation from

$$A_g = 1 - \sum_m T_m - \sum_{m'} R_{m'} \tag{2}$$

where $m, m' = 0, \pm1, \pm2, \pm3\ldots$ are the diffraction orders in the Si and the air space, respectively. The wavelength dependence of the refractive index coefficients used in the RCWA analysis here was taken from Palik *et al.* [14]. The reflection and transmission for a planar polished Si surface as a function of wavelength are necessarily restricted to the zero order. On the other hand, for nanostructured surfaces as a result of the high Si refractive index ($n \sim 3.77$ at 700 nm), there are numerous orders propagating in the Si even at the longest wavelengths studied. For an 850-nm period grating, there are four propagating diffracted orders in the Si at a wavelength of 800 nm and nine at 400 nm, while in the air there are two and three, respectively.

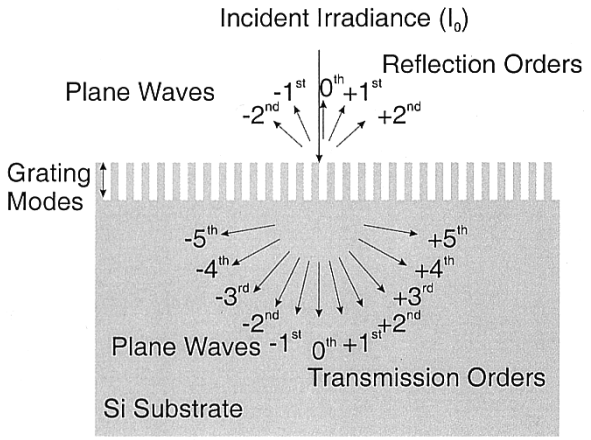
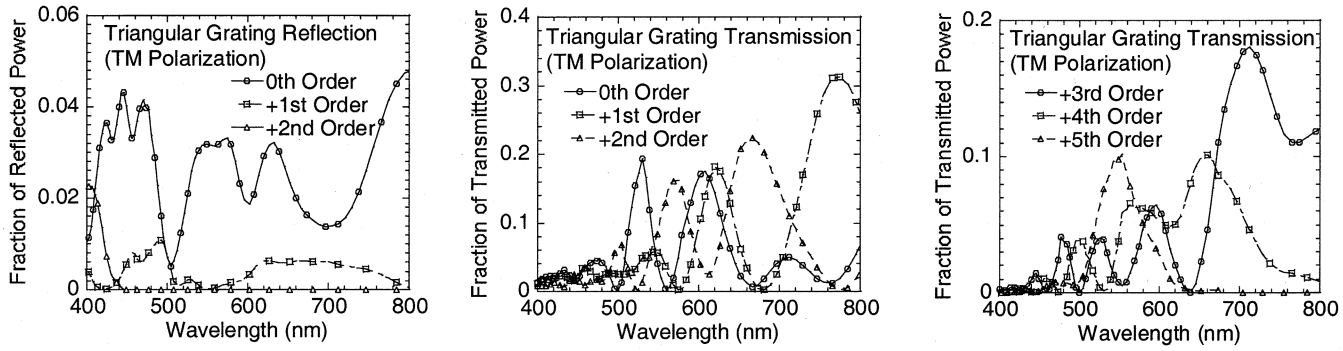


Fig. 1. Reflection and transmission diffraction orders for normal incident irradiance for silicon nanostructured gratings.

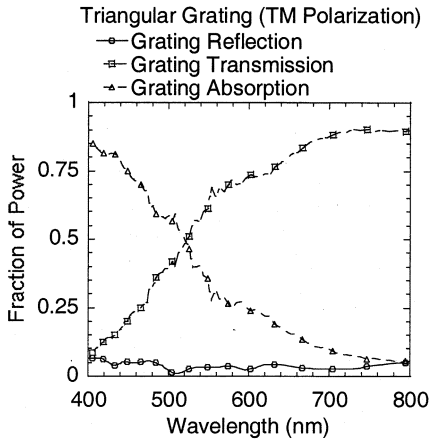
Fig. 1 shows a schematic of the reflection and transmission diffraction orders as well as the region of the grating resonant modes. The optical power in the reflected zeroth to second diffraction orders and transmitted zeroth to fifth diffraction orders as a function of wavelength are shown Fig. 2. These RCWA plots are for transverse magnetic (TM) polarization (defined relative to the grating wavevector), normal-incidence irradiance on a triangular shaped structure with period $\sim$ 850 nm, linewidths: top $\sim$ 50 nm, middle $\sim$ 300 nm, bottom $\sim$ 830 nm, and depth $\sim$ 1000 nm. The behavior is quite complex, exhibiting numerous resonances and the intensity shifts between the various diffracted orders as the wavelength is varied. These resonances are due, at least in part, to Fabry–Perot resonances of the grating modes in the vertical direction in these 1000-nm-deep gratings.

Since the grating is not blazed, the corresponding negative order powers are equivalent. The cutoff of the fifth order around 675 nm is evident. There are higher orders at shorter wavelengths that are not shown, but are included in the data analysis. For nanostructured grating surfaces, the higher diffracted orders contain a significant amount of energy and cannot be neglected. Fig. 3(a) shows the sum of all reflection orders, sum of all transmission orders, and the absorption in the triangular grating region, as a function of wavelength for TM polarization. The overall trend to larger grating absorption at shorter wavelengths is due to the increasing Si absorption coefficient. For comparison, the reflection and transmission power as a function of wavelength of a planar polished surface is plotted in Fig. 3(b). Note the dramatic reduction in the reflection for the triangular-structured surface compared to the planar surface.

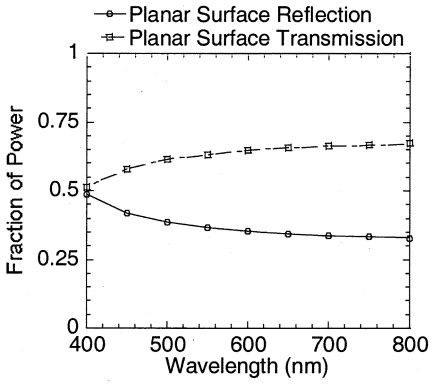


(a) (b) (c)

Fig. 2. RCWA power distribution profiles for triangular gratings as a function of wavelength for normal incidence TM polarization. (a) Zeroth, first, and second reflection diffraction orders. (b) Zeroth, first, and second transmitted diffraction orders. (c) Third, fourth, and fifth transmited diffraction orders.



(a)



(b)

Fig. 3. (a) Sum of all reflected and all transmitted diffraction-order powers and the absorption in the nanostructured grating region as a function of wavelength for a triangular grating for normal-incidence, TM-polarization irradiation. (b) Reflected and transmitted power for a planar surface as a function of wavelength.

Fig. 4 shows comparable results for the TM-incidence reflection, transmission and grating absorption for the rectangular structures. The overall behavior is quite similar to the triangular profile structures with some additional structure resulting from the more well-defined mode structure in the grating region. Similarly, TE-polarization reflection, transmission, and absorption profiles have been evaluated for the triangular and rectangular shaped nanostructured grating using the RCWA method. In Section VI, we will use the diffractive-order power distribution of the structures to evaluate the photocurrents of the structured devices.

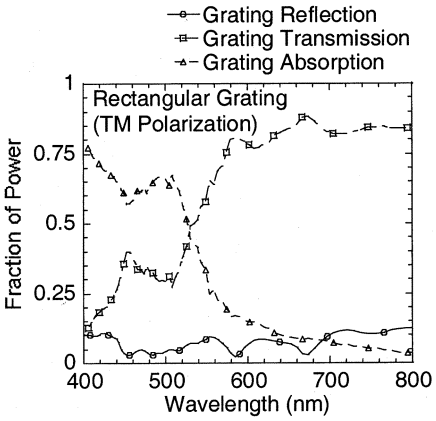


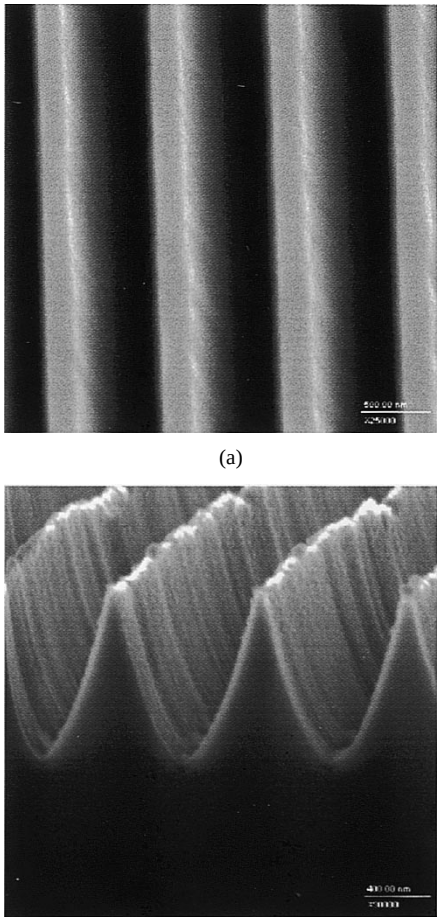
Fig. 4. The sum of all reflected and all transmitted diffraction-order powers and the absorption in nanostructured grating region as a function of wavelength for a rectangular grating.

III. Nanostructure Grating Fabrication

Interferometric lithography (IL) is ideally suited to inexpensive, large-area nanopatterning and was used in the nanostructure fabrication process. [15] A typical IL configuration consists of a collimated laser beam incident on a Fresnel mirror (FM) arrangement mounted on a rotation stage for period variation. [16] The interference between two coherent laser beams results in a periodic pattern at a period of $\lambda/2\sin\theta$ where λ is the optical wavelength and 2θ is the included angle between the interfering beams. The nanoscale grating patterns were first formed in photoresist followed by pattern transfer onto the underlying substrate using reactive-ion-etching (RIE) in a parallel plate reactor using CHF_3/O_2 plasma chemistry [17]. By controlling the plasma parameters, the profile can be varied from vertical to triangular. Fig. 5(a) and (b) shows rectangular and triangular-shaped Si gratings a period of $\sim$ 850 nm and depth of $\sim$ 1000 nm.

IV. MSM Detector Fabrication

Both rectangular and triangular shaped grating structures were fabricated in localized active areas of the MSM devices using IL, along with conventional I-line contact mask printing [18] for defining the device active areas and metal contact regions. For comparison purposes, we fabricated identical



(a)

(b)

Fig. 5. Scanning electron microscope (SEM) images of the nanostructured gratings in the active regions of the MSM photodetectors: (a) Top view of rectangular shaped gratings (b) Side view of triangular shaped gratings.

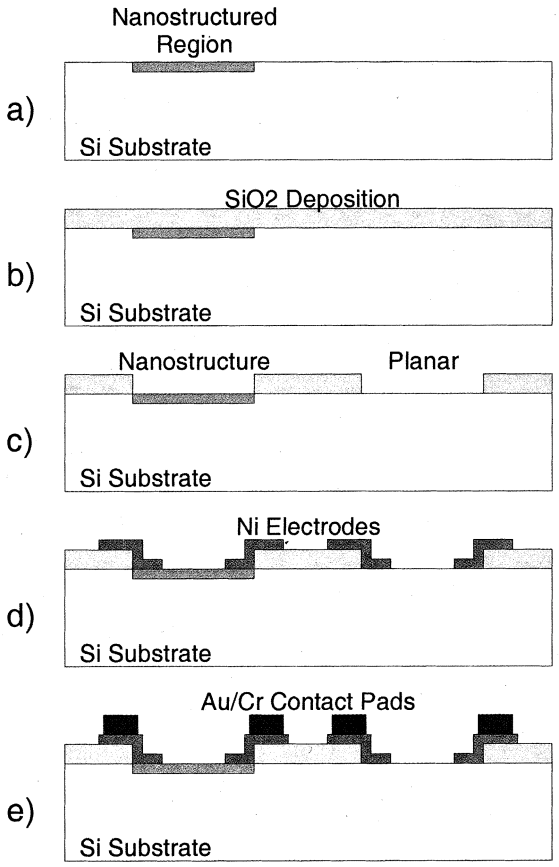


Fig. 6. Process flow sequence for fabrication of nanostructured and planar MSM photodetectors.

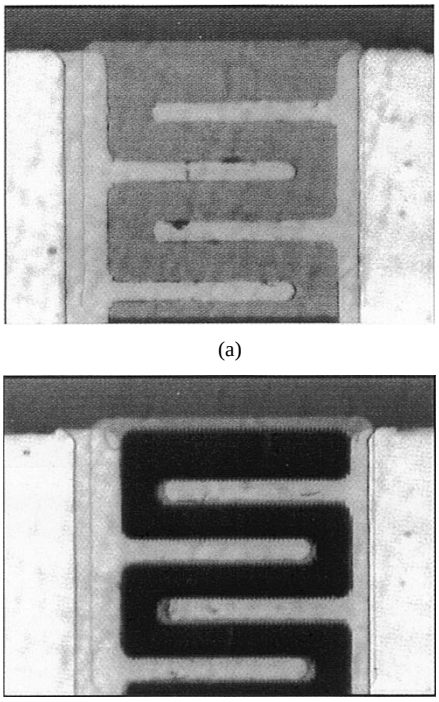
devices with planar active regions side-by-side with the nanostructured devices.

Fig. 6 shows a process flow sequence. A bare Si wafer was coated with photoresist (PR) and exposed to an interference pattern, as described above. A 30-nm-thick blanket layer of Cr was then deposited by e-beam evaporation and a lift-off step was used to create an array of Cr lines that act as an effective RIE etch mask. The wafer was again coated with PR layer and a mask was then used to selectively pattern the PR to expose the Cr lines in localized regions where the nanostructures were etched into the Si substrate by RIE, thus defining the nanostructured active areas. After the RIE step, the PR and the Cr etch mask lines are removed. There is considerable damage left on the nanostructured region by the RIE step. Two rapid thermal anneal (RTA) steps were performed next. The first RTA was performed at 900 °C for 5 min in a nitrogen environment to anneal the damaged surface [19], followed by a second RTA step for 3 min at 450 °C in a hydrogen environment to passivate the Si surface.

A low-temperature, evaporated SiO_2 isolation layer ($\sim$ 100-nm thick) was e-beam evaporated on the sample as an insulating layer [Fig. 6(b)]. A second mask was used to define active area windows in the oxide. Windows (ranging from 50×50 μm^2 to 150×150 μm^2) were opened in the oxide using a buffered oxide etch. Using a third mask, an interdigitated electrode pattern with contact pads extending onto the oxide was defined and 50-nm-thick nickel Schottky barrier electrodes were deposited by e-beam evaporation and lift-off with gaps ranging from 10 to 20 μm and aspect ratios of 5:1. Thicker 50:300 nm Cr:Au bonding metallization was then defined and evaporated on the contact pads above the oxide using a fourth mask and deposited by evaporation and lift-off. Fig. 7(a) and (b) shows optical micrographs of fully fabricated structured MSM and planar surface MSM devices, respectively. The active areas of the devices shown are 50×50 μm^2. The interdigitated electrode gaps are 10 μm, with the grating lines running perpendicular to the interdigitated contact fingers. Note that the active area of the structured device is much darker than the planar device due to the reduced grating reflection.

V. Dark Currents

Dark currents as a function of applied bias are shown in Fig. 8 for nanostructured and planar surface devices. Typical devices had 150×150 μm^2 active areas and 10-μm interdigitated electrode gaps. For both devices, the dark currents were comparable up to 27 V. At higher biases, the structured devices typically exhibit lower leakage than the planar devices. This difference is not fully understood at this time. However, it is worth mentioning the lower dark current in the textured devices suggests the possibility of higher sheet resistance in the nanostructured active regions compared to the planar surfaces due to



(a)

(b)

Fig. 7. Optical micrographs of fully fabricated MSM photodetectors with 50×50-μm^2 active regions and 10-μm interelectrode gaps. (a) Planar surfaced active area device. (b) Nanostructured (triangular) active area device.

the smaller cross-sectional area of the former. In addition, reactive-ion-etching (RIE) of the Si active areas introduces deep energy levels (surface states) which act as efficient energy recombination centers that can lead to further reduction in the dark current of the structured devices [20]–[22]. However, the photoresponse results suggest that surface recombination does not play a major role in the response.

VI. Spectral Response Measurements, Modeling, and Analysis

The photoresponse as a function of wavelength was measured using a continuous-wave (CW) xenon lamp source, 1/4-m monochrometer, polarizer, chopper, transimpedance amplifier (TIA), detector biasing-tee, and a lock-in amplifier. The lamp beam was focused onto the monochrometer entrance slit. The spectrally filtered output from the monochrometer was focused at normal incidence onto the active region of the MSM devices. The focused optical beam spot overfilled the active area of the MSM detector. Both structured and planar $150 \times 150\ \mu$m^2 active area devices with interdigitated 10-μm electrode gaps were characterized at a 40-V bias, depletion width $d_{\text{depletion}} \sim 10\ \mu$m (for $n \sim 5 \times 10^{14}$ cm^{-3}) extending fully across the active area.

The photoresponse signal for both the rectangular and triangular grating MSM devices was normalized to an adjacent planar device photoresponse signal. Since the structured and planar devices were identical in all aspects except for the nanostructures in the active region, we expect the normalized signal to reflect the photocurrent ratio profile between structured/planar devices, although device-to-device variations in the photocurrent of about $\Delta < \pm 15\%$ were observed within a nominally uniform ~ 1 cm^2 processed area, presumably due to contaminants/defects in the metal/semiconductor interface junction. These variations contribute a small dc shift to the normalized curves. We sampled six devices of each kind and report best results.

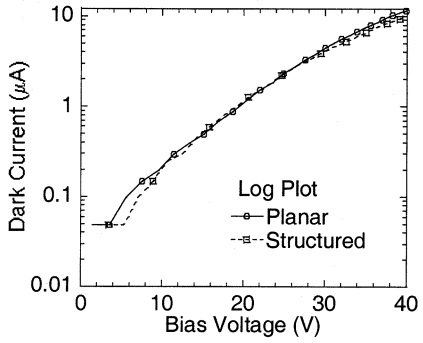
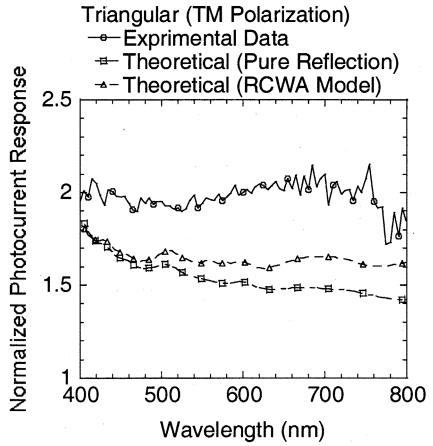
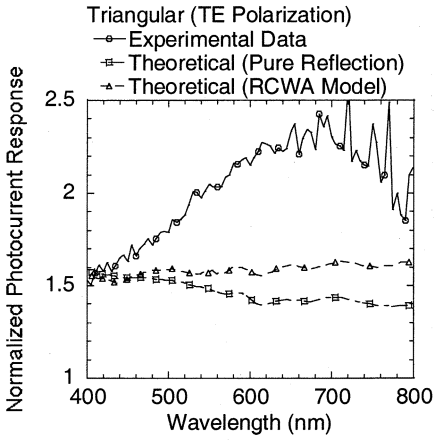


Fig. 8. Log scale-plot of dark current for structured and planar devices as a function of bias voltage.



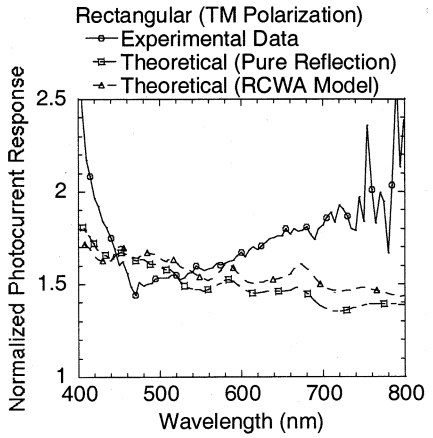
(a)



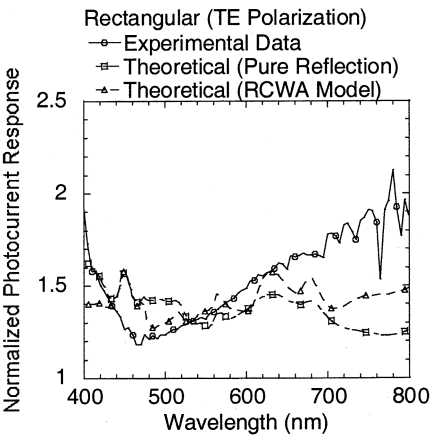
(b)

Fig. 9. Experimental and theoretical plots of normalized (triangular-structured/planar) MSM device photocurrent responses for normal incidence: (a) TM polarization and (b) TE polarization irradiation. The lower theoretical curve includes the anti-reflection properties of the grating. The upper theoretical curve (closer to the experiment) shows the results of a calculation accounting for the smaller absorption depth of the diffracted radiation (see text for details).

The grating structure (rectangular and triangular) parameters were given in Section II for the characterized device results presented in this section. Figs. 9 and 10 show experimental results of normalized spectral responses as a function of wave-



(a)



(b)

Fig. 10. Experimental and theoretical plots of normalized (rectangular-structured/planar) MSM device photocurrent responses for normal incidence: (a) TM and (b) TE polarizations. Theoretical curves as in Fig. 9.

length for the (triangular and rectangular) shaped nanostructured MSM devices for both TM and TE polarized normal incident irradiance. As seen from Fig. 9(a), the experimental TM normalized spectral response of the triangular structured device shows a broadband enhancement of $\sim 2\times$ over the entire 400- to 800-nm range. The rectangular profile grating devices show a somewhat lower overall enhancement, but with larger enhancements at both the shortest and longest wavelengths investigated. The noise at the long-wavelength end results from the low signal levels in this experimental arrangement.

Some, but not all, of the enhancement is due to the anti-reflection properties of the nanostructured surfaces. For example, at a wavelength of 700 nm, the reflectivity of a planar Si surface is 33% compared with $\sim 3\%$ (triangular) and $\sim 7\%$ (rectangular) for the structured devices. In order to evaluate the absolute quantum efficiency (η) at this wavelength, a CW mode-locked Ti:Al_2O_3 laser was used and the beam was focused using a 10 $\times$ lens to a spot size smaller than the 10-μm interelectrode gap. The calculated enhancement for the triangular structure is 1.44. From the experiment, the photocurrent density of $\sim$ 0.25 A/cm^2 (planar) and $\sim$ 0.46 A/cm^2 (triangular) was measured for an incident optical power of 0.24 mW or (1.067 W/cm^2), giving an enhancement of $\sim$ 1.85. This results in an external quantum efficiency η^{ext} of $\sim$ 44% (planar) and

TABLE I
EXPERIMENTAL RESULTS OF SURFACE REFLECTANCE, RESPONSIVITY, EXTERNAL AND INTERNAL Q.E. FOR PLANAR AND NANOSTRUCTURED DEVICES AT $\lambda = 700$ nm FOR TM POLARIZATION AT NORMAL INCIDENCE

Device	Surface Reflection	Responsivity (A/W)	External Q.E.	Internal Q.E.
Planar	~33%	~0.23	~44%	~66%
Structured	~3%	~0.43	~81%	~84%

81% (triangular) and an internal quantum efficiency [23] $\eta^{\text{int}} = \eta^{\text{ext}}/(1-R)$ of 66% (planar) and 84% (triangular) (see Table I).

Simply correcting the photoresponse for the anti-reflection properties of the gratings gives the lower theoretical curves shown along with the experimental curves in Figs. 9 and 10. As can be seen from Fig. 9, at the shorter wavelengths, the results are close to the experimental results, however as the wavelength increases the gap between the theory and experiment increases for both the TE and TM polarizations. A similar trend is found for the rectangular grating case in Fig. 10. The RCWA calculations are sensitive to the details of the grating lineshape for both the rectangular and triangular cases. The triangular structures [Fig. 5(b)] are closer to ideal then the rectangular structures, which have curved rather than square bottoms. Therefore, the data analysis will be concentrated on the triangular gratings.

We postulate that the additional enhancement in the quantum efficiency, beyond the antireflection properties of the grating, is due to carriers generated in the grating region and to redirection of a significant part of the power transmitted into the Si into higher diffraction orders that propagate at an angle to the normal and therefore are absorbed closer to the surface where the probability of collection is higher. Note from the above analysis that η^{int} of the planar device is only 66%, i.e., that 34% of the photoexcited carriers are not collected. We are not aware of any models for the dependence of the internal quantum efficiency on depth within the semiconductor, but it is reasonable to assume that this efficiency is higher for carriers excited at the surface where the bias fields of the MSM are large and that it decreases as the carriers are generated deeper within the semiconductor. To allow a calculation, we make the simple ansatz

$$\eta = e^{-\alpha_C z} \tag{3}$$

where z is the depth into the semiconductor and α_C is a scaling parameter, determined from the planar device experimental results, that reflects the assumption that carriers generated near the surface to contribute a higher quantum efficiency than those generated deeper into the material.

Then the total external quantum efficiency is given by

$$\begin{aligned}\eta_{\text{total}}^{\text{ext}} =& \eta_{\text{grating}} \\ &+ \sum_{\text{allorders}} \frac{\cos\theta_m P_m}{\alpha P_i}\left[\int_0^{\infty} e^{-(\alpha/\cos\theta_m)z} e^{-\alpha_C z} dz\right] \\ =& \eta_{\text{grating}} + \sum_{\text{allorders}} \frac{P_m}{P_i}\frac{1}{1+\frac{\alpha_C\cos\theta_m}{\alpha}}\end{aligned} \tag{4}$$

where P_m and θ_m are the power diffracted into and the propagation angle of the m^{th} order, respectively, P_i is the incident power, and α is the Si absorption coefficient. In the limit of $\alpha_C \to 0$, i.e,.

TABLE II
THEORETICAL CALCULATIONS OF COMPONENT AND TOTAL EXTERNAL QE AT $\lambda = 700$ nm FOR TM POLARIZATION AT NORMAL INCIDENCE BASED ON RCWA PLOTS OF FIG. 2

Contributing factor TM polarization λ=700nm	Fraction of Power	Angle of propagation (θ_t)	Photocurrent density per contributor factor	Component Q.E.	Partial Q.E.
Surface reflection	~0.03	--	--	--	--
Nanostructured absorption (α)	~0.10	--	~ 0.06A/cm^2	~ 100%	~ 10%
0th order	~0.04	~0^0	~ 0.016A/cm^2	~ 63%	~ 2.5%
±1st order	~0.05	~13^0	~ 0.02A/cm^2	~ 63.6%	~ 3.2%
±2nd order	~0.31	~26^0	~ 0.12A/cm^2	~ 65.4%	~ 20.3%
±3rd order	~0.33	~41^0	~ 0.14A/cm^2	~ 69.2%	~ 23%
±4th order	~0.14	~61^0	~ 0.07A/cm^2	~ 77.8%	~ 11%
Total Power 1.00		Total Photocurrent density ~ 0.43A/cm^2		Total Σ(partial Q.E) η_{ext} ~ 70%	

in the absence of any dependence of the collection efficiency on the absorption depth, this reduces to the usual result.

Recall that the planar and structured devices are identical in all aspects except for the grating structures; therefore, the carrier collection mechanisms in the two devices should be the same. If we estimate the fitting parameter α_C from the planar device results, it should be directly applicable to the structured device. For normal incident irradiance on a planar surfaced device, where there are no high-order modes, all of the energy is contained in the zeroth-diffraction order mode, which propagates at $\theta_0 = 0$. As shown in Fig. 3(b), at 700 nm, a planar silicon surface reflects $\sim 33\%$ of the incident light, or $P_0/P_i \sim 0.67$. The absorption coefficient $\alpha = 1.8 \times 10^3$ cm^{-1} at this wavelength, and therefore from (3), $\alpha_C \sim 8.8 \times 10^2$ cm^{-1}. We can now estimate the total external efficiency for the structured devices by using the RCWA modeling results. From Figs. 2 and 3, at 700 nm, the modeled reflection loss is $\sim 3\%$, and $\sim 10\%$ of the incident irradiance energy is absorbed in the nanostructured grating region. The rest of the incident irradiance energy is distributed as follows: 1) $\sim 4\%$ is coupled into the zeroth order; 2) $\sim 2 \times 2.5\%$ is coupled into the $\pm 1^{\text{st}}$ order with a propagating transmitted angle of $\theta \pm_1 \sim \pm 13^\circ$ with respect to the normal; 3) $\sim 2 \times 15.5\%$ is coupled into the $\pm 2^{\text{nd}}$ orders propagating at $\sim \pm 26^\circ$; 4) $\sim 2 \times 16.5\%$ is coupled into the $\pm 3^{\text{rd}}$ orders propagating at $\sim \pm 41^\circ$; and 5) $\sim 2 \times 7\%$ of the energy is coupled into the $\pm 4^{\text{th}}$ order propagating at $\sim \pm 61^\circ$. Using (4), we can calculate the partial quantum efficiencies for each order, with the results shown in Table II. The ratios of photocurrents between the structured and planar devices for this case are: $\sim 1.7\times$ ($\sim 70\%$ total external quantum efficiency) for TM polarization, which is in the correct direction, but insufficient to fully explain the experimental results. The curves in Figs. 9 and 10 labeled theoretical (RCWA calculation) show similar calculations across the spectral region for both TE and TM polarizations for the triangular and rectangular gratings.

TABLE III
DEPLETION DEPTH FOR $n \sim 5 \times 10^{14}$ cm^{-3} DOPED SILICON AS A FUNCTION OF APPLIED BIAS VOLTAGE ACROSS A 10-μm ELECTRODE GAP

Bias (V)	2.5	5	10	15	20	25	30	35	40
Depletion Depth (μm)	~2.5	~3.6	~5.1	~6.2	~7.2	~8.0	~8.8	~9.5	~10.1

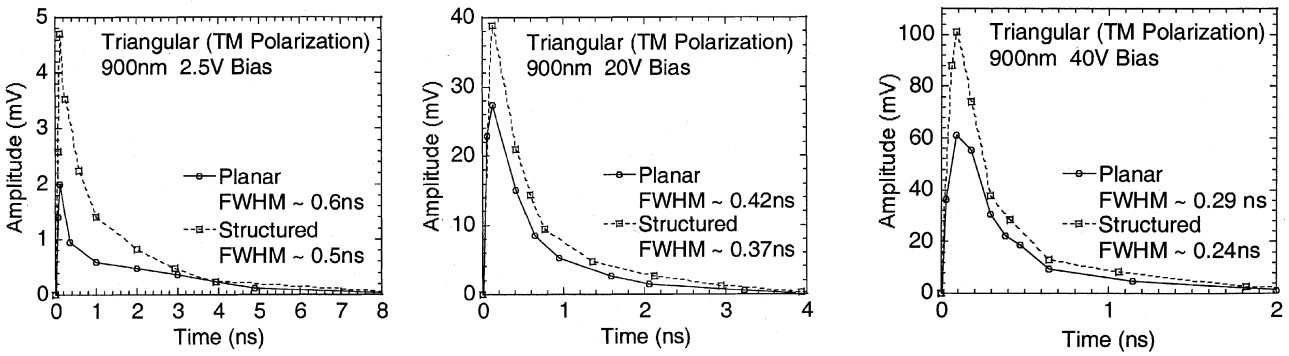
VII. TIME RESPONSE

In the previous section, we analyzed the impact of nanostructuring the Si surface on the CW quantum efficiency of the photodetectors. Another beneficial aspect of generating carriers closer to the surface is faster carrier collection.

Pulsed response measurements are taken using $\sim$ 150-fs duration excitations at two different wavelengths (900 and 400 nm) from a CW mode-locked Ti:Al_2O_3 laser (doubled for the short wavelength, 0.24-mW average power for both wavelengths, 77-MHz repetition rate). Photodetectors were probe tested using an 18 GHz probe and a high-speed sampling oscilloscope. The laser spot size was ~ 50 μm in diameter. Although the spot size was larger than the interdigitated electrode gap (10 μm) the total focal spot was well within the 150×150 μm^2 active region of the devices tested. Normal-incidence TM polarization relative to the grating wavevector excitation was used for both wavelengths. At 900 nm, the absorption depth extends to $\sim$ 20 μm, while at 400 nm the absorption depth is < 1 μm. These two absorption depths provide insight regarding the relative contributions of carriers generated within or below the depletion region. For both wavelengths, the time response measurements were taken for bias voltages varying from 2.5 to 40 V. Table III shows the depletion width at each bias voltage.

Before the experimental data and analysis is provided, it is useful to review [24] the three primary factors that impact MSM photodetector response speed.

1) *Drift of carriers through the depletion region.* At high E fields, the drift velocities of both electrons and holes



(a) (b) (c)

Fig. 11. Time-response measurements for TM-polarization, normal-incidence irradiance at 900 nm as a function of bias voltage. (a) 2.5 V. (b) 20 V. (c) 40 V.

in Si saturate [25] at about 1×10^7 cm/s. Provided the field within the depletion region exceeds the saturation value for most of its length, we assume that the carriers move with a constant velocity $v_{\rm sat}$. The longest transit time $\tau_{\rm drift} = W/v_{\rm drift}$ will result when carriers are generated near one edge of depletion layer width W. For a 10-μm drift region in Si, $\tau_{\rm drift} \sim 100$ ps.

2) *Diffusion of carriers to the depletion region.* The time it takes for carriers to diffuse a distance d is $\tau_{\rm diff} = d^2/2D$, where D is the carrier diffusion coefficient, for holes in Si and $d \sim 10$ μm, $\tau_{\rm diff} \sim 40$ ns $\gg \tau_{\rm drift}$. Those carriers generated within the depletion region respond rapidly, while those generated outside give rise to a "slow" tail.

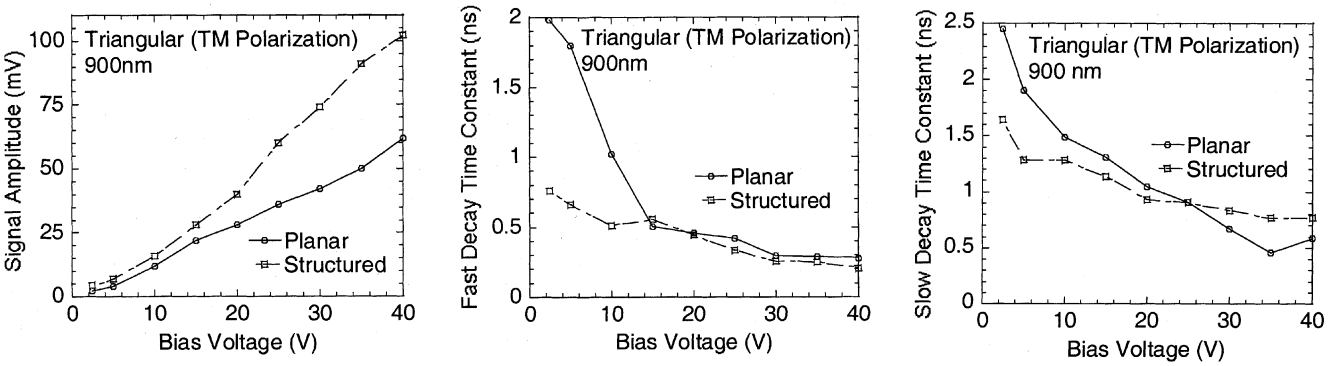
3) *Junction and parasitic capacitance effects.* A metal-semiconductor junction under reverse bias exhibits a voltage-dependent capacitance caused by the variation in stored charge at the junction represented by the relation $C_j = A/2(2e\varepsilon_s N_d)^{1/2}{\rm V}^{-1/2}$. This capacitance is usually quite small for MSM devices as a result of their planar design (A small, depletion-width large) There are also parasitic circuit capacitances associated with the packaging and cabling that often dominate the junction effects, as well as the limiting response of the electronics. Both structured and planar detectors have an identical $\tau_{\rm circuit}$ and measurement limitations for these measurements.

We first present the time response measurements for an excitation wavelength of 900 nm (absorption length of ~ 20 μm, planar device reflection loss $\sim 28\%$, and the structured device reflection loss $\sim 5\%$). For a low bias voltage of 2.5 V, the depletion width is only $\sim$ 2.5 μm, extending only part way between the electrodes. In this case, large number of carriers are generated outside the depletion region and the device time is limited by carrier diffusion. The time response signals for both structured and planar devices are shown in Fig. 11. Both the peak signal amplitude (planar 2.0 mV; triangular 4.7 mV) and total area under the curve (planar $\sim$3.3 V·ps; triangular $\sim$7.8 V·ps) are significantly smaller for the planar device. The ratio of the integrated areas is 2.4, higher than the ratio of the external quantum efficiencies of about 1.8 at this wavelength, suggesting that there is an even longer response tail for the planar devices due to carriers generated deep in the material that contribute to the total response. The peak response reflects this same ratio, the high speed response is somewhat slower for the textured device. The long tail for both devices is due to carrier diffusion from electron-hole pairs being generated outside the depletion region.

As the bias voltage is increased, the depletion region extends across the active region thus increasing the drift component signal strength as seen in Fig. 11(b) and (c) for both planar and structured devices. The peak amplitude of the time response versus applied voltage is plotted in Fig. 12(a), as the bias is increased more carriers are collected and the signal becomes larger. At this long wavelength of 900 nm, no saturation in carrier collection is observed up to the highest applied bias where the depletion region extends fully across the device, but only to a depth of ~ 10 μm (fringing fields from the surface electrodes), whereas the absorption depth is > 20 μm.

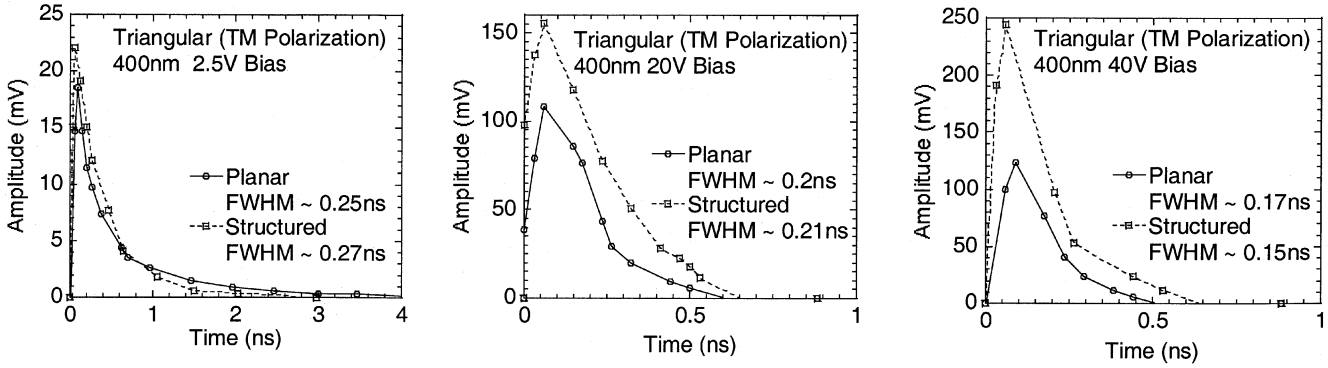
To further analyze these results, fast and slow components of the time constants were evaluated from the slopes of the log (time response), as plotted in Fig. 12(b) and (c). The structured device has a shorter fast response time at low voltages; the time responses are similar at high applied biases.

For a 400-nm excitation, the absorption occurs within the first micrometer from the surface, e.g., almost totally within the grating for the textured device. At this wavelength, the planar device reflects about $\sim 49\%$ of the incident light and the structured device reflects about $\sim 7\%$. Fig. 13 shows the time response for the structured and planar device at biases ranging from 2.5 to 40 V. While the absorption depth is short, the depletion width extends only part way across the transverse dimension of the devices at low biases and there is still a significant diffusion contribution to the response. At the highest bias, the depletion region extends across the active region of the device and there is no diffusion contribution. The time responses are significantly faster than at the longer wavelength, corresponding to the shortened transit times for this short wavelength absorption. The structured device has a larger area near the upper part of the signal (fast component) indicating a larger total power collection due to the improved light coupling and low reflection losses compared to the planar device. The integrated time responses [Fig. 14(a)] are roughly equal at low bias, but show the expected factor of $\sim$ 2 (from the CW results) at high biases. As the bias is increased, more carriers are collected until saturation is reached, in contrast to the long-wavelength results where saturation was not observed [Fig. 13(a)]. Fig. 14(b) shows the carrier time constant curve as a function of bias voltage; at



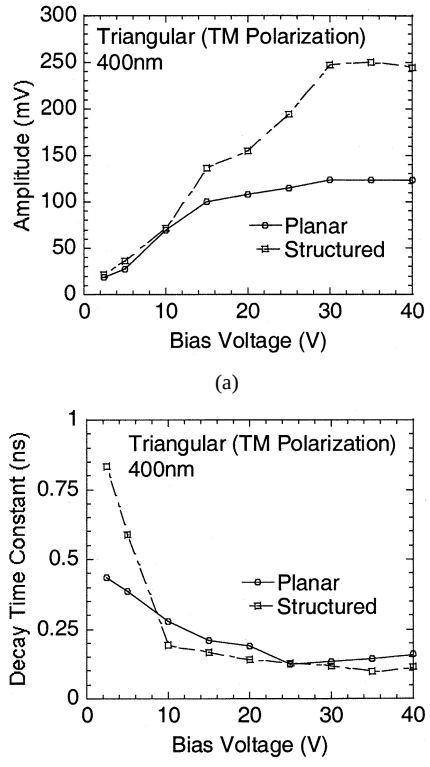
(a) (b) (c)

Fig. 12. Comparative analysis of the time response of triangular-structured and planar devices for normal incident irradiance. (a) Signal amplitude as a function of bias voltage. (b) Fast time-decay constant as a function of bias voltage. (c) Slow time-decay constant as a function of bias voltage.



(a) (b) (c)

Fig. 13. Time-response measurements for TM polarization, normal-incidence irradiance at 400 nm as a function of bias voltage. (a) 2.5 V. (b) 20 V. (c) 40 V.



(a)

(b)

Fig. 14. Comparative analysis of the time response of triangular-structured and planar devices for normal incident irradiance. (a) Signal amplitude as a function of bias voltage. (b) Time-decay constant as a function of bias voltage.

higher biases the response time reaches an instrumental limit of ~100–150 ps for both planar and structured devices. The evidence of undershoot seen in Fig. 13 also confirms that the response is instrumentally limited.

VIII. Analysis and Conclusion

We employed the optical reflection, transmission, and absorption properties of Si nanoscale structures to improve the performance of silicon MSM photodetectors. MSM photodetectors were fabricated with and without nanostructures in the active area for comparison purposes. Interferometric lithography was used for the nanopatterning the structures and conventional optical lithography was used for defining the device active and metal contact areas. Devices characterized had a $150 \times 150\ \mu\text{m}^2$ area with 10-μm interdigitated electrode gaps.

RCWA was used to model the reflection and transmission diffraction order modes as a function of wavelength. The power and angle of transmission of the transmitted diffraction higher orders as a function of nanostructure geometry is the key mechanism that improved the device performance by generating carriers near the surface of the active region. The internal quantum efficiency increased from $\sim 66\%$ to $\sim 84\%$ at $\lambda = 700$ nm. By increasing the absorption near the surface, we not only increase the total carrier collection efficiency, but also improve the time response of the signal. This was shown by the time response data signals at $\lambda = 900$ nm that show the fast part of the signal originating from higher diffraction order photo generated carriers

that are collected near the surface have a narrower width. However, the slower component of the time response signal, originating from the zeroth-order mode, is comparable in both planar and structured devices, due to the same absorption depth from which the photogenerated carriers are diffusing from into the depletion region. In order to improve the complete time response signal, we need to investigate a structured design that eliminates the deep carriers. For future work, we propose to fabricate similar devices in silicon-on-insulator materials where the absorption depth is restricted by the device structure rather than the intrinsic Si absorption.

ACKNOWLEDGMENT

The authors would like to thank A. Frauenglass for assisting in the setting up of the spectral measurement setup, Y. Wo for assisting in the automation of the monochrometer, and G. Liecthy for assisting in the high-speed data acquisition and computer simulation.

REFERENCES

[1] S. Y. Chou and M. Y. Liu, "Nanoscale tera-hertz metal-semiconductor-metal photodetectors," *IEEE J. Quantum Electron.*, vol. 28, pp. 2358–2368, Oct. 1992.

[2] B. W. Mullins, S. F. Soares, K. A. McArdle, C. Wilson, and S. R. J. Brueck, "A simple high-speed Si Schottky diode," *IEEE Photon. Technol. Lett.*, vol. 3, pp. 360–362, 1991.

[3] M. Ghioni, F. Zappa, V. P. Kesan, and J. Warknock, "A VLSI-compatible high-speed silicon photodetector for optical data link applications," *IEEE Trans. Electron Devices*, vol. 43, pp. 1054–1060, 1996.

[4] J. S. Wang, C. G. Shih, W. H. Chang, J. R. Middleton, P. A. Apostolakis, and M. Feng, "11-GHz bandwidth optical integrated receivers using GaAs MESFET and MSM technology," *IEEE Photon. Technol. Lett.*, vol. 5, pp. 316–318, 1993.

[5] A. Sayles and J. Uyemura, "An optoelectronic CMOS memory circuit for parallel detection and storage of optical data," *IEEE J. Solid-State Circuits*, vol. 23, pp. 1110–1115, 1991.

[6] S. Alexandrou, C. Wang, T. Hsiang, M. Y. Liu, and S. Y. Chou, "A 75 GHz silicon metal-semiconductor-metal Schottky photodiode," *Appl. Phys. Lett.*, vol. 62, pp. 2507–2509, 1994.

[7] Y. L. Ho and K. S. Wong, "Bandwidth enhancement in Si metal-semiconductor-metal photodetectors by trench formation," *IEEE Photon. Technol. Lett.*, vol. 8, pp. 1064–1066, 1996.

[8] L.-H. Laih, T.-C. Chang, Y.-A. Chen, W.-C. Tsay, and J.-W. Hong, "Characteristics of MSM photodiodes with trench electrodes on p-type Si wafer," *IEEE Trans. Electron Devices*, vol. 45, pp. 2018–2023, 1998.

[9] M. Y. Liu, E. Chen, and S. Y. Chou, "140 GHz metal-semiconductor-metal photodetectors on silicon-on-insulator substrate with a scaled layer," *Appl. Phys. Lett.*, vol. 65, pp. 887–888, 1994.

[10] H. C. Lee and B. Van Zeghbroeck, "A novel high-speed silicon MSM photodetector operating at 830 nm wavelength," *IEEE Trans. Electron Devices*, vol. 16, pp. 175–177, 1995.

[11] N. F. Mott and E. A. Davis, *Electronic Processes in Noncrystalline Materials*, 2nd ed. Oxford, U.K: Oxford Univ. Press, 1979, ch. 6.

[12] S. Kemme, S. H. Zaidi, and J. M. Gee, "Submicron diffractive gratings for thin film solar cell applications," presented at the 9th Workshop on Crystalline-Silicon Materials and Processes, Breckenridge, CO, Aug. 1999.

[13] M. G. Moharam and T. K. Gaylord, "Rigorous coupled-wave analysis of planar-grating diffraction," *J. Opt. Soc. Amer.*, vol. 71, no. 7, pp. 811–818, 1981.

[14] E. D. Palik, *Handbook of Optical Constants of Solids*. New York: Academic, 1985, pp. 555–568.

[15] X. Chen and S. R. J. Brueck, "Imaging interferometric lithography: A wavelength division multiplex approach to extending optical lithography," *J. Vac. Sci. Technol.*, vol. B16, pp. 3392–3397, 1998.

[16] S. H. Zaidi and S. R. J. Brueck, "Multiple exposure interferometric lithography," *J. Vac. Sci. Technol.*, vol. B11, p. 653, 1993.

[17] ——, "Si-texturing with sub-wavelength structures," *IEEE PVSC*, vol. 26, pp. 171–174, 1997.

[18] S. H. Zaidi, S. R. J. Brueck, T. Hill, and R. N. Shagam, "Mix and match interferometric and optical lithographies for nanoscale structuring," in *Proc. SPIE*, vol. 3331, 1998, p. 406.

[19] S. Wolf and R. N. Tauber, *Silicon Processing for the VLSI Era*. Sunset Beach, CA: Lattice Press, 1986, vol. 1, pp. 57–58.

[20] S. K. Ghandhi, *VLSI Fabrication Principles*, 2nd ed. New York: Wiley, 1994, pp. 638–639.

[21] S. H. Zaidi, R. Douglas, and J. M. Gee, "Characterization of random reactive ion etched-textured silicon solar cells," *IEEE Trans. Electron Devices*, vol. 48, p. 1200, 2001.

[22] N. Yabumoto, M. Oshima, O. Michikani, and S. Yoshi, "Surface damage on silicon substrates caused by reactive sputter etching," *Jpn. J. Appl. Phys.*, vol. 20, p. 893, 1981.

[23] A. L. Fahrenbruch and R. H. Bube, *Fundamentals of Solar Cells*. New York: Academic, 1983, p. 81.

[24] J. Wilson and J. F. B. Hawkes, *Optoelectronics*. Englewood Cliffs, NJ: Prentice-Hall, 1989, ch. 7, pp. 286–289.

[25] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. New York: Wiley, 1981, p. 46.



A. K. Sharma (F'01) received the B.S. and M.S. degrees in electrical engineering in 1990 and 1994, respectively, from the University of New Mexico, Albuquerque, where he is currently working toward the Ph.D. degree in electrical engineering at the Center for High Technology Materials.

Since 1994, he has been an Electronics Engineer at the Air Force Research and Laboratory, Space Vehicles Directorate, Kirtland AFB, NM. His research interests include nanoscale semiconductor device fabrication techniques, physics of nanostructures, and nanodevices for high-speed and low-power-consumption electronic applications.



S. H. Zaidi received the Ph. D. degree in physics from the University of New Mexico, Albuquerque, in 1989. His dissertation was on interferometric lithography and optical coupling to surface plasma waves.

He was a Researcher at the Center for High Technology Materials, University of New Mexico, for a decade, helping develop several applications of interferometric lithography. Since 2000, he has been with Gratings, Inc., Albuquerque, NM, where his focus has been on the development and application of microrscale and nanoscale random and periodic lithography techniques for a wide range of materials and devices. His current research interests include Si solar cells, nanoscale MOSFETs, and heteroepitaxial growth on nanoscale features.



P. C. Logofatu received the M.S. degree in technological physics from University of Bucharest, Bucharest, Romania, in 1992 and the Ph.D. degree in optical sciences from the University of New Mexico, Albuquerque, in 2000.

Between 1992 and 1996, he was with the National Institute for Laser, Plasma and Radiation Physics, Bucharest, Romania. Presently, he is a post-doctoral Fellow at the Center for High Technology Materials, University of New Mexico, where his research interests include scatterometry and error analysis.



S. R. J. Brueck (S'63–M'71–SM'89–F'93) was born in New York City in 1944. He received the B.S. degree from Columbia University, New York, in 1965, and the M.S. and Ph.D. degrees from Massachussetts Institute of Technology (MIT), Cambridge, in 1967 and 1971, all in electrical engineering.

From 1971 to 1985, he was a Staff Member at MIT Lincoln Laboratory. In 1985 he moved to the University of New Mexico, Albuquerqe, where he is currently a Professor in the EECE and Physics and Astronomy Deparments and Director of the Center for High Technology Materials. His current research interests include nanoscale lithography, the physics of nanostructures, the nonlinear optics of poled glasses, and semiconductor laser physics.

Radiation from a dipole embedded in a multilayer slab

S. R. J. Brueck,*,† V. A. Smagley,† and P. G. Eliseev
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106, USA
(Received 21 March 2003; published 12 September 2003)

An analytical solution for the radiation emitted from a dipole embedded in an arbitrary, planar dielectric film stack is presented. The calculation uses a rigorous Hertz-vector formalism to treat the electromagnetic boundary conditions. The radiation fields are then evaluated in a far-field approximation to get the radiated fields far from the dipole. Both two-dimensional (2D) emission into bound modes of the dielectric stack and three-dimensional (3D) emission into radiation fields above and below the stack are evaluated. These solutions are explored for two simple cases: a InGaAs slab symmetrically clad with up to four high-contrast (Al_2O_3/GaAs) Bragg mirror pairs and semi-infinite air spaces, and a similar asymmetric structure with a GaAs substrate on one side. The symmetric structure supports both 2D bound and 3D radiation fields. The asymmetric structure only supports 3D radiation fields since there are no strictly bound modes, but ''leaky'' modes appear that are very similar to the bound modes in the symmetric structure except that the radiated power ultimately is transmitted into the substrate in a very highly directional beam. This calculation is applicable to a wide range of solid-state photonic devices, including vertical-cavity and edge-emitting lasers, spontaneous light-emitting diodes, and photodetectors.

DOI: 10.1103/PhysRevE.68.036608 PACS number(s): 41.20.−q, 42.25.−p, 42.55.−f, 42.82.−m

I. INTRODUCTION

The coupling of an electric dipole with electromagnetic fields in solids is fundamental to solid-state light emitting and detecting devices (LEDs, lasers, displays, photodetectors, etc.). In a uniform infinite medium the dipole radiation fields are very well known. The study of these radiation fields in the presence of metallic and dielectric interfaces has been a recurring topic in electromagnetism. Sommerfeld [1,2] made seminal contributions to the solution that provide the basis of the work presented here. He was concerned with issues of long-wavelength radio wave propagation over the surface of the earth and addressed the single interface problem of a dipole above a lossy dielectric. As is very well known [3,4], the addition of a second interface, e.g., a dipole embedded within a dielectric slab, adds significant complexity to the problem, since the radiation is now distributed between three-dimensional (3D) radiation out the sides of the slab and the two-dimensional (2D) bound modes supported by the slab. With the development of multilayer structures with very high, and precisely tailorable, reflectivities, such as vertical-cavity lasers, resonant cavity LEDs, and photodetectors, there is a need for a more detailed understanding of the distribution of the radiation between these components and the corresponding angular and thickness/wavelength dependencies. Both approximate analytic treatments [5,6] and numerical solutions [7] have appeared in the literature. Recently, two equivalent detailed analytical treatments for the simple slab case have been reported [4,8]. Both of these treatments presented fully analytic results for the radiation into both the 2D bound modes and the 3D radiation modes and investigated in detail the dependence on dielectric contrast and slab thickness.

The purpose of this paper is to extend these results [4] to arbitrary multilayer dielectric structures and to investigate in particular the impact of high-reflectivity Bragg mirrors on the dipole radiation. The analysis is extended to asymmetric cases, including mode-free cases that support ''leaky'' modes that are not bound in the strict sense, but rather build up significant intensity within the multilayer structure and ultimately radiate their power into 3D radiation within the substrate.

Section II presents the Hertz vector formulation and introduces a significant algebraic simplification by introducing both electric and magnetic Hertz vectors. Section III presents the closed form analytic results for arbitrary multilayer structures. Detailed investigation of the results for a symmetric structure with high-reflectivity Bragg reflectors is presented in Sec. IV A. A related asymmetric structure with a high-index substrate is analyzed in Sec. IV B. Finally, conclusions are presented in Sec. V.

II. HERTZ VECTOR FORMULATION

A. Geometry of the problem

The geometry of the problem is shown in Fig. 1. A dipole is embedded in an arbitrary film stack taken as infinite in the (x,y) directions. The z direction is perpendicular to the films. The dipole is oriented at an angle θ_d from the z direction in the (x,z) plane. The layer stacks above and below the dipole have arbitrary numbers of layers with arbitrary relative dielectric constants ($\kappa_i = n_i^2$) and thicknesses (L_i). The total physical thicknesses of the top and bottom cladding layers from the position of the dipole to the edge of the outermost layer are $L_{t\text{-}tot}$ and $L_{b\text{-}tot}$. For the sample calculations presented below, the inner cladding layers (shaded)

*Also at Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico 87131, USA. Email address: brueck@chtm.unm.edu

†Also at Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA.

 ©2003 The American Physical Society

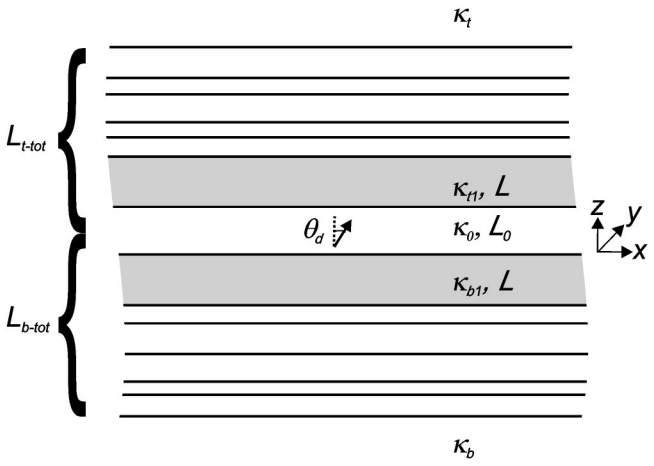


FIG. 1. Geometry of the problem. A radiating dipole oriented at an angle θ_d in the (x,z) plane is embedded in the center of a infinite slab of material of relative dielectric constant κ_0 and thickness L_0. The slab is bound with arbitrary film stacks ending in top and bottom semi-infinite media of relative dielectric constant κ_t and κ_b. For the example calculations, the shaded inner cladding thicknesses L are identical and variable. The total physical thicknesses of the upper and lower claddings from the position of the dipole to the edge of the semi-infinite media are indicated by $L_{t\text{-tot}}$ and $L_{b\text{-tot}}$.

have identical dielectric constants, $\kappa_{t1}=\kappa_{b1}=\kappa=n^2$, and an equal thickness, $L_{t1}=L_{b1}=L$ that is varied in the calculation. The thickness of the layer containing the dipole is denoted as L_0.

B. Hertz vector formalism

In our previous paper [4] we used the Hertz vector formalism introduced by Sommerfeld [1] to evaluate the radiation from a simple dielectric slab structure. Lukosz [9,10] has developed an alternate formulation of the Hertz vector source terms that has the significant advantage that the boundary conditions are uncoupled for the TE and TM radiation from a horizontal dipole. This dramatically simplifies the algebra of the calculation; the final results are identical. Lukosz's formulation is briefly described here and is followed in the evaluations of the radiated power.

Lukosz [9,10] has shown that the electromagnetic field for any orientation of the dipole can be represented by two scalar functions, $\phi(\vec{x})$ and $\psi(\vec{x})$, which are the z components (perpendicular to the layers) of an electric and a magnetic Hertz vector, respectively,

$$\vec{\Pi}^{(E)}(\vec{x})=(0,0,\phi(\vec{x})),$$

$$\vec{\Pi}^{(H)}(\vec{x})=(0,0,\psi(\vec{x})). \tag{1}$$

Both scalar functions satisfy the homogeneous Hemholtz equation in source-free regions, viz.,

$$(\nabla^2-\kappa_i\varepsilon_0k_0^2)\,\phi(\vec{x})=(\nabla^2-\kappa_i\varepsilon_0k_0^2)\,\psi(\vec{x})=0, \tag{2}$$

where ε_0 is the free-space permittivity, $\kappa_i=n_i^2$ is the relative dielectric permittivity of the medium, and $k_0=\omega/c$ is the magnitude of the free-space photon wave vector. The fields are given by

$$\vec{E}(\vec{x})=i\omega\vec{\nabla}\times\vec{\Pi}^{(H)}(\vec{x})+(\kappa\varepsilon_0)^{-1}\vec{\nabla}\times\vec{\nabla}\times\vec{\Pi}^{(E)}(\vec{x}),$$

$$\vec{H}(\vec{x})=-i\omega\vec{\nabla}\times\vec{\Pi}^{(E)}(\vec{x})+(\mu_0)^{-1}\vec{\nabla}\times\vec{\nabla}\times\vec{\Pi}^{(H)}(\vec{x}). \tag{3}$$

where ε_0 and μ_0 are the free-space values of the electric and magnetic permeabilities, respectively.

From Eqs. (1) and (3) it is straightforward to show that $\phi(\vec{x})$ corresponds to a TM wave and $\psi(\vec{x})$ describes a TE wave; since these are orthogonal there is no coupling between them. In the Sommerfeld basis [1], there is only an electric Hertz vector with components both in the z direction and in the x direction; the boundary conditions at interfaces are coupled which algebraically complicates the general solution for an arbitrary film stack.

C. Expansion of the dipole field into TE and TM plane waves

To begin the calculation, it is necessary to express the field of a dipole, located at the origin in an infinite medium, in a superposition of TE and TM plane waves. Then the boundary value problem for each independent plane wave can be directly solved by standard thin-film electromagnetic methods; finally the results are transformed back to real space using the techniques developed by Sommerfeld [1,2,11].

The infinite-medium real-space electric-Hertz vector corresponding to a dipole is simply

$$\vec{\pi}(\vec{x})=p_o\frac{e^{in_ik_0r}}{4\pi\varepsilon_0\kappa_ir}(\sin\theta_d\hat{e}_x+\cos\theta_d\hat{e}_z), \tag{4}$$

where p_o is the dipole moment, $(\hat{e}_x,\hat{e}_y)$ are unit vectors in the corresponding directions. The total radiated power for this dipole in an infinite medium is given by the well-known result [12]

$$P_{\text{inf}}=\frac{k_o^4\,\text{Re}(n_i)|p_o|^2}{12\pi\eta_0\varepsilon_0^2}, \tag{5}$$

where $\eta_0=\sqrt{\mu_0/\varepsilon_0}$. In the following, radiated powers are shown normalized to this value. The corresponding Fourier transform of the infinite-medium Hertz vector is

$$\vec{\Pi}_\infty^{(E)}(\vec{K})=\frac{p_o}{\varepsilon_0\kappa_ik_0^2}\frac{1}{K^2-\kappa_i}(\sin\theta_d\hat{e}_x+\cos\theta_d\hat{e}_z)$$
$$\equiv\Phi(K)(\sin\theta_d\hat{e}_x+\cos\theta_d\hat{e}_z), \tag{6}$$

where k_0K is the transform coordinate. This expression can be put in a more suitable form for the planar geometry of the problem by carrying out the inverse transform in the z direction and expressing the remaining integrals in cylindrical coordinates,

$$\Phi(\rho,z)=\frac{p_o}{\kappa_i}\int_0^\infty \frac{k_0}{\gamma_i}e^{-\gamma_i|\tilde{z}|}J_0(\lambda\tilde{\rho})\lambda d\lambda$$

$$=\frac{p_o}{2\kappa_i}\int_{-\infty}^{\infty}\frac{k_0}{\gamma_i}e^{-\gamma_i|\tilde{z}|}H_0^1(\lambda\tilde{\rho})\lambda d\lambda, \qquad (7)$$

where $k_0\lambda$ is the radial transform coordinate [$\lambda^2=\xi^2+\eta^2$ with $k_0\xi$ ($k_0\eta$) the $x(y)$-directed transform coordinate], $\tilde{\rho}=k_0\rho$ and $\tilde{z}=k_0z$ are the dimensionless radial and vertical spatial coordinates, and

$$\gamma_i=\sqrt{\lambda^2-\kappa_i}\xrightarrow[\lambda\to 0]{}-in_i, \qquad (8)$$

where the sign of the square root is chosen to ensure $\mathrm{Re}(\gamma_i)>0$ for convergence of Eq. (7). J_0 is the zeroth-order Bessel function and H_0^1 is the Hankel function of the first kind of zeroth order. The second form of this expression will be useful for evaluation of the complex integrations since the integration extends along the entire λ axis.

The scalars ϕ and ψ for the unbounded medium are found in terms of Φ by equality of the z components of the electric and magnetic fields from Eqs. (3) and (6), giving

$$\phi(\xi,\eta;z)=\Phi(\lambda;z)\left(\cos\theta_d\mp i\frac{\xi\gamma_i}{\lambda^2}\sin\theta_d\right) \qquad (9a)$$

and

$$\psi(\xi,\eta;z)=\Phi(\lambda;z)\left(\eta_0\frac{\eta}{\lambda^2}\sin\theta_d\right), \qquad (9b)$$

where the $(-)$ sign in Eq. (9a) is appropriate in the upper half plane ($z>0$) and the $(+)$ sign in the lower half plane ($z<0$). For a vertical dipole ($\sin\theta_d=0$) there are only TM waves (*only a single* scalar function $\phi(\xi,\eta;z)$ is required); for a horizontal dipole ($\sin\theta_d=1$) there are both TM and TE waves [*both* scalar functions $\phi(\xi,\eta;z)$ and $\psi(\xi,\eta;z)$ are required]. The source terms for a horizontal dipole ($\sin\theta_d=1$) depend on the transverse transform coordinates (ξ,η) which results in a dependence of the radiated fields on the polar angle φ where $\varphi=0$ is the radial direction parallel to the dipole.

III. EVALUATIONS OF THE RADIATED FIELDS AND POWER

A. Boundary value problem and evaluation of the Hertz potential; vertical dipole ($\cos\theta_d=1$)

The boundary conditions for ϕ at an interface between media 1 and 2 are [9]

$$\phi_1=\phi_2,$$

$$\frac{1}{\kappa_1}\frac{\partial\phi_1}{\partial z}=\frac{1}{\kappa_2}\frac{\partial\phi_2}{\partial z}. \qquad (10)$$

It is a standard multiple-interface thin-film matrix-analysis problem to solve for the Hertz potential everywhere. The result is

$$\phi_0^V=\frac{p_o}{2\kappa_1}\int_{-\infty}^{\infty}\frac{k_0}{\gamma_0}H_0^1(\lambda\tilde{\rho})\lambda d\lambda$$
$$\times\left[\frac{e^{-\gamma_0\tilde{L}_0}R_b^p(1+e^{-\gamma_0\tilde{L}_0}R_t^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}e^{-\gamma_0\tilde{z}}\right.$$
$$\left.+\frac{e^{-\gamma_0\tilde{L}_0}R_t^p(1+e^{-\gamma_0\tilde{L}_0}R_b^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}e^{\gamma_0\tilde{z}}+e^{-\gamma_0|\tilde{z}|}\right] \qquad (11)$$

for the slab region (subscript 0), and

$$\phi_t^V=\frac{p_o}{2\kappa_t}\int_{-\infty}^{\infty}\frac{k_0}{\gamma_t}H_0^1(\lambda\tilde{\rho})\lambda d\lambda$$
$$\times\left[\frac{T_t^pe^{-\gamma_0\tilde{L}_0/2}(1+e^{-\gamma_0\tilde{L}_0}R_b^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}\right]e^{-\gamma_t(\tilde{z}-\tilde{L}_{t-tot})},$$

$$\phi_b^V=\frac{p_o}{2\kappa_b}\int_{-\infty}^{\infty}\frac{k_0}{\gamma_b}H_0^1(\lambda\tilde{\rho})\lambda d\lambda$$
$$\times\left[\frac{T_b^pe^{-\gamma_0\tilde{L}_0/2}(1+e^{-\gamma_0\tilde{L}_0}R_t^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}\right]e^{+\gamma_b(\tilde{z}+\tilde{L}_{b-tot})} \qquad (12)$$

for the top [above the slab and top film stack (subscript t)] and bottom [below the slab and bottom film stack (subscript b)] semi-infinite outer cladding regions. In these expressions, $k_0^{-1}\tilde{L}_{t-tot}$ and $k_0^{-1}\tilde{L}_{b-tot}$ are the total physical thicknesses of the top and bottom cladding layers from the position of the dipole to the edge of the outer cladding regions as shown in Fig. 1. The Hertz potential within the film stacks is readily evaluated from these results. Here the R's and T's refer to the magnetic field reflectivity looking from inside the slab from the multilayer top (bottom) stacks for a TM wave propagating upwards (downwards) in the slab $R_t^p(R_b^p)$ and the T's are the corresponding transmissions through the entire film stacks. In the limit of a single dielectric interface at the top of the slab these reduce to the familiar results for a TM-wave incident from the slab towards the top dielectric (and similar expressions for the bottom interface)

$$R_t^p\to\frac{\kappa_t\gamma_0-\kappa_0\gamma_t}{\kappa_t\gamma_0+\kappa_0\gamma_t},$$

$$T_t^p=\frac{\kappa_t}{\kappa_0}(1-R_t^p)=\frac{2\kappa_t\gamma_t}{\kappa_t\gamma_0+\kappa_0\gamma_t}. \qquad (13)$$

Comparing this result with Eq. (10) of Ref. [1] for a simple dielectric slab, the only changes are (1) the substitution of the multilayer refection/transmission coefficients for their single interface values; (2) the separate identification of top and bottom reflection and transmission coefficients re-

quired for this generalized asymmetric calculation; and (3) setting the dipole position to the center of the slab ($a=0$ from Ref. [1]) since any shift of the dipole position can be accommodated by adding additional films (with the same dielectric constant). The bracketed terms have simple physical interpretations. There are two contributions to the fields in the top region, one direct transmission from the upward directed plane waves of the dipole and a second reflected from the bottom slab interface. The resonant denominator accounts for the multiple reflections within the slab. The zeros of this denominator correspond to poles of the integrand and, thus, the radiation into 2D bound waveguide modes. There are branch cuts in the λ plane as a result of the multivalued γ_i's. Integration around these branch cuts gives the 3D free-space radiation. The inverse transform will be discussed following the presentation of the boundary value solutions for the horizontal dipole in the next section.

B. Boundary value problem and evaluation of the Hertz potential; horizontal dipole ($\cos\theta_d=0$)

For a horizontal dipole, both ϕ and ψ must be evaluated. Since the boundary conditions are uncoupled, the solution proceeds straightforwardly using the source terms in Eq. (9) for $\theta_d=\pi/2$. The boundary conditions are [9]

$$\phi_1=\phi_2, \quad \psi_1=\psi_2,$$
$$\frac{1}{\kappa_1}\frac{\partial\phi_1}{\partial z}=\frac{1}{\kappa_2}\frac{\partial\phi_2}{\partial z}, \quad \frac{\partial\psi_1}{\partial z}=\frac{\partial\psi_2}{\partial z}. \tag{14}$$

This decoupling of the boundary conditions for the two scalar functions is a major algebraic advantage of the Hertz vector representation introduced by Lukosz [9]. The solutions again are obtained by straightforward thin-film analysis based on the boundary conditions Eq. (14) and the source terms Eq. (9),

$$\psi_0^H=\frac{p_0}{2}\eta_0\int_{-\infty}^{\infty}\frac{1}{\gamma_0}\sin\varphi H_0^1(\lambda\tilde{\rho})d\lambda \times\left[\frac{e^{-\gamma_0\tilde{L}_0}R_b^s(1+e^{-\gamma_0\tilde{L}_0}R_t^s)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^sR_t^s}e^{-\gamma_0\tilde{z}} + \frac{e^{-\gamma_0\tilde{L}_0}R_t^s(1+e^{-\gamma_0\tilde{L}_0}R_b^s)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^sR_t^s}e^{\gamma_0\tilde{z}}+e^{-\gamma_0|\tilde{z}|}\right],$$

$$\phi_0^H=\frac{ip_0}{2\kappa_0}\int_{-\infty}^{\infty}k_0\cos\varphi H_0^1(\lambda\tilde{\rho})d\lambda \times\left[\frac{e^{-\gamma_0\tilde{L}_0}R_b^p(1-e^{-\gamma_0\tilde{L}_0}R_t^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}e^{-\gamma_0\tilde{z}} - \frac{e^{-\gamma_0\tilde{L}_0}R_t^p(1-e^{-\gamma_0\tilde{L}_0}R_b^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}e^{\gamma_0\tilde{z}}(\mp)e^{-\gamma_0|\tilde{z}|}\right] \tag{15}$$

for the slab, and

$$\psi_t^H=\frac{p_0}{2}\eta_0\int_{-\infty}^{\infty}\frac{1}{\gamma_t}\sin\varphi H_0^1(\lambda\tilde{\rho})d\lambda \times\frac{T_t^se^{-\gamma_0\tilde{L}_{0'}/2}(1+e^{-\gamma_0\tilde{L}_0}R_b^s)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^sR_t^s}e^{-\gamma_t(\tilde{z}-\tilde{L}_{t-tot})},$$

$$\phi_t^H=\frac{-ip_0}{2\kappa_t}\int_{-\infty}^{\infty}k_0\cos\varphi H_0^1(\lambda\tilde{\rho})d\lambda \times\frac{T_t^pe^{-\gamma_0\tilde{L}_{0'}/2}(1-e^{-\gamma_0\tilde{L}_0}R_b^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}e^{-\gamma_t(\tilde{z}-\tilde{L}_{t-tot})}, \tag{16a}$$

$$\psi_b^H=\frac{p_0}{2}\eta_0\int_{-\infty}^{\infty}\frac{k_0}{\gamma_b}\sin\varphi H_0^1(\lambda\tilde{\rho})d\lambda \times\frac{T_b^se^{-\gamma_0\tilde{L}_{0'}/2}(1+e^{-\gamma_0\tilde{L}_0}R_t^s)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^sR_t^s}e^{+\gamma_b(\tilde{z}+\tilde{L}_{b-tot})},$$

$$\phi_b^H=\frac{ip_0}{2\kappa_b}\int_{-\infty}^{\infty}k_0\cos\varphi H_0^1(\lambda\tilde{\rho})d\lambda \times\frac{T_b^pe^{-\gamma_0\tilde{L}_{0'}/2}(1-e^{-\gamma_0\tilde{L}_0}R_t^p)}{1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p}e^{+\gamma_b(\tilde{z}+\tilde{L}_{b-tot})} \tag{16b}$$

for the top and bottom semi-infinite outer cladding regions. In these expressions, $R_{t,b}^s, T_{t,b}^s$ are the electric field reflectivities for TE polarized waves incident from the slab on the respective interfaces. In the limit of a simple dielectric interface these reduce for the top interface to

$$R_t^s=\frac{\gamma_0-\gamma_t}{\gamma_0+\gamma_t}, \quad T_t^s=1-R_t^s=\frac{2\gamma_t}{\gamma_0+\gamma_t}. \tag{17}$$

In Eq. (15), the minus sign in the parentheses is appropriate for $\tilde{z}>0$ and the plus sign for $\tilde{z}<0$. As mentioned above, the polar angle φ, defined relative to the orientation of the dipole in the slab, arises from the presence of the transverse transform coordinates (ξ,η) in the source terms for the horizontal dipole.

The physical interpretation of these equations is similar to that for the vertical dipole. Now the coupling is to both TE and TM radiation and bound modes. The zeros of the denominators correspond to the bound modes (TM for $1-e^{-2\gamma_0\tilde{L}_0}R_b^pR_t^p=0$; TE for $1-e^{-2\gamma_0\tilde{L}_0}R_b^sR_t^s=0$). In each numerator there is a term that corresponds to the radiation directly incident from the dipole and a second term from the radiation reflected from the opposite interface. For the potential within the slab there are upward and downward reflected

waves and the source terms. For the potential in the half-spaces the numerators correspond to transmission from the source at the center of the slab. The denominators account for the multiple reflections, and include all near field terms; no approximations to the full electromagnetic calculation have been made to this point. These approximations will be made in the branch cut integrals, which are evaluated by the method of steepest descents that is only valid far (many wavelengths) from the source point, and in the evaluation of the contributions from the poles using the asymptotic limits of the Hankel function, again valid many wavelengths from the source point. There are branch cuts corresponding to each of the thin film layers. In the half-spaces above and below the structure, only the integral around the branch cut corresponding to that specific medium contributes to the radiation fields; the other branch cuts contribute only to the near fields and thus are excluded from the subsequent analysis.

C. Evaluation of the inverse transform

The apparatus of a complex analysis can now be applied to the remaining integrals. As noted above, there are contributions due to integration around the branch cuts (corresponding to 3D radiation into the semi-infinite half-spaces above and below the structure) and due to poles of the denominators (2D bound modes). The manipulations have been presented for a single interface in great detail in Ref. [11] for a single interface and in Ref. [4] for the simple slab case. Only the final results will be presented here.

1. Vertical dipole

2D bound modes. For the vertical dipole, the Hertz potential within the slab is evaluated using the asymptotic limit of the Hankel function $[H_0^1(z)\rightarrow -ie^{iz}\sqrt{2/\pi z}]$ and taking the residues at the zeros of the denominator,

$$\phi_0^{2D}=p_0k_0\sum_l \sqrt{\frac{8\pi}{k_l^p\tilde{\rho}}}e^{ik_l^p\tilde{\rho}}\frac{k_l^p}{\kappa_0\gamma_{0,l}^p\delta D_l^p} \times[(1+R_{sym}^p e^{-\gamma_{0,l}^p\tilde{L}_0})\cosh(\gamma_{0,l}^p\tilde{z}) + R_{asym}^p e^{-\gamma_{0,l}^p\tilde{L}_0}\sinh(\gamma_{0,l}^p\tilde{z})], \tag{18}$$

where the denominator is expanded around the roots (k_l^p, where the subscript is the mode index) of the TM (p superscript) modal dispersion relation, viz.,

$$D_l^p=1-e^{-2\gamma_1\tilde{L}_0}R_b^pR_t^p\approx(\lambda-k_l^p)\left.\frac{\partial D_p}{\partial k}\right|_{k_l^p}\equiv(\lambda-k_l^p)\,\delta D_l^p \tag{19}$$

and

$$R_{sym}^p\equiv\frac{R_t^p+R_b^p}{2},\quad R_{asym}^p\equiv\frac{R_t^p-R_b^p}{2}. \tag{20}$$

The superscript on the γ's takes note of the fact that they are evaluated at the roots of the TM dispersion relation and the subscripts on the γ's refer to the layer index (j) and the mode index (l) [$\gamma_{j,l}^p\equiv\gamma_j(\lambda)|_{\lambda=k_l^p}$]. The corresponding indices on the reflection coefficients are suppressed for convenience.

From this result, the next step is to evaluate the fields and to integrate across the mode profile to get the radiated power with the result for the central slab containing the dipole,

$$\hat{P}_{V,TM}^{2D}=\frac{3\pi k_0}{\mathrm{Re}(\kappa_1)\sqrt{\varepsilon_0}}\sum_l\frac{|k_l^{p\prime}|^4}{|\delta D_l^p|^2} \times\left[\begin{array}{l}[|c_{0,l}|^2+|d_{0,l}|^2]\dfrac{\sinh(\gamma_{0,l}^{p\prime}\tilde{L}_0/2)}{\gamma_{0,l}^{p\prime}} \\ +[c_{0,l}d_{0,l}^*+c_{0,l}^*d_{0,l}]\dfrac{\sin(\gamma_{0,l}^{p\prime\prime}\tilde{L}_0/2)}{\gamma_{0,l}^{p\prime\prime}}\end{array}\right], \tag{21}$$

where the summation is over all modes (l), $\gamma_{0,l}^p=\gamma_{0,l}^{p\prime}+i\gamma_{0,l}^{p\prime\prime}$, and the coefficients $c_{0,l}$ and $d_{0,l}$ are the complex amplitudes of the upward and downward directed fields in each layer at the center of the slab. From Eq. (18)

$$\begin{aligned}c_{0,l}&=\frac{1}{2}(1+R_{sym}^p e^{-\gamma_{0,l}\tilde{L}_0}+R_{asym}^p e^{-\gamma_{0,l}\tilde{L}_0})\\&=\frac{1}{2}(1+R_t^p e^{-\gamma_{0,l}^p\tilde{L}_0}),\\ d_{0,l}&=\frac{1}{2}(1+R_{sym}^p e^{-\gamma_{0,l}\tilde{L}_0}-R_{asym}^p e^{-\gamma_{0,l}\tilde{L}_0})\\&=\frac{1}{2}(1+R_b^p e^{-\gamma_{0,l}^p\tilde{L}_0}).\end{aligned} \tag{22}$$

Similar expressions need to be evaluated over the entire film stack, including the top and bottom semi-infinite regions, and summed to get the total 2D radiated power.

3D radiated energy. The calculation proceeds by substituting the integral expression for the Hankel function

$$H_0^1(z)=\frac{4}{\pi}e^{iz}\int_0^\infty(4iz-y^2)^{-1/2}e^{-y^2/2}dy\approx e^{iz}\sqrt{\frac{2}{i\pi z}}, \tag{23}$$

where the last approximation is valid in the radiation zone, many wavelengths from the dipole. Then the Hertz potential in the semi-infinite cladding regions can be evaluated by integrating around the branch cut by the method of steepest descents with the result for the top half-space:

$$\phi_t=\frac{p_0k_0e^{in_t\tilde{R}}}{\tilde{R}}\frac{1-R_t^p}{\kappa_0}\frac{[1+e^{in_0\tilde{L}_0\cos(\vartheta)}R_b^p(\theta)]}{1-e^{2in_0\tilde{L}_0\cos(\vartheta)}R_t^p(\theta)R_b^p(\theta)} \times e^{i(n_0\cos\vartheta-n_t\cos\theta)\tilde{L}_0/2}, \tag{24}$$

where $\tilde{R}$ is the dimensionless radial distance coordinate, ϑ is the propagation angle inside the slab for an external angle of θ [$\cos^2\vartheta\equiv1-(\kappa_t/\kappa_1)\sin^2\theta$] and the reflection coefficients are evaluated at the external angle θ.

For the radiated power, normalized to the power radiated by the dipole in an infinite medium, the final result is

$$\hat{P}_{TM}^{V}=\frac{3}{4}\left|\frac{\kappa_t}{\kappa_0}\right|^{5/2}\int_0^{\pi/2}\left|\frac{[1-R_t^p(\theta)][1+e^{in_0\tilde{L}_0\cos\vartheta}R_b^p(\theta)]e^{i(n_0\cos\vartheta-n_t\cos\theta)\tilde{L}_0/2}}{1-e^{2in_0\tilde{L}_0\cos\vartheta}R_t^p(\theta)R_b^p(\theta)}\right|^2\sin^3\theta d\theta \tag{25}$$

with a similar expression for the power radiated into the bottom half-space with the obvious interchange ($t\Leftrightarrow b$). As expected, this result is identical to that for the simple slab [4] with the substitution of the film-stack reflectivities for those of the simple dielectric interface and the generalization to an asymmetric cladding structure.

2. Horizontal dipole

A horizontal dipole couples to both TE and TM radiation. The calculations are very similar to those presented above and only the results are given here.

2D bound modes. For TE coupling

$$\psi_0^{2D,H}=\sin\varphi p_0\eta_0\sum_l\sqrt{\frac{8\pi}{k_l^s\tilde{\rho}}}e^{ik_l^s\tilde{\rho}}\frac{1}{\delta D_l^s}\times[(1+R_{sym}^s e^{-\gamma_{0,l}^s\tilde{L}_0})\cosh(\gamma_{0,l}^s\tilde{z})+R_{asym}^s{}^{-\gamma_{0,l}^s\tilde{L}_0}\sinh(\gamma_{0,l}^s\tilde{z})], \tag{26}$$

$$\hat{P}_{TE}^{2D,H}=\frac{3\pi}{4\,\mathrm{Re}(n_0)}\sum_l\left|\frac{k_l^s}{\delta D_l^s}\right|^2\times\left[\begin{array}{l}[|c_{0,l}|^2+|d_{0,l}|^2]\dfrac{\sinh(\gamma_{0,l}^{s\prime}\tilde{L}_0/2)}{\gamma_{0,l}^{p\prime}}\\ +[c_{0,l}d_{0,l}^*+c_{0,l}^*d_{0,l}]\dfrac{\sin(\gamma_{0,l}^{s\prime\prime}\tilde{L}_0/2)}{\gamma_{0,l}^{p\prime\prime}}\end{array}\right],$$

$$c_{0,l}=\frac{1}{2}(1+R_t^s e^{-\gamma_{0,j}^s\tilde{L}_0}),$$

$$d_{0,j}=\frac{1}{2}(1+R_b^s e^{-\gamma_{0,j}^s\tilde{L}_0}). \tag{27}$$

For TM coupling

$$\phi_0^{2D,H}=\cos\varphi p_0 k_0\sum_l\sqrt{\frac{8\pi}{k_l^p\tilde{\rho}}}e^{ik_l^p\tilde{\rho}}\frac{1}{\delta D_l^p}\times[(1-R_{sym}^p e^{-\gamma_{0,l}^p\tilde{L}_0})\sinh(\gamma_{0,l}^p\tilde{z})-R_{asym}^p e^{-\gamma_{0,l}^p\tilde{L}_0}\cosh(\gamma_{0,l}^p\tilde{z})] \tag{28}$$

$$\hat{P}_{TM}^{2D,H}=\frac{3\pi}{4\,\mathrm{Re}(\kappa_0^{3/2})}\sum_l\left|\frac{k_l^p}{\delta D_l^s}\right|^2\times\left[\begin{array}{l}[|c_{0,l}|^2+|d_{0,l}|^2]\dfrac{\sinh(\gamma_{0,l}^{p\prime}\tilde{L}_0/2)}{\gamma_{0,l}^{p\prime}}\\ +[c_{0,l}d_{j,l}^*+c_{0,l}^*d_{j,l}]\dfrac{\sin(\gamma_{0,l}^{p\prime\prime}\tilde{L}_0/2)}{\gamma_{0,l}^{p\prime\prime}}\end{array}\right],$$

$$c_{0,j}=\frac{1}{2}(1-R_t^p e^{-\gamma_{0,j}^p\tilde{L}_0}),$$

$$d_{0,j}=-\frac{1}{2}(1-R_b^p e^{-\gamma_{0,j}^p\tilde{L}_0}). \tag{29}$$

3D radiation into top half-space. For TE coupling

$$\psi_t^H=p_0k_0\eta_0\frac{\sin\varphi}{\sin\theta}\left(\frac{e^{ik_t\tilde{R}}}{n_t\tilde{R}}\right)T_t^s\frac{1+R_b^s e^{in_0\tilde{L}_0\cos\vartheta}}{1-R_t^sR_b^s e^{2in_0\tilde{L}_0\cos\vartheta}}\times e^{i(n_0\cos\vartheta-n_t\cos\theta)\tilde{L}_0/2}, \tag{30}$$

$$\hat{P}_{TE,t}^{3D,H}=\frac{3}{8}\mathrm{Re}\left(\frac{n_t}{n_0}\right)\int_0^{\pi/2}\times\left|T_t^s\frac{1+R_b^s e^{in_0\tilde{L}_0\cos\vartheta}}{1-R_t^sR_b^s e^{2in_0\tilde{L}_0\cos\vartheta}}e^{i(n_0\cos\vartheta-n_t\cos\theta)\tilde{L}_0/2}\right|^2\times\sin(\theta)d\theta. \tag{31}$$

For TM coupling

$$\phi_t^H=p_0k_0\frac{\cos\theta\cos\varphi}{\sin\theta}\left(\frac{e^{in_t\tilde{R}}}{n_t\tilde{R}}\right)T_t^p\frac{-1+R_b^p e^{in_0\tilde{L}_0\cos\vartheta}}{1-R_t^pR_b^p e^{2in_0\tilde{L}_0\cos\vartheta}}\times e^{i(n_0\cos\vartheta-n_t\cos\theta)\tilde{L}_0/2}, \tag{32}$$

$$\hat{P}_{TM,t}^{3D,H}=\frac{3}{8}\mathrm{Re}\left(\frac{n_t}{n_0}\right)\int_0^{\pi/2}\times\left|n_tT_t^p\frac{-1+R_b^p e^{in_0\tilde{L}_0\cos\vartheta}}{1-R_t^pR_b^p e^{2in_0\tilde{L}_0\cos\vartheta}}e^{i(n_0\cos\vartheta-n_t\cos\theta)\tilde{L}_0/2}\right|^2\times\cos^2(\theta)\sin(\theta)d\theta. \tag{33}$$

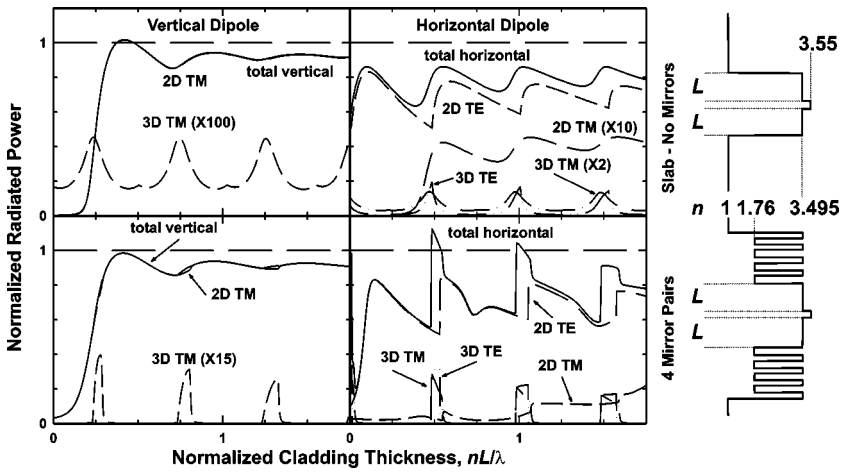


FIG. 2. Radiated power from a dipole embedded in a slab (top) and in the same slab with four mirror stacks (bottom). The results for a vertical dipole are shown in the left panels and for a horizontal dipole in the right panels. In each case the dipole is embedded in a thin (10-nm-wide) InGaAs ($n=3.55$) section clad top and bottom with GaAs ($n=3.495$) spacers of thickness L. For the slab, the semi-infinite outer claddings are air ($n=1$); for the mirror stacks are $\lambda/4$ pairs of Al_2O_3 ($n=1.76$) and GaAs, again with semi-infinite air outer claddings. See text for detailed discussion.

with equivalent expressions for the radiation into the bottom half space.

Again, it is instructive to compare these expressions for a horizontal dipole with the equivalent expressions derived for the symmetrically clad slab with Sommerfeld's Hertz vector approach [4]. For the TE coupling the results are again very similar with the straightforward substitutions of the multilayer reflectivities, the more careful tracking of reflection coefficients necessary for nonequivalent surfaces, and the automatic placement of the dipole in the center of the slab. The TM results are significantly simpler in this formulation, particularly for the Hertz potential and the 3D radiation contributions. In the previous treatment, based on the Sommerfeld formulation, the possible algebraic simplification was only recognized for the bound mode contributions.

IV. RESULTS

In the following two sections, results using these expressions are presented for two related vertical-cavity structures. First, we consider a slab bound by two symmetric quarter wavelength mirror stacks with air on both sides. The power is radiated into both bound modes and 3D radiation. In the second case, the same structures are atop a GaAs substrate, again with air cladding the top surface. Now there are no bound modes, but there are ''leaky'' modes with field distributions very similar to the bound modes, but whose energy leaks into radiation modes in the substrate. Absorption has been set to zero (all indices taken as pure real quantities), interface roughness has been neglected, and the structure is taken as infinite within the plane. Additionally, the dipole radiation has been assumed to be at a continuous single frequency ω, without any dephasing (T_2) or decay (T_1) events. The impact of all of these assumptions is to allow the coherent addition of fields over many transits of the multilayer structure, e.g., to investigate very high finesse resonances. This allows beautiful examples of resonance effects; however, due consideration of these limiting effects must be taken in applying this analysis to the interpretation of experimental results.

A. Slab with DBR mirrors—bound modes and radiation

The slab structure consists of a thin (10 nm) InGaAs quantum well ($n=3.55$) with the dipole in the center of the well, symmetric GaAs inner cladding layers ($n=3.495$) whose thicknesses are equal and are varied in the calculation, symmetric $\lambda/4$ Al_2O_3/GaAs ($n=1.76/3.495$) mirror pairs (results for zero to four mirror pairs are presented), and air ($n=1$) upper and lower outer cladding half-spaces. The spatial index profile is shown as an inset in Fig. 2. These index values are appropriate to a wavelength of 980 nm; results are shown normalized to the wavelength in the GaAs layers.

The calculated radiated power, normalized to the radiated power in an unbounded InGaAs medium as a function of the GaAs inner cladding thickness for a simple slab (top, no mirror stacks) and for a slab bound on each side by four mirror stacks (bottom) is shown in Fig. 2 for both a vertical dipole source (left) and a horizontal dipole source (right). For the vertical dipole source, all of the radiation is into TM bound and radiation modes. For the air-clad slab, the 3D radiation shows strong Fabry-Perót resonances for slab thicknesses near $2n_1\tilde{L}\sim(2j+1)\lambda/2$, $j=0,1,\ldots$, and much weaker resonances for even numbers of half-wave resonances. This alternation arises from the factoring of the resonant denominator for this symmetric case and the cancellation of one set of resonances by the numerator in Eq. (25). The horizontal dipole has a similar set of resonances but now at even numbers of half-waves and the cancellation for odd numbers of half-waves, for both the TE and TM radiation components. These resonance conditions are a direct consequence of the boundary conditions in Eqs. (10) and (14). The TM radiation boundary conditions for a vertical dipole require that the Hertz potential be zero at the edges of the inner cladding as the magnitude of the reflectivity approaches unity, whereas the TE and TM boundary conditions for a horizontal dipole force the derivative of the potential to zero at the same locations. The high symmetry of a dipole located in the precise center of the slab results in these alternating resonance selection rules; more complex behavior is found for a dipole displaced from the center of the slab. For the 2D bound modes, the vertical dipole exhibits a strong suppression of the radiation for very thin slabs ($L\ll\lambda/n$) while the radiation is more pronounced for the horizontal dipole at these thicknesses. In both cases, the total radiated power rises approximately to that for the dipole in an infinite medium as the inner cladding thickness is increased; the slightly lower value (<1) than in an infinite InGaAs medium is be-

cause the normalization is to the dipole in the higher index InGaAs, but the majority of the bound mode radiation is confined to the GaAs layers. Each of the contributions to the total radiated power has been multiplied by various factors to make them visible in the figure. As is well known, for the vertical dipole, only very weak radiation into the air spaces is observed ($3D\times100$). More interesting is the result for the horizontal dipole, where most of the radiation is into the TE mode; the 2D TM radiation is multiplied by 10, while the 3D TM radiation contribution to the total power is multiplied by 2.

The impact of the mirror pairs on the distribution of power between the individual radiation components is significant, whereas the total radiated power is only modestly impacted. For a vertical dipole, there is a noticeable increase in the radiated power into the 2D TM mode for very thin slabs. This is a result of the increased mode volume due to the mirrors (see the left side of Fig. 2). For the air-clad slab, only the lowest-order mode is propagating for $2n_1L\leqslant\lambda$ while for the four-mirror-clad slabs, there are six propagating modes even for $\tilde{L}=0$. For both dipole orientations, the 3D radiation is significantly increased at the alternating half-wave resonances, and suppressed away from these resonances. For a horizontal dipole at the first two resonances, the total radiated power is slightly larger ($\sim15\%$) than that for a dipole in an infinite medium corresponding to a decrease in the radiative lifetime. The 3D radiation is strongly peaked at the even integral resonances for both the TE and TM components. Away from these resonances, essentially all of the radiated power is in the 2D bound modes. As the number of mirror pairs is increased, more of the energy is shifted into the TM radiation, although it remains a significantly smaller fraction of the total power than that radiated into the TE mode. There is a more equal balance of radiated power between TE and TM for the 3D radiation.

The evolution of the total radiated power as the number of mirror pairs is shown more clearly in Fig. 3. Results are shown for both vertical (top) and horizontal (bottom) dipoles for the slab case and for one through four mirror pairs. The curves are offset for clarity. In each case, the radiation from a dipole in an infinite medium is shown as the associated dotted line.

For a vertical-cavity laser, the radiated power for a horizontal dipole in the normal direction, perpendicular to the layers, is of particular interest. The maximum intensity into the normal direction occurs when the TE and TM powers are equal, just at the leading edge of the resonances of Fig. 2 (see also Fig. 5B). Figure 4 (top) shows the normalized forward direction emission for a horizontal dipole as a function of the slab thickness for the five cases of a slab and one through four mirror pairs. The angular variation at the peak is also shown in Fig. 4 (bottom). The variation with slab thickness shows the expected Lorentzian line shape from the expansion of the resonant denominator at an increasing cavity finesse F as the reflectivity increases with the number of mirror pairs. (The calculated normal emission power reflectivities and corresponding F's are $|R|^2=0.31$, $F\sim2.5$, slab; $|R|^2=0.75$, $F\sim11$, one mirror pair; $|R|^2=0.93$, $F\sim43$, two mirror pairs; $|R|^2=0.982$, $F\sim173$, three mirror pairs; and $|R|^2=0.995$, $F=628$, four mirror pairs.) The slight shift of the resonance away from $2nL/\lambda=1$ is due to the optical thickness of the InGaAs quantum well. As ex-

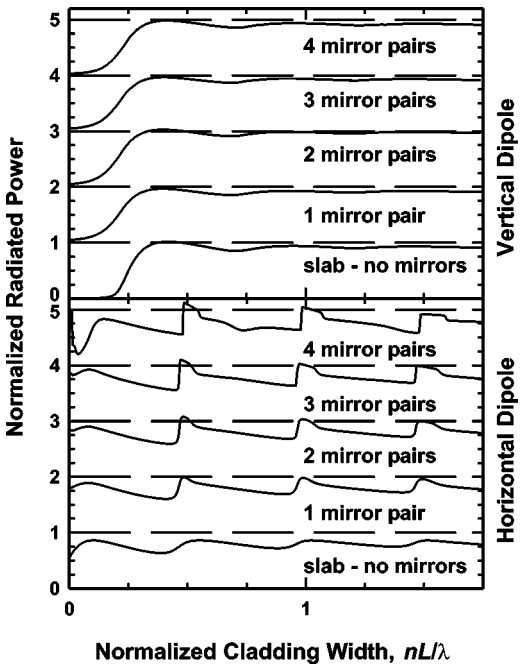


FIG. 3. Evolution of the total radiated power for a symmetric structure as the number of mirror pairs is varied. The curves are offset for clarity; in each case the dashed line represents the power radiated by the dipole into an infinite InGaAs medium.

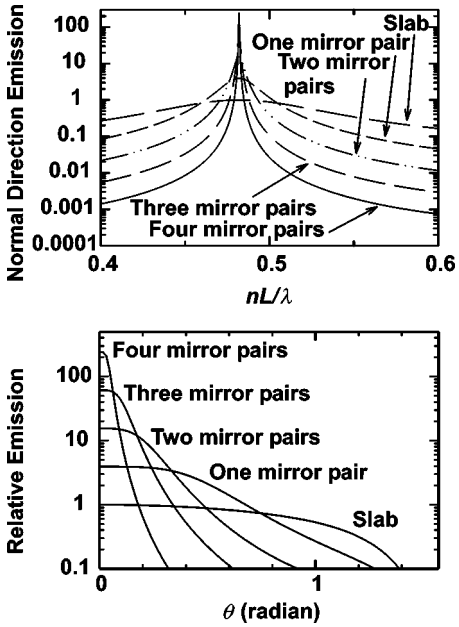


FIG. 4. Vertical emission (normal to the layer structure) into 3D radiation modes for a horizontal dipole embedded in a symmetrically clad slab as a function of the inner cladding thickness (top) and the angular variation for the peak vertical emission (bottom), both with the number of mirror pairs as a parameter. As expected, the increasing finesse of the resonance with increasing numbers of mirror pairs is reflected in both plots.

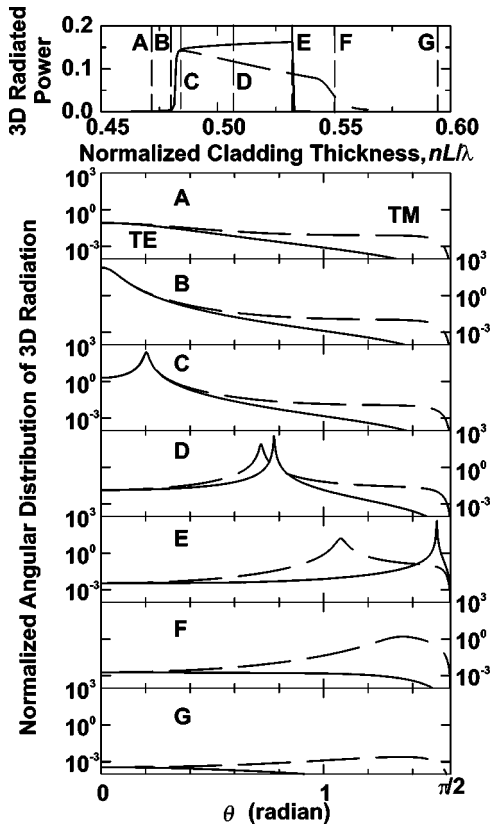


FIG. 5. Angular variation of the 3D radiation from a horizontal dipole in a symmetrically clad slab with four mirror pairs as the inner cladding thickness is varied. The top panel shows an expanded view of the integrated TE and TM 3D radiation in the vicinity of the first-resonance inner cladding thickness. The angular traces (A though G) show the evolution of the angular dependence of the emission as the inner cladding thickness is varied in this region.

pected, the angular variation of the emission at the peak of each of these resonances narrows significantly as F increases.

As the inner cladding thickness L increases beyond this resonance, the peak of the 3D emission shifts to steeper angles for both TE and TM radiation, and the total 3D radiated power increases dramatically. The top panel of Fig. 5 shows an expanded view of the total integrated 3D radiated power emitted from both surfaces around the first peak region ($nL/\lambda \sim 0.5$) for the four-mirror-pair case for a horizontal dipole. From the symmetry of this geometry, the same power is emitted from both surfaces. The angular dependence of the radiated power is shown for various values of L (A through G) as indicated in the top panel. Case A is before the resonance; the radiated power is quite small and does not show a pronounced angular dependence. Case B is the maximum of the normal-direction emitted power shown previously in Fig. 4. Notice that this peak normal emission corresponds to a very small fraction of the power radiated from the dipole. Case C is just past the sharp increase in the 3D emitted power for both TE and TM radiation. There is a peak in the angular dependence, for both TE and TM radiation, at a small angle away from normal. This peak can be traced to the round-trip resonance condition in the denominators [Eqs. (31) and (33)] that must move away from normal as the thickness is increased. The explicit dependence in Eqs. (31) and (33) is on the thickness of the central slab medium containing the dipole, which is fixed in this calculation. The dependence on the thickness of the cladding is contained in the phase of the reflection coefficients. This lowest order round trip resonance condition is simply, $4n_{GaAs}\cos\vartheta_{GaAs}L + 2n_{InGaAs}\cos\vartheta L_{InGaAs} + 2\phi(\theta) = \lambda/2\pi$, which clearly exhibits the dependence of ϑ_{GaAs}, the angle of propagation in the inner cladding, on L. The final phase term $\phi(\theta)$ represents the phase shift in the reflector stack beyond the inner cladding. The external angle θ is related to ϑ_{GaAs} by the Fresnel condition $n_{GaAs}\sin\vartheta_{GaAs} = \sin\theta$ for the present case of a semi-infinite air ($n_t = 1$) outer cladding. The change in the external angle for a given change in L is amplified by the GaAs refractive index. For case D, corresponding roughly to the middle of the thickness region with substantial radiation, the peaks for TE and TM radiation occur at different angles and the TE peak is significantly sharper. This is because of the lower TM reflectivity related to Brewster's angle for TM radiation incident on a single interface. The splitting is due to phase shifts as the angle is increased. This trend is continued for case E, just before the cutoff of the TE radiation. The angular peak in the TE emission is close to $\pi/2$. For a further increase in the inner cladding thickness, the peak shifts beyond the angle for total internal reflection and the emitted power shifts from 3D radiation out the sides of the structure to 2D modes propagating within the structure. This same transition is more gradual for the TM radiation (case F) consistent with the more gradual onset of the 2D TM emission with inner cladding thickness seen in Fig. 2. Finally, case G is for a thickness beyond the significant 3D emission region where the total 3D power is weak and again does not show any strong angular dependence.

B. Active region and mirrors on substrate—No bound modes

We now make a change in the structure, replacing the bottom semi-infinite air cladding with a semi-infinite GaAs cladding. This simple change greatly modifies the calculation, but of course the results should not be dramatically different since, at least as the number of mirrors is increased, the dipole should not ''see'' the substrate refractive index. The major change in the calculation is that there are no bound modes in this structure, in the sense that the zeros of the dispersion relation all correspond to ''modal'' indices that are lower than the GaAs substrate refractive index for the entire thickness region L investigated. Thus the radiated power is either emitted out the top surface, or ''leaks'' from the film stack into the GaAs substrate. As before, some of the power is radiated into the upper air space. Mathematically, having the substrate index higher than the solution to the dispersion relation puts the poles on a different sheet of the Riemann surface of the inverse transform integrand (e.g., the negative sign must be taken for the square root corresponding to the substrate propagation constant whereas a positive sign was taken in the true bond mode case) so that the poles do not contribute to the inverse transform integration.

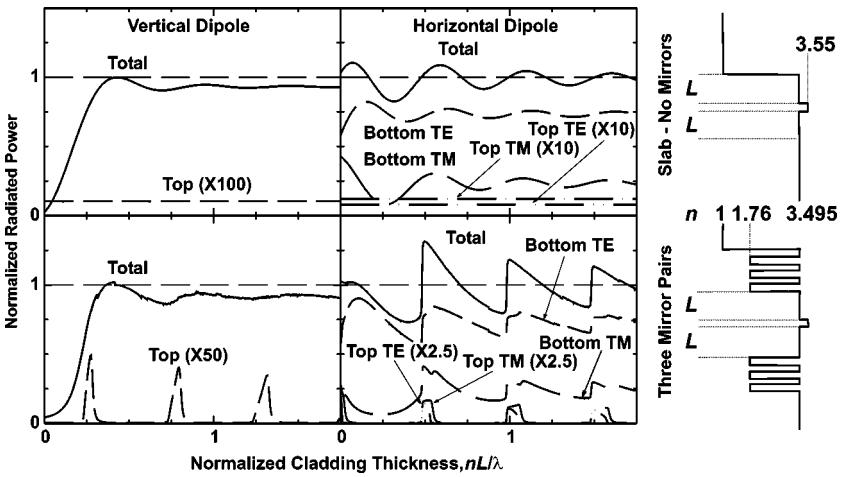


FIG. 6. Radiated power from a dipole embedded in an asymmetric film structure with air upper and GaAs lower outer claddings. The top panels are for a simple InGaAs/GaAs structure, the bottom panels include three Al_2O_3/GaAs $\lambda/4$ mirror pairs. In each case the dipole is embedded at the center of a thin InGaAs region. The left panels are for a vertically oriented dipole; the right for a horizontal dipole. The only emission is into 3D radiation for these cases. There are no strictly bound modes, however, for the three mirror-pair case; there are ''leaky'' modes largely confined to the film stack but with radiation that leaks into the substrate.

Figure 6 shows the total radiated power for a simple slab case (top), with just a buried InGaAs layer, and for three mirror pairs (bottom). Results are shown for both a vertical dipole (left) and a horizontal dipole (right). As there are no bound modes, there are no 2D contributions; the calculation proceeds only from Eqs. (31) and (33) and their bottom surface counterparts. For a vertical dipole, the results are very similar to those for the symmetrically clad slab with air on both sides. There are somewhat larger differences for a horizontal dipole. The roughly factor of two decrease compared with the equivalent figure for the symmetrically clad geometry is due to the counting of the radiated power out of both top and bottom surfaces in Figs. 2 and 5, whereas the top and bottom surface radiated powers are shown individually in this figure. There are differences in the details of the emission dependence on inner cladding thickness that will be discussed in conjunction with Figs. 8 and 9. The total horizontal dipole emission does not exhibit the sharp transitions between 2D and 3D emission that characterized the symmetric air-clad slab, but the overall dependence of the emission on the inner cladding thickness is similar. The total emission at the peak around $nL/\lambda \sim 0.5$ is larger than for the similar symmetric air-clad three-mirror-pair case. There is also a significant change in the distribution of energy between TE and TM emission, the TM emission being more pronounced in this substrate geometry.

Figure 7 shows the progression of the total emitted power as a function of the inner cladding thickness with the number of mirror pairs as a parameter. The curves have been offset for clarity. The radiated power for a dipole in an infinite medium is shown as the dashed line associated with each curve.

An expanded view of the first top surface emission region for a horizontal dipole, in the vicinity of $nL/\lambda \sim 0.5$, is shown in the top panel of Fig. 8. This should be compared with the comparable panel in Fig. 5 for the symmetrically air-clad slab. Again, the overall results are similar, but there are some noticeable differences. The overall shapes are inverted, in Fig. 5 the TE power increases slightly as the inner cladding thickness increases up to a very sharp cutoff while the TM power decreases and shows a much more gradual cutoff. In Fig. 8 the TM power increases while the TE power decreases as the inner cladding thickness increases. The limits of the emission with inner cladding thickness are more gradual and less well defined for this case. The TM emission is larger than the TE, the opposite of the symmetrically clad slab case.

The angular dependencies of the emission out the top (air-clad) surface are shown in panels A through G. Panel A is for inner cladding thicknesses smaller than those that correspond to the peak emission. The emission is relatively weak and featureless. Panel B corresponds to the peak of the normal emission. Notice that this peak occurs for a thickness further up the knee of the emission than was the case for the symmetrically air-clad slab (Fig. 5B). Nonetheless, the emission is weaker in the present asymmetric case. In comparable units, the emission at normal for the air-clad case with three mirror pairs was 60, both top and bottom; in this case it is 11

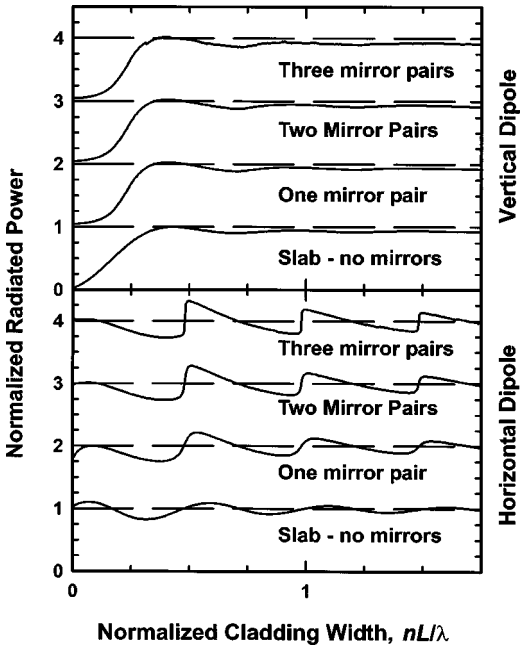


FIG. 7. Evolution of the total radiated power for an asymmetrically clad slab with an air upper and a GaAs substrate lower cladding. The curves have been offset for clarity; in each case the dashed line represents the power radiated by the dipole into an infinite InGaAs medium.

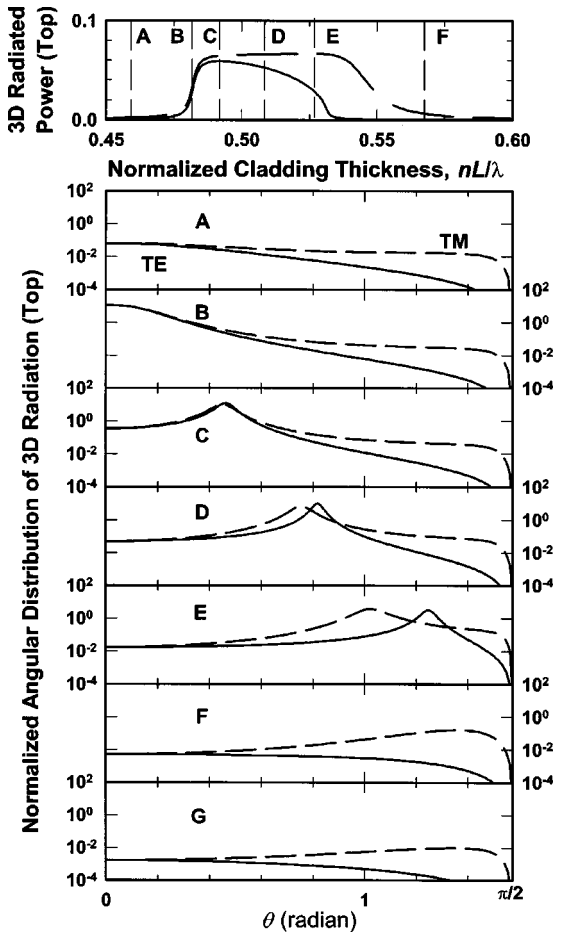


FIG. 8. Angular variation of the 3D radiation (horizontal dipole) from the top (air) surface of an asymmetrically clad slab (air upper and GaAs lower outer claddings) with three mirror pairs as the inner cladding thickness is varied. The top panel shows an expanded view of the integrated TE and TM 3D radiation in the vicinity of the first-resonance inner cladding thickness. The angular traces (A though G) show the evolution of the angular dependence of the emission as the inner cladding thickness is varied in this region. Curve G is for a normalized inner cladding thickness $nL/\lambda \sim 0.65$, beyond the thicknesses plotted in the top curve.

top and 134 bottom. Overall the normal emission is comparable, but much is lost to substrate emission in this asymmetric case. Panel C is for an inner cladding thickness near the peak TE emission. Both TE and TM emissions show comparable peak emission angles as was the case for Fig. 5. Panel D is for an inner cladding thickness near the mid point of the significant top-surface emission. As was the case for the symmetrically clad structure, at these higher angles a splitting is observed between the TE and TM peak emission angles as a result of different phase contributions from the TE and TM reflectivities. Panel E is for a thickness near the upper edge of the TE emission. The peak of the TE emission is close to $\pi/2$ as expected. Panels F and G are at the upper edge of the TM emission and beyond both TE and TM emission regions. Panel G is for a nL/λ of 0.65, outside the range of the top panel, which was plotted on the same horizontal scale as Fig. 5 for comparison. Since these curves are for emission out the top surface, the same amplification of the external angle relative to the internal angle as was found for the previous symmetrically clad structure is also found in these cases.

In contrast to the relatively orderly behavior of the top surface emission shown in Fig. 8, the bottom surface emission shows a much more pronounced angular dependence as shown in Fig. 9 for five different values of the normalized inner cladding thickness, again for a horizontal dipole. Note the much larger vertical axis scale from 10^{10} to 10^{-5}. The TE and TM emission angular dependencies are shown side by side for clarity in this figure. The top pair of panels is for the same inner cladding thickness as panel B of Fig. 8, just at the peak of the normal emission. In addition to the angular peak at $\phi=0$, both the TE (left) and TM (right) emission show additional structure. There are sharp dips, which are relatively independent of inner cladding thickness and are related to changes in reflectivity. A notable feature is the very strong peak in the TE emission at $\phi \sim 1.25$ (labeled TE1) and the comparable, but significantly smaller peak in the TM emission at $\phi \sim 0.66$ (TM1). These correspond to energy transmitted into the substrate from the lowest order ''leaky'' mode. That is from a field distribution in the upper cladding layers that looks very similar to the zero-order bound mode for the symmetrically air-clad structure. Compared to the peak normal emission of ~ 134 into the substrate, the value of this emission is 5×10^7! The resonance linewidth is correspondingly narrow, so the normalized integrated radiated power is ~ 0.5 as shown in Fig. 6. As the inner cladding thickness is increased, both TE1 and TM1 peaks shift to steeper angles. Since the refractive index of the substrate (3.495) is very similar to that of the InGaAs slab (3.55), the amplification of the angle by refraction is small. As the inner cladding thickness is further increased, both of these peaks saturate at angles close to $\pi/2$. The last pair of panels in Fig. 9 is for an inner cladding thickness such that there are two propagating leaky modes and consequently two pairs of sharp peaks, TE1 and TE2 and TM1 and TM2. These very sharp angular dependences posed some numerical integration challenges. It was easy to miss the peak in the overall integrals [the bottom surface equivalents of Eqs. (31) and (33)]. Care had to be taken to track the peaks as a function of inner cladding thickness and integrate carefully over the peak with a very fine step size to get accurate results. This is also the reason that only the three-mirror-pair rather than the four-mirror-pair case was evaluated for the asymmetric structure; the resonances became too narrow and too peaked to deal with within the 16-bit precision of the present computer program. Adding losses to the dielectric constant limits the peak value, but also results in energy loss to absorption within the dielectric stack that must be accounted for in evaluating the total radiated energy.

V. SUMMARY AND CONCLUSIONS

A fully analytic treatment of the radiation from a dipole embedded in an arbitrary dielectric stack has been presented. The treatment starts from a Hertz-vector formalism introduced by Sommerfeld [1] and extended by Lucosz [9]. The analysis proceeds by finding the Fourier transform of the

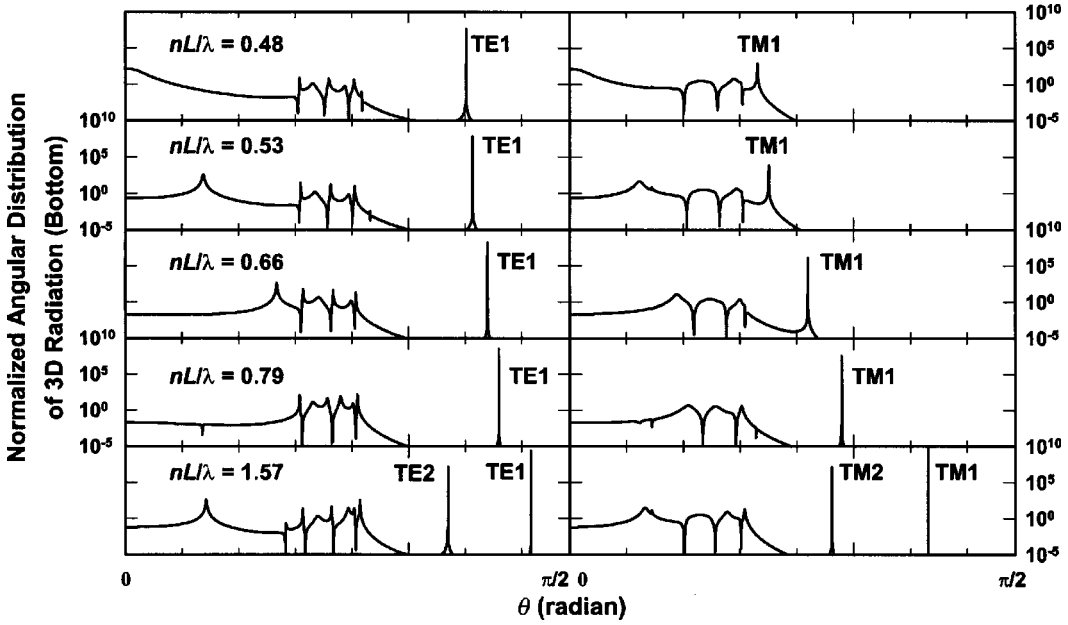


FIG. 9. Angular variation of the 3D radiation from the bottom (GaAs substrate) of an asymmetrically clad slab (air upper and GaAs lower) with three mirror pairs as the inner cladding thickness is varied for a horizontal dipole. The various values of nL/λ are given in the figure. The top panels, $nL/\lambda = 0.48$, correspond to position B, the peak normal emission, in Fig. 8 (top). The second set of panels, $nL/\lambda = 0.53$, corresponds to position E, the edge of the TE emission peak, in Fig. 8 (top). The third row of panels, $nL/\lambda = 0.79$, corresponds to a cladding thickness just before the emergence of the second top-surface emission peak (Fig. 6). The final row of panels, $nL/\lambda = 1.57$, corresponds roughly to the middle of the third top-surface emission peak in Fig. 6.

dipole fields for an unbounded medium, solving the independent boundary conditions across the multilayer stack for each Fourier component, and transforming back to real space. This procedure provides a rigorous solution to the multilayer boundary conditions. Two different classes of singularities are present in the integrand for the inverse transform: poles that correspond to bound 2D waveguide modes confined within the multilayer stack, and branch cuts that correspond to 3D radiation emitted from the sides of the stack into the upper and lower outer cladding half-spaces. While the boundary conditions are strictly valid, approximations are made in order to evaluate the radiated fields that are accurate only many wavelengths from the position of the dipole, e.g., in the radiation zone away from the near fields. The results are directly related to previous work [4] that investigated a simple dielectric slab geometry with the straightforward replacement of single layer reflectivities with their multilayer counterparts.

This formalism was applied to two closely related structures, both of which are germane to the design of semiconductor vertical-cavity lasers. Both structures contain a thin active region containing the dipole with symmetric inner claddings and an equal numbers of Bragg reflector mirror pairs on both sides. In the first case, the structure is fully symmetric with an air cladding both above and below the stack. In the second case, the lower cladding is replaced with a substrate with the same refractive index as the inner cladding. These two cases illustrate some interesting points. For the symmetrically air-clad structure, the dipole couples to both the bound modes of the slab and the radiation fields in the air spaces. For the asymmetric structure, on the other hand, there are no bound modes because of the high-index outer cladding and only the 3D radiation calculation is operative. However, for high reflectivities, e.g., a sufficient number of Bragg reflector pairs, the results should be essentially equivalent since the dipole can not ''see'' the substrate because of the high reflectivity. Instead of true bound modes, there are leaky modes that look very much like the bound modes of the air-clad structure, but whose power is transmitted into the substrate after many round trips within the film stack. Because of the very high reflectivities in these structures, especially at steep angles, very sharp angular dependences are found for this radiated power, corresponding to the bound modes in the symmetric air-clad case. In both cases, the 3D radiation into the upper low-index (air) half-space is a sharp function of the thickness of the inner cladding with TM and TE resonances at alternating half-wave inner cladding thicknesses. The asymmetric case is very close to that of some LEDs on solid substrates. For example, visible-emission, nitride-based quantum-well LEDs [13] are examples of mode-free structures [14]. Modes similar to guided modes but losing power due to leakage into the substrate are observed in such structures.

Finally, we present the following conclusions.

(1) Analytic solutions are presented using a rigorous Hertz-vector formalism for dipole emission within a multilayer dielectric structure. These solutions are suitable for application to edge-emitting and vertical-cavity surface-emitting lasers (VCSELs), resonant light-emitting diodes and photodetectors, and other multilayer optoelectronic devices. Expressions are presented as functions of a normalized optical thicknesses of the various layers and can be used to ana-

lyze either thickness or wavelength variations.

(2) Enhancement/suppression of the dipole radiation relative to that in an infinite medium is obtained. The results are very dependent on the structure; however, even with a very high finesse planar cavity, the maximum enhancement is limited to $\sim 31\%$ for a horizontal dipole at the first resonance peak, e.g., for a total inner cladding thickness of $\sim n\lambda$. The inner cladding thickness for peak emission is close to, but slightly longer than, the peak normal emission thickness.

(3) A larger fractional suppression of the dipole emission is found for a vertical dipole for very short inner-cladding thicknesses lengths $\ll n\lambda/2$. This is due to destructive interference between the radiated and reflected fields at the position of the dipole.

(4) The distribution of the radiated power between 3D radiation into the outer claddings and 2D radiation into the waveguide modes shows a larger variation with structure than does the total radiated power. For a free-standing symmetric structure, the addition of high-reflectivity Bragg mirrors modifies significantly the fraction of power emitted into the air space and its inner cladding thickness and angular variations. As expected, the angular width and inner cladding thickness tolerance of the normal-emission resonance decreases with increasing mirror reflectivities.

(5) In the asymmetric case, with a high-index substrate, the bound modes of the structure evolve into leaky modes whose energy is radiated into the substrate. The top-surface emission for this case is similar to that for the symmetric case. The bottom (substrate) 3D emission is characterized by a complex dependence at small angles corresponding to the reflectivity coefficient for this multilayer structure, and by very intense and narrow resonance peaks at higher angles corresponding to the leaky modes.

[1] A. Sommerfeld, Ann. Phys. (Leipzig) **28**, 665 (1909).
[2] A. Sommerfeld, *Partial Differential Equations in Physics* (Academic Press, New York, 1949).
[3] D. Marcuse, *Theory of Dielectric Optical Waveguides* (Academic Press, San Diego, 1991).
[4] S. R. J. Brueck, IEEE J. Sel. Top. Quantum Electron. **JSTQE-6**, 899 (2000).
[5] G. Björk, S. Machida, Y. Yamamoto, and K. Igeta, Phys. Rev. A **44**, 669 (1991).
[6] H. Rigneault and S. Monneret, Phys. Rev. A **54**, 2356 (1996).
[7] Y. Xu, R. K. Lee, and A. Yariv, Phys. Rev. A **61**, 033807 (2000).
[8] S.-T. Ho, Y. Chu, J.-P. Zhang, S. Wu, and M.-K. Chin, in *Optical Processes in Microcavities*, edited by R. K. Chang and A. J. Campillo (World Scientific, Singapore, 1996), pp. 339–387.
[9] W. Lukosz and R. E. Kunz, J. Opt. Soc. Am. **67**, 1607 (1977).
[10] W. Lukosz, Phys. Rev. B **22**, 3030 (1980).
[11] A. Baños, *Dipole Radiation in the Presence of a Conducting Half-Space* (Pergamon Press, Oxford, 1966).
[12] J. A. Stratton, *Electromagnetic Theory*, 1st ed. (McGraw-Hill, New York, 1941), pp. 573–575.
[13] S. Nakamura, M. Senoh, N. Iwasa, and S. Nagahame, Jpn. J. Appl. Phys., Part 2 **34**, L797 (1995).
[14] I. V. Akimova, P. G. Eliseev, M. A. Osinski, and P. Perlin, Quantum Electron. **26**, 1039 (1996).

Single-longitudinal-mode emission from interband cascade DFB laser with a grating fabricated by interferometric lithography

J.L. Bradshaw, J.D. Bruno, J.T. Pham, D.E. Wortman, S. Zhang and S.R.J. Brueck

Abstract: A distributed-feedback, type-II interband cascade laser is demonstrated in pulsed mode, emitting near 3.145 μm at temperatures between 50 and 80 K. Feedback is provided by a surface-etched grating formed using interferometric lithography. Between 50 and 80 K, the device's single longitudinal mode red-shifts with temperature at ∼0.1 nm/K. At 70 K and 190 mA of pulsed injection current, side-mode suppression exceeds 30 dB. The characteristics of related Fabry–Perot IC lasers are also described and contrasted with those of the distributed feedback device.

1 Introduction

In this paper, we report on the first demonstration of a distributed feedback (DFB) interband cascade (IC) laser. Light–current–voltage (L–I–V) and spectral characteristics of the DFB device are reported, along with the L–I–V characteristics of related Fabry–Perot (FP) devices for comparison. Possible improvements to the DFB device structure are also identified.

The results reported below are significant because they represent another step forward in the ongoing development of quantum-cascade-based semiconductor diode lasers (SDLs), which include the type-I intersubband quantum cascade (QC) lasers and the type-II IC lasers. These cascade laser designs partially overcome fundamental limitations of more conventional mid-IR laser designs, and hold the potential of providing single-mode SDLs throughout the mid-IR region that can operate under CW conditions at ambient temperatures (or temperatures accessible with thermoelectric coolers).

The QC laser has recently demonstrated CW operation at 300 K from a FP device, emitting at 9.1 μm [1], and single-mode DFB QC lasers have been demonstrated between 4.5 and 16.5 μm (see [2] for a recent review). FP IC lasers have been demonstrated in the 3 to 5 μm wavelength region, and recently, room temperature, pulsed operation was demonstrated at 3.51 μm [3]. However, single-mode IC lasers have not, to our knowledge, been reported previously.

2 Experimental details

The laser sample used to fabricate the DFB device, referred to below as M2, was grown by molecular beam epitaxy (MBE) on a p-type (001) GaSb substrate. The layer structure and DFB device geometry are shown schematically in Fig. 1. A 2.26-μm-thick n-type InAs/AlSb superlattice region serves as the waveguide's lower cladding and is followed by a 1.34-μm-thick core region. The core consists of 18 identical cascaded active regions separated by n-type injection regions. A 0.3-μm-thick, p-doped GaSb layer was grown above the core region and serves as a separate confinement/top contact layer and host layer for the etched grating of the DFB devices.

The active region of the laser's core consists of InAs and GaInSb quantum well (QW) layers and GaSb and AlSb blocking/barrier layers, and the injection regions consist of digitally graded n-InAs and AlSb layers. A schematic energy band diagram of a single period of a typical IC laser's cascaded active and injection regions is shown in Fig. 2. Electrons injected into the active region's InAs QW relax to the lowest heavy-hole state in the neighbouring GaInSb QW with the emission of a photon. Subsequent transport, facilitated by interband tunnelling, moves the electrons into the adjacent injection region, where they are transported through the region's conduction miniband for injection into the next active region. The active/injection regions of M2 and the MBE procedures used in its growth are the same as those used in the growth of the sample in [3] and have been described elsewhere [4].

After MBE growth, the M2 sample was cleaved to provide laser material for the fabrication of both DFB and FP devices. Prior to wet-etching for mesa definition of DFB devices, a grating was etched into the top of the DFB material as follows. First, an antireflection coating (ARC) and photoresist (PR) layer were deposited onto the top p-GaSb epilayer of the M2 sample. Two beams formed from the third harmonic of a yttrium-aluminium-garnet (YAG) laser at 355 nm were then directed from opposite positions onto the sample, each making angles of ∼45° relative to the sample's surface. The angle was chosen so that the beam's interference formed a 465-nm-period standing-wave pattern in the PR above the ARC layer, exposing the PR. The grating period was determined by dividing the expected emission wavelength of the M2 device (3.32 μm at 80 K) by twice the waveguide's modal index n. The latter was estimated at 3.47 by measuring the FP fringe spacing, $\Delta\lambda$,

© IEE, 2003

IEE Proceedings online no. 20030613

doi: 10.1049/ip-opt:20030613

Paper received 9th January 2003

J.L. Bradshaw, J.D. Bruno, J.T. Pham and D.E. Wortman are with Maxion Technologies, Inc., Hyattsville, MD 20782-2003, USA

S. Zhang and S.R.J. Brueck are with the Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106, USA

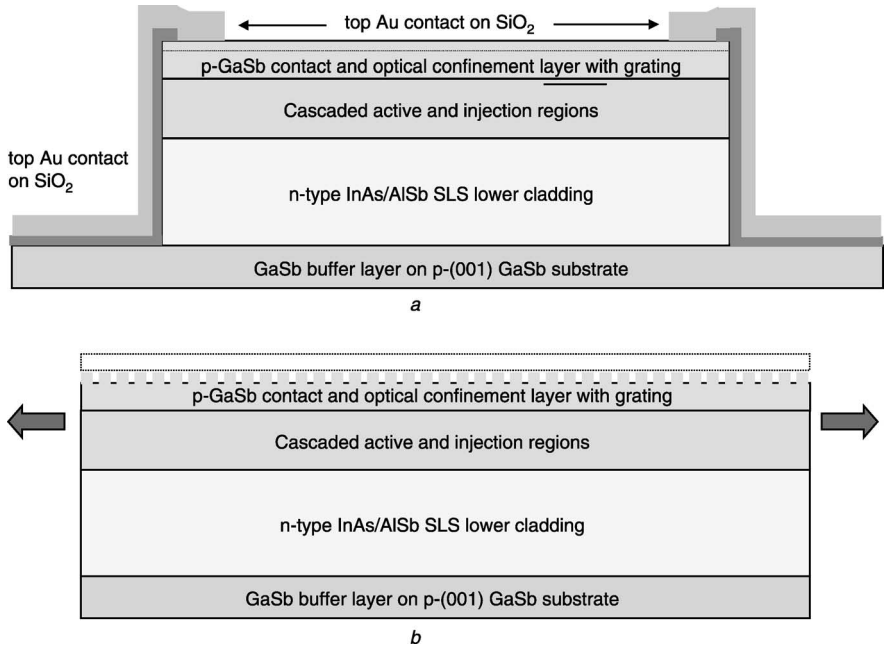


Fig. 1 *Schematic layer structure and device geometry used for DFB devices*

a Cross sectional view showing DFB edge-contacting scheme
b Side view showing top-grating orientation

of a similar device's below-threshold amplified spontaneous emission spectrum and using the relation $n = \lambda^2/(2L\Delta\lambda)$, where $\Delta\lambda$ is the separation between adjacent FP fringes, λ is the expected emission wavelength and L is the device's cavity length.

The exposed PR and underlying ARC layers were then removed, and an ion beam was used to etch a shallow vertical trench into the exposed p-GaSb layer. After cleaning, this process resulted in an ~40-nm-deep, 465-nm-period grating with a duty cycle of approximately 50% on the top p-GaSb layer of the DFB laser material.

The FP and DFB materials were processed into ridge mesa devices, with mesa widths of 65 and 115 μm, defined with optical lithography and wet chemical etching. Insulating layers of SiO_2 were deposited after the mesas were defined, leaving openings on top of the ridges to allow for electrical contacts to the devices (Fig. 1). In the case of DFB devices, only 5-μm-wide windows were opened for metal contact along each top edge of the mesa ridge. This left most of the top of the (wider) DFB mesas without a metal contact over the grating; a step intended to reduce optical losses. Metal contacts were then deposited onto the tops of the mesas, as appropriate, and onto the bottom substrate for electrical contact. Some DFB devices were intentionally fabricated with metal contacts covering their gratings for the purpose of performance comparisons. Laser bars were cleaved into cavities approximately 1-mm long, and both facets were left uncoated.

The fabricated lasers were indium-soldered onto a copper heatsink and mounted onto the temperature-controlled cold finger of an optical cryostat. L–I–V and spectral characterisations were carried out using a liquid-nitrogen-cooled InSb detector and a grating monochromator with lock-in electronic detection.

3 Performance of Fabry–Perot lasers

Figure 3 shows the L–I–V characteristics of a 115-μm-wide by 0.96-mm-long cavity length M2 FP device. This device had the entire top of the mesa contacted with a 0.3-μm-thick Au contact layer. The threshold currents at

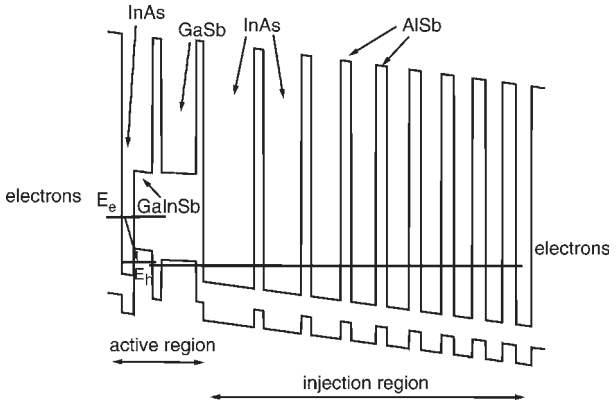


Fig. 2 *Energy band diagram showing single stage of a typical IC laser structure at operating bias*

Active region (left) shows interband optical transition and interband tunnelling transition; injection region (right) transports electrons to next stage of device

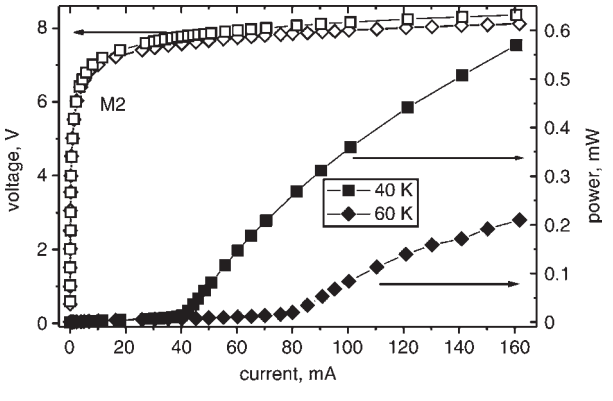


Fig. 3 *CW L–I–V characteristics of a 115-μm × 0.96-μm M2 FP device at heatsink temperatures of 40 and 60 K*

Device does not have a grating structure, and the top mesa is covered with a thick Au contact

heatsink temperatures of 40 and 60 K are 40 and 80 mA, respectively (corresponding to current densities of 36.2 A/cm^2 and 72.4 A/cm^2, respectively). The output power is very low: at 100 mA of CW injection current at 60 K, the measured power from a single facet is below 0.1 mW. The device slope efficiencies at 40 and 60 K are also low.

For comparison purposes, Fig. 4 shows the L–I–V characteristics of a 65-μm-wide by 0.98-mm-long cavity length FP IC laser, fabricated from a laser structure that included an upper cladding region. The Fig. 4 device is the one described in [3] (referred to below as M1) and has active and injection regions that are nominally identical to those of the M2 FP device. The differences lie above the active/injection regions: in the M1 device, an upper cladding region separates the laser core from the top metal contact, whereas in the M2 device the optical mode is confined between the lower cladding and the metal contact and/or air, depending on how the mesa top is contacted. The M1 characteristics show a threshold current density of 13.2 A/cm^2 and an initial slope efficiency of 0.65 W/A at a heatsink temperature of 80 K. The initial slope efficiency corresponds to a differential external quantum efficiency of ~350% (approximately 19% differential efficiency per stage) when the output from both facets is assumed to be equal. At 100 mA of CW current, the output power measured from a single facet was 57 mW. The I–V characteristics of M1 exhibit a high resistance below threshold and an abrupt decrease in differential resistance at threshold, with the decrease occurring at 6.95 V. Based on the observed 80 K laser emission at 3.315 μm (373.9 meV) and the presence of 18 cascaded stages, this corresponds to a voltage efficiency of 96%. By comparison, the output power and slope efficiencies of the M2 FP device are greatly reduced. It is clear that this reduction in M2's performance relative to M1 is not due to the active/injection region designs, but is instead related to the waveguide structure and/or to the quality of the electrical contacts made to the core region. In the M2 FP design, the optical mode is larger at the top metal contact than in the M1 device. Consequently, optical losses were expected to be larger in the M2 FP and DFB devices. Because of this, the p-GaSb separate-confinement/contact layer was doped with Be at a relatively low level of 5-6 × 10^{17} cm^{-3}. This doping level was chosen as a compromise between the simultaneous requirements of high electrical conductivity in the layer and low free-carrier losses. Despite the generally inferior characteristics of the M2 device relative to the M1 device, the M2 device still has a reasonably good diode characteristic, showing a high initial resistance to current flow and a rapid drop in resistance above approximately 7 V.

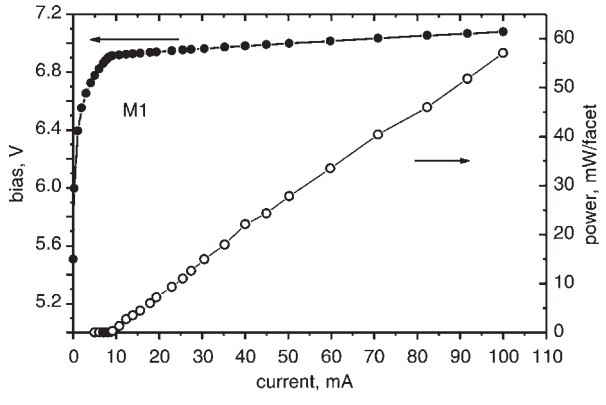


Fig. 4 *CW L–I–V characteristics of the 65-μm × 0.98-μm FP device of [3] at a heatsink temperature of 80 K*

4 Performance of DFB lasers

L–I–V characteristics of the DFB laser structure processed with the M2 grating material and with metal contacts along only the top edges of the mesa are shown in Fig. 5. The laser exhibited a threshold current of ~165 mA in the pulsed mode at 60 K. Note that the I–V characteristics do not show the high resistance at low current that was observed for the M2 FP device (Fig. 3) whose mesa was covered with a thick Au contact. To determine whether this might be related to the relatively low doping in the Be-doped contact layer, another DFB device was fabricated from the same M2 DFB material, but a 100-Å-thick Au contact was deposited over the entire top of the grating structure (electrically connecting the thicker Au edge contacts). The L–I–V characteristics of this device are shown in Fig. 6, and the data shows significantly improved I–V characteristics with a high initial resistance and improved forward bias diode characteristics. Additionally, the device exhibited a lower threshold current of ~10 mA at 80 K, corresponding to a threshold current density of ~27 A/cm^2. This threshold current density compares favourably with the threshold current density of 72 A/cm^2 at 60 K of the device shown in Fig. 5, presumably due to the thinner gold contact layer resulting in a lower optical loss and reduced spreading resistance in the top GaSb layer. Note, however, that this same device did not emit in a single longitudinal mode but exhibited FP emission characteristics.

Lasing spectra acquired under pulsed conditions are shown in Fig. 7 for the DFB laser device shown in Fig. 5. The lasing spectrum exhibits a single longitudinal mode near 3.142 μm at 60 K. When the device is processed with the grating-free laser material in a FP geometry, it exhibits a FP lasing spectra centred near a wavelength of 3.30 μm (multimode spectra at right-hand side of Fig. 7). The separation between these lasing energies is nearly 19 meV; an energy comparable to the width of the device's spontaneous-emission spectrum. This difference could be explained if our estimate of the modal index in the DFB material were ~5% larger than its actual value. This would also be consistent with the high threshold current density seen in Fig. 5. Note also that the cavity length for the DFB laser is half that of the FP device. Thus, the longitudinal mode spacing in the DFB device would be twice the spacing

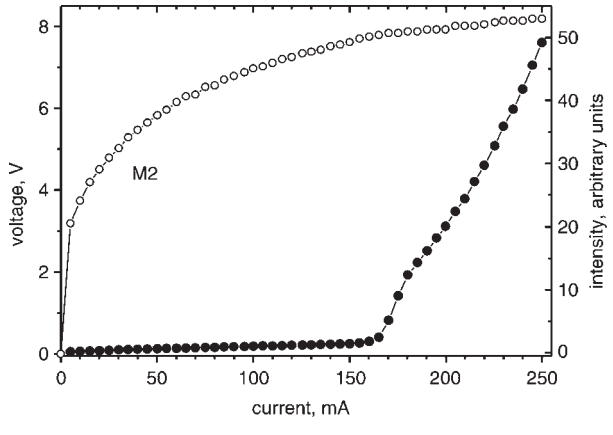


Fig. 5 *L–I–V characteristics of an edge-contacted, 115-μm-wide by 0.54-mm-long mesa DFB laser*

Data were acquired under pulsed conditions using a 1-μs pulse width and a 10 kHz repetition rate at a heatsink temperature of 60 K

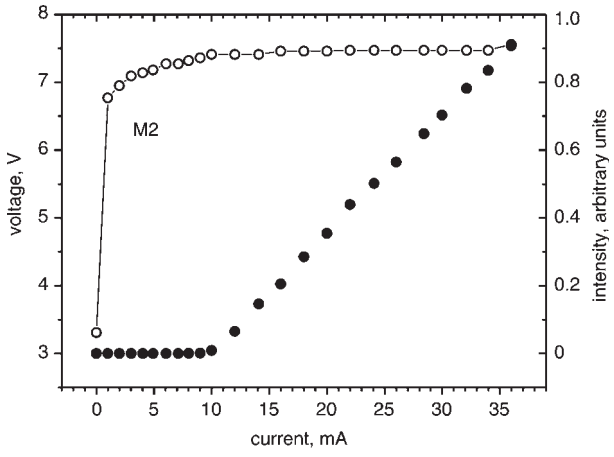


Fig. 6 *L–I–V characteristics of an edge-contacted, 35-μm-wide by 1.04-mm-long mesa DFB laser acquired under pulsed conditions using a 1-μs pulse width and a 10 kHz repetition rate at a heatsink temperature of 80 K*

Top of grating is covered with an 80-Å-thick Au contact

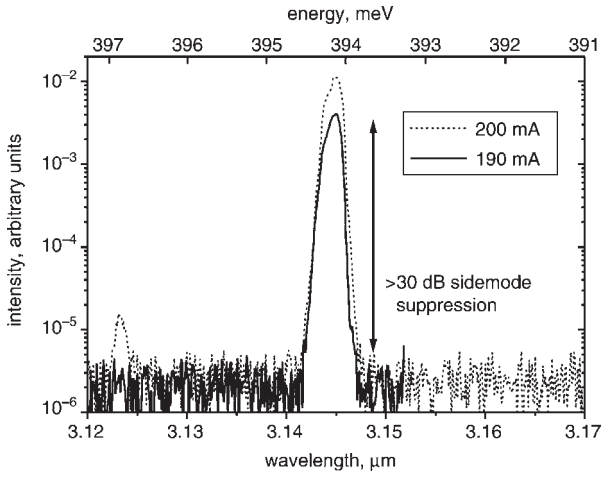


Fig. 8 *High-resolution emission spectrum of the DFB device described in Fig. 5 under pulsed conditions at 70 K*

Vertical scale is logarithmic showing side-mode suppression exceeding 30 dB at 190 mA injection current. The spectra were acquired using 1-μs pulses at a 10 kHz repetition rate

of the cavity modes shown in Fig. 7 for the FP device. Consequently, the DFB lasing spectra clearly consists of a single longitudinal mode. However, because the processed mesa widths are rather broad ($>65\,\mu m$), all the lasers investigated here probably emitted several lateral modes.

Higher resolution lasing spectra are shown in Fig. 8. At a temperature of 70 K and a pulsed injection current of 190 mA, the DFB laser exhibits a side-mode suppression ratio in excess of 30 dB. At 200 mA injection current, the emergence of another mode can be seen near 3.122 μm. Note that the vertical axis is logarithmic. Figure 9*a* shows spectra acquired under pulsed conditions at temperatures ranging from 50 to 80 K. The DFB lasing wavelength is plotted as a function of temperature in Fig. 9*b*. The laser exhibits a temperature tuning rate of 0.1 nm/K, which is consistent with expectations based on the variation of the refractive index of III–V materials with temperature.

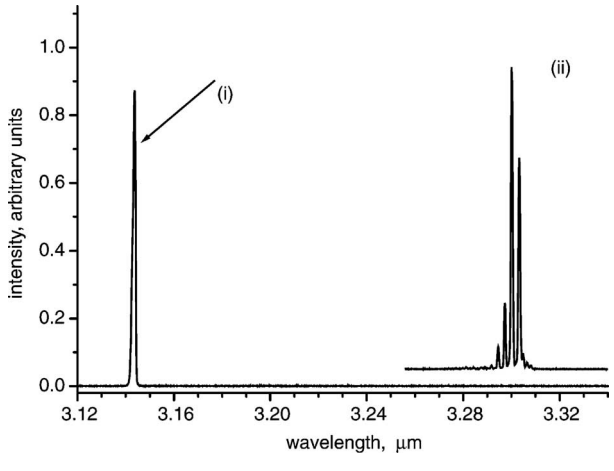


Fig. 7 *Emission spectra of the DFB Single longitudinal mode emission at 60 K (left) from edge-contacted DFB laser device described in Fig. 5 and multimode emission is observed from FP device fabricated from the same M2 laser material (right) at 80 K. Both spectra were acquired using a 1-μs pulse width and a 10 kHz repetition rate*

(i) Single-longitudinal-mode DFB operation, 115 μm × 0.54 mm mesa, $T = 60$ K, $I = 175$ mA
(ii) 65 μm × 1.01 mesa, Fabry–Perot laser, $T = 80$ K, $I = 60$ mA

At higher pulsed-current levels, the DFB laser exhibits linewidth broadening. Several lasing spectra at high pulsed-current levels are shown in Fig. 10. Note the significant broadening of the laser linewidth at higher current levels. The broadening is attributed to heating during the current pulse as well as the influence of the higher order lateral modes at higher injection-current levels.

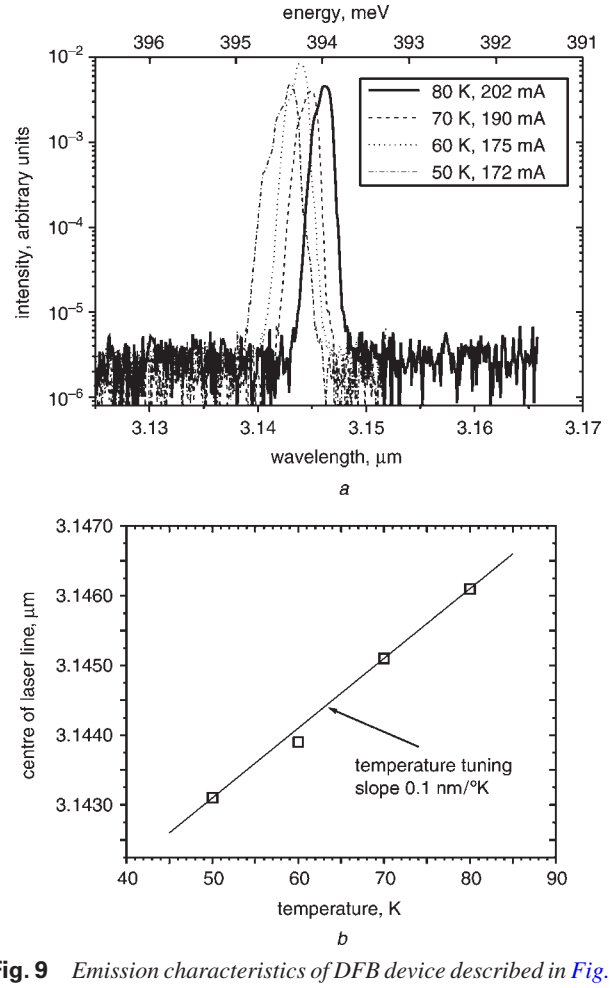


Fig. 9 *Emission characteristics of DFB device described in Fig. 5*

a High-resolution emission spectrun under pulsed conditions between 50 K and 80 K
b Emission wavelength against heatsink temperature between 50 K and 80 K

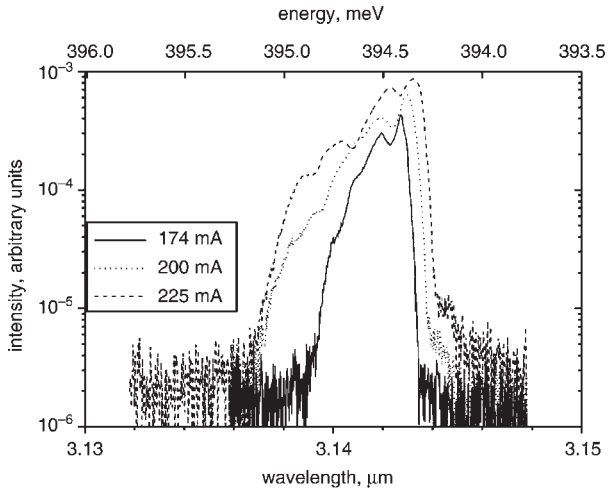


Fig. 10 *High-resolution emission spectra of the DFB device described in Fig. 5 under pulsed conditions at 50 K and injection currents of 174, 200 and 225 mA*

Effects of heating during the pulse and multiple lateral modes are evident

5 Summary and conclusions

We have described the fabrication and performance characteristics of the first DFB IC laser. The laser exhibits single-longitudinal mode operation in pulsed mode between 50 and 80 K, emitting near 3.144 μm at 60 K. Results show that improved performance can be expected from designs with a slightly larger $(\sim+5\%)$ grating period and a more heavily p-doped GaSb contact layer. The larger period will provide feedback closer to the M2 material's gain peak, and the improved top GaSb layer conductivity will allow for a reduced spreading resistance and eliminate the need for covering the grating with Au to achieve improved I–V characteristics. This should significantly reduce optical losses, and improve device output power, slope efficiency and temperature performance.

6 Acknowledgments

This work was supported in part by the Army Research Laboratory through a Cooperative Research and Development Agreement (CRDA# 0105-A-C980) and by a Phase I SBIR Contract funded by the Missile Defense Agency and managed by Dr R. Soref of the Air Force Research Laboratory, Hanscom, MA (Contract F19628-01-C-0035). The authors also wish to thank Dr R.Q. Yang for his contributions and Dr J.R. Meyer and I. Vurgaftman of the Naval Research Laboratory for helpful discussions.

7 References

1 Beck, M., Hofstetter, D., Aellen, T., Faist, J., Oesterle, U., Ilegems, M., Gini, E., and Melchoir, H.: 'Continuous wave operation of a mid-infrared semiconductor laser at room temperature', *Science*, 2002, **295**, p. 301
2 Gmachl, C., Straub, A., Colombelli, R., Capasso, F., Sivco, D.L., Sergent, A.M., and Cho, A.Y.: 'Single-mode, tunable distributed-feedback and multiple-wavelength quantum cascade lasers', *IEEE J. Quantum Electron.*, 2002, **38**, p. 569
3 Yang, R.Q., Bradshaw, J.L., Bruno, J.D., Pham, J.T., Wortman, D.E., and Tober, R.L.: 'Room temperature type-II interband cascade laser', *Appl. Phys. Lett.*, 2002, **81**, p. 397
4 Yang, R.Q., Bradshaw, J.L., Bruno, J.D., Pham, J.T., and Wortman, D.E.: 'Mid-infrared type-II interband cascade lasers', *IEEE J. Quantum Electron.*, 2002, **38**, p. 559

Fabrication of 1D and 2D vertical nanomagnetic resonators

Shuang Zhang, Wenjun Fan, B. K. Minhas, A. Frauenglass,
K. J. Malloy, and S. R. J. Brueck[a)]
Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, New Mexico 87106

(Received 3 June 2004; accepted 5 October 2004; published 13 December 2004)

We have successfully fabricated the first resonant magnetic nanostructures exhibiting a negative permeability in the midinfrared. These metal-dielectric structures exhibit local resonances based on metallic inductive-capacitive effects that are controlled by the dimensions of the individual nanostructures and are independent of the periodicity, which is much smaller than the resonance wavelengths. Compared to other commonly used structures for obtaining negative permeability in the microwave and THz regions which are based on a planar fabrication paradigm, this structure is oriented vertically so that the smallest features are controlled by deposition rather than by lithography. © *2004 American Vacuum Society.* [DOI: 10.1116/1.1824067]

I. INTRODUCTION

Double negative materials (DNM), or left-handed materials, are of intense current interest as a result of their unconventional optical properties[1] that promise enhanced linear and nonlinear optical effects.[2] A DNM has both a negative permittivity (ε) and a negative permeability (μ). While metals below the plasma frequency provide a negative permittivity, naturally occurring materials with negative permeability are not available. Composite electromagnetic materials with a structural scale much smaller than an electromagnetic wavelength can act as effective media with negative permeability.

The most widely investigated structure for establishing magnetic resonance and thus a negative permeability is the split ring resonator (SRR) proposed by Pendry *et al.* in 1999.[3] The SRR consists of two concentric metal rings each with a gap facing towards the opposite directions. This antenna structure can be viewed as an equivalent inductor-capacitor (LC) tank circuit; the inductance is due to the loops while the capacitance is due to the gaps and the mutual coupling between the two loops. When a time-varying magnetic field penetrates the structure, currents are induced in the metal rings, the magnetic field generated by the current either opposes or adds to the external magnetic field; at frequencies just above the resonance, negative permeability can be realized. Since this initial proposal, many experiments have been reported at microwave frequencies.[4]

There has been a strong interest in extending these magnetic resonance structures to higher frequencies to extend the domain of application. Recently, Yen *et al.* fabricated a SRR structure with a resonance around 1 THz.[5] They used conventional planar processing technologies with minimum features sizes in the range of ~ 1 μm, defined by optical lithography. Using conventional technologies, it is difficult to scale this structure by the factors of ~ 100 to 1000 needed to shift the resonance to the infrared and optical spectral regions. In this article, we describe the fabrication and modeling of an array of nanostructures with a resonance in the mid-IR region (to 65 THz) and properties that are indicative of a negative effective permeability. This structure has a vertical configuration, the smallest features are defined by self-aligned deposition instead of lithography, so they are relatively easy to control at the ~ 1 nm level, and the processing steps are very robust and readily extended to large areas. To our knowledge, this is the first experimental work on negative permeability reported for the mid-IR.

The structure (schematic is shown in Fig. 1) consists of an array of upper open gold "staples" each with two footings, separated from a lower continuous gold layer by a dielectric layer of ZnS. The light is incident along the normal to the wafer (z direction in Fig. 1), and the polarization is defined as: TM(TE) for E(H) field along the x direction. Each staple is an LC circuit with the major part of the inductor associated with the loop of the staple and with capacitors formed between the top gold footings, the dielectric layer, and the bottom gold layer. In the simplest model, the inductance and the capacitance of the LC circuit can be written as $L=\mu_0 A = \mu_0 WH$ and $C=\varepsilon d_c/t_c$, where A is the area of the loop, ε is the dielectric constant of ZnS, and the other parameters are labeled in Fig. 1. However, due to the finite electron inertia of the metal at optical frequencies and edge effects of the structure geometry, both the inductance term and capacitance term cannot be simply expressed by analytical expressions. Still, the geometrical parameters qualitatively explain how resonance frequency changes; for example, when the footing size d_c is increased, the capacitance increases, and the resonance frequency $f=1/\sqrt{LC}$ decreases. The same reasoning also applies to the change of frequency due to a the change in the inductance.

As a result of the thick Au film ($\gg$IR skin depth), light can only be reflected or absorbed from the structure; there is no transmission across the IR. For analysis, it is simpler to model the transmission through the staple and its mirror image in the metal surface, as shown in the bottom panel of

[a)]Author to whom correspondence should be addressed; electronic mail: brueck@chtm.unm.edu

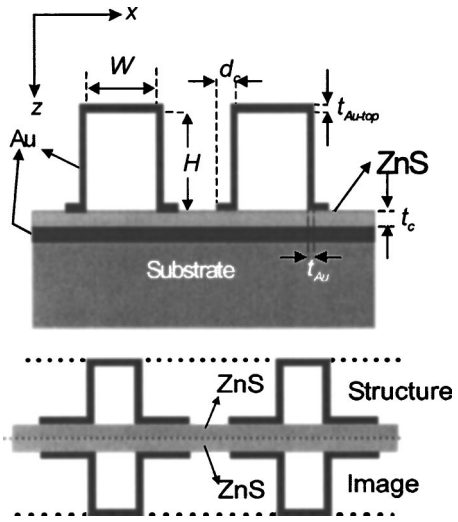


FIG. 1. Top: Schematic of the nanostructures. The light is incident along the normal from the top; for TM polarization, ***H*** is in the *y*-direction, perpendicular to the loop. Bottom: The equivalent *LC* circuit formed between the top staple structure and its image in the metal. The dashed lines are the reference planes for the effective permeability calculation.

Fig. 1. The period of the array is 600 nm, which is much smaller than the resonant wavelength that varied between 4- and 7-μm for these experiments. Three 1D samples (magnetic structures extending $\gg\lambda$ into the plane of Fig. 1) (A, B, C) and one 2D (as in Fig. 2; extent of each individual structure $\ll\lambda$) sample (D) were fabricated with different geometric parameters as shown in Table I. Figure 2 shows the FEM pictures of the 2D structure (D) and 1D structure (B). Samples A and C have different footing sizes (d_c) while samples A and B have different staple heights (H), with other parameters held constant. Sample D has the same parameters as B except for the segmentation in the *y*-direction and the removal of the material inside the loop.

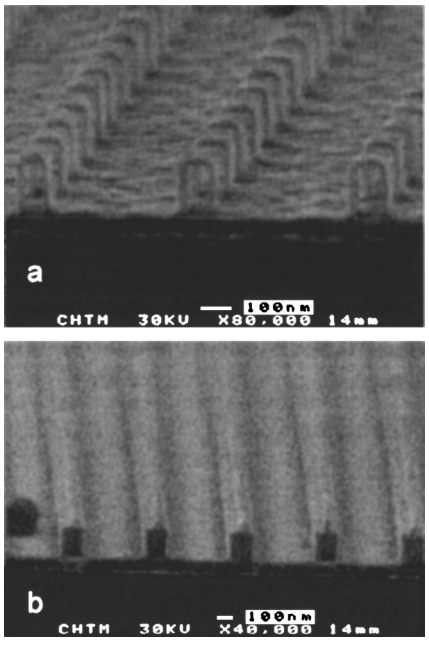


FIG. 2. SEM pictures of the arrays of 2D [sample D, shown in (a)] and 1D [sample B, shown in (b)] staple-shaped nanostructures.

TABLE I. Sample geometrical parameters (all dimensions in nm).

Sample	W	H	d_c	t_{Au}	$t_{\mathrm{Au\ top}}$	t_c
A	130	280	190	~15	~30	80
B(D)	130	180	190	~15	~30	80
C	130	280	90	~15	~30	80

II. FABRICATION

The processing steps for the 1D samples are as follows. First, a ~100-nm thick layer of gold is evaporated onto a silicon wafer, followed by dielectric evaporation of a 80-nm thick layer of ZnS ($n\sim2.26$), which is transparent throughout the ~2- to 10-μm IR region. The Au layer forms a ground plane for the structure and the ZnS is the dielectric spacer for the capacitor. Next, a layer of Si_3N_4($n\sim1.7$) is grown on the sample by plasma enhanced chemical vapor deposition (PECVD), followed by spun-on layers of a polymer bottom antireflection coating (BARC) and an I-line photoresist (PR). A large-area, 600-nm pitch, one-dimensional grating is fabricated by interferometric lithography (IL) using a 355-nm third-harmonic YAG laser source [Figs. 3(a) and 3(b)].[6] The pattern is transferred into the Si_3N_4 layer by reactive ion etching (RIE) using CHF_3 and O_2 gases and the residual PR and ARC are subsequently removed by plasma ashing with O_2 gas [Fig. 3(c)]. The etching is selective and leaves the ZnS layer untouched. Metal evaporation is carried out at an angle to deposit Au on the top and one sidewall of

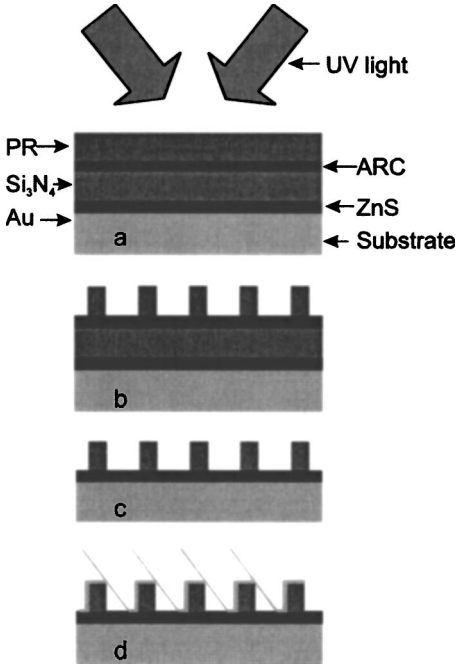


FIG. 3. Schematic diagrams of fabrication steps for 1D structure. (a) Interferometric lithography with two coherent UV beams. (b) The grating in PR after development. (c) The grating in the nitride layer after RIE etching. (d) Au deposition at an angle from one side.

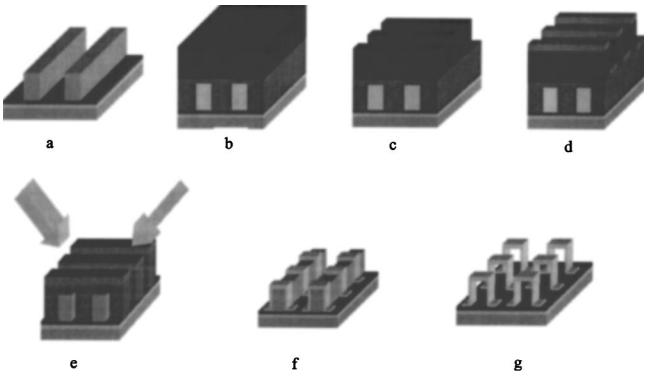


FIG. 4. Schematic flow charts of processing steps for 2D structure. (a) Gratings in the nitride layer. (b) Arc and PR are deposited by spin coating. (c) Gratings in PR perpendicular to the nitride gratings are fabricated. (d) Lift-off process to obtain metal gratings. (e) Metal evaporations at an angle are carried out on both sides. (f) ARC is removed by O_2 plasma ashing. (g) Nitride removed by CHF_3 plasma ashing.

the Si_3N_4 gratings, and to form the footings on one side [Fig. 3(d)]. A second evaporation is then carried out from the other side to form the completed "staple" structure. Because the top of the staple structure receives two Au depositions, the thickness of the Au on the top is twice that on the sidewalls and footings. The angle of the metal evaporation along with the nitride dimensions controls the footing size, and hence the capacitance. The dimensions of the nitride pattern control the area of the loop and, thus, the inductance.

For the 2D samples, the same procedures are followed up to Fig. 3(c), which is repeated as Fig. 4(a). A thick layer of ARC is then spun on the wafer to cover the Si_3N_4 grating and planarize the surface. Then a layer of PR is added by spinning [Fig. 4(b)]. Another exposure is carried out to produce gratings perpendicular to the buried Si_3N_4 gratings [Fig. 4(c)], followed by a metal evaporation at normal incidence [Fig. 4(d)]. The wafer is then immersed in acetone to remove the PR, and a grating pattern is obtained in the metal. Using the metal grating as mask, RIE etching in an O_2 plasma is done to selectively etch away the thick ARC layer locally exposing the Si_3N_4 gratings which are not etched in the O_2 plasma. Two evaporations of Au at an angle are carried out to form the staple structures [Fig. 4(e)]. The ARC layer is then removed by a high pressure, isotropic O_2 plasma ashing step [Fig. 4(f)]. Finally, the Si_3N_4 gratings are removed by CF_4 plasma ashing leaving the 2D array of hollow staple nanostructures [Fig. 4(g)].

III. INFRARED REFLECTANCE CHARACTERIZATION

FTIR reflectance measurements were performed on these four samples for both TE and TM polarizations. Due to the anisotropy of the staple structure, we expect to optically couple to the resonant structure only when the magnetic field of the incoming light is perpendicular to the loop of the staple (TM polarization). The IR transmission is negligible due to the thick bottom metal, so the nanostructure optical properties are solely characterized by the resonance absorption. As expected, the TE measurements on all the samples did not show any structure [to avoid clutter in the figure, the TE reflectance is shown only for sample C in Fig. 5(top)]. The TM measurements show reflectance dips that vary according to the sample geometrical dimensions directly confirming the existence of structure-dependent magnetic activity for these staple arrays. The strong resonance, with as much as 50% absorption in a single layer of nanostructures, suggests that very high fields are generated in the nanostructures that will be of interest for future nonlinear optics experiments. As shown in Fig. 5 (top), sample A with the largest inductance and capacitance, has the lowest resonant frequency. Sample C has the same inductance as sample A but a smaller capacitance (smaller footing width *dc*), giving a higher frequency resonance (shorter wavelength). Sample D (and B) has the same capacitance as sample A and a lower inductance (shorter loop height, *H*), also giving a higher frequency resonance. Comparison between the 1D sample and 2D sample (samples B and D) with the same geometric parameters (not shown) was made; there is no significant difference either in the location of the resonance dip or on the depth of the resonance. Experimentally, samples B and D have relatively smaller reflection dips; the reason for this variation is under investigation.

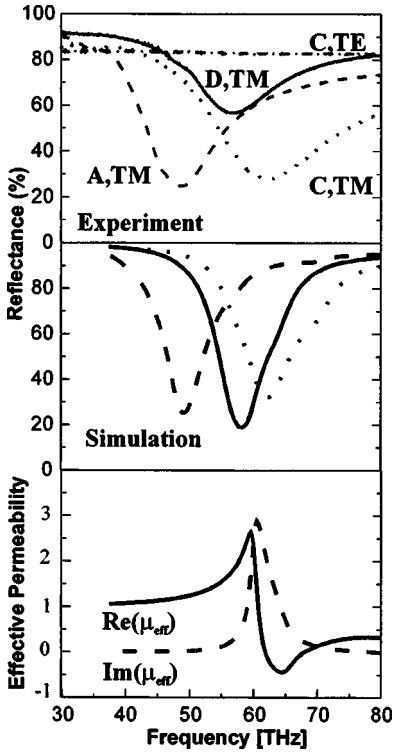


FIG. 5. Experimental (top) and simulation (middle) results for the reflectivity as a function of frequency for three different structures (cf. Table I). The resonance frequency is a function of the nanostructure parameters and is independent of the array period, which is the same for all three structures. The resonance is observed only for TM polarization, i.e., incident magnetic field coupled into the inductive structure. The bottom panel shows the effective permeability extracted from the model for sample C.

IV. DISCUSSION

The simulation results for the three structures are shown in Fig. 5 (middle). The simulation is based on the rigorous coupled wave analysis, or RCWA, which is commonly used for analysis of scattering of electromagnetic waves from periodic structures.[7] Bulk values of the wavelength dependence of the Au optical properties were used.[8] The modeling explains the experimental data well except that the width of absorption dip of the experimental data is somewhat larger than that of the model. Both nonideal material parameters (particularly electron scattering in the thin Au films) nonuniformity of the structures across the ~1-cm diameter measurement area can contribute to this observation.

To further investigate the strength of this magnetic activity, we extracted the effective permeability of sample C from the simulated results of the complex transmission and reflection coefficient of the equivalent *LC* circuit (Fig. 1, bottom).[9,10] For a uniform layer of metamaterials with incident and outgoing media of air, the complex transmission and reflection coefficient are expressed as

$$\frac{1}{t} = \left[\cos(nkd) - \frac{i}{2}\left(z + \frac{1}{z}\right)\sin(nkd)\right], \tag{1}$$

$$\frac{r}{t} = -\frac{1}{2}i\left(z - \frac{1}{z}\right)\sin(nkd), \tag{2}$$

where n and z are the refractive index and impedance of the metamaterial. By inverting Eqs. (1) and (2), we can obtain n as a function of r and t,

$$n = \frac{1}{kd}\cos^{-1}\left(\frac{1 + r^2 - t^2}{2t}\right). \tag{3}$$

After n is evaluated, z is obtained from

$$z = \frac{i\left[\frac{1+r}{t} - \cos(nkd)\right]}{\sin(nkd)}. \tag{4}$$

Finally, the complex value of the effective permeability is $\mu_{\text{eff}} = nz$. As shown in Fig. 5 bottom, the real part of the permeability μ shows a strong modulation and is negative over a restricted frequency range; the imaginary part shows the complementary Lorentzian-type peak—features characteristic of a strong magnetic resonance. Extending the modeling (not shown) across a range of structure parameters suggests that these results can be extended to a more negative permeability by increasing the width W of the staple structures, i.e., increasing the fill factor.

In summary, nanostructured magnetic resonators have been fabricated with resonances in the mid-IR spectral region. The measured reflectivity shows a strong absorption dip in good agreement with the simulation results. The simulation shows strong magnetic response and indicates negative magnetic permeability over a limited frequency range. Combining this negative permeability structure with a negative permittivity structure, for example by segmenting the underlying Au film, will lead to metamaterials with negative refractivity in the IR suitable for transmissive as well as reflective measurements. Further work is underway to investigate the scaling of these structures to the near-IR and visible spectral regions.

ACKNOWLEDGMENT

Support for this work was provided by the ARO/MURI program in Deep Subwavelength Optical Nanolithography.

[1] V. G. Veselago, Sov. Phys. Usp. **10**, 509 (1968).
[2] J. B. Pendry, Phys. Rev. Lett. **85**, 3966 (2000).
[3] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, IEEE Trans. Microwave Theory Tech. **47**, 2075 (1999).
[4] R. Shelby, D. R. Smith, and S. Schultz, Science **292**, 77 (2001).
[5] T. J. Yen, W. J. Padilla, N. Fang, D. C. Vier, D. R. Smith, J. B. Pendry, D. N. Basov, and X. Zhang, Science **303**, 5663 (2004).
[6] X. Chen, S. H. Zaidi, S. R. J. Brueck, and D. J. Devine, J. Vac. Sci. Technol. B **14**, 3339 (1996).
[7] M. G. Moharam and T. K. Gaylord, J. Opt. Soc. Am. **72**, 1385 (1982).
[8] J. H. Weaver, C. Krafka, D. W. Lynch, and E. E. Koch, *Optical Properties of Metals* (Fachinformationzentrum, 7514 Eggenstein-Leopoldshafen 2, Karlsruhe, Germany), Vols. I and II, Physics Data, Nos. 18-1 and 18-2 (1981).
[9] D. R. Smith and S. Schultz, Phys. Rev. B **65**, 195104 (2002).
[10] S. O'Brien and J. B. Pendry, J. Phys.: Condens. Matter **14**, 6383 (2002).

Midinfrared Resonant Magnetic Nanostructures Exhibiting a Negative Permeability

Shuang Zhang, Wenjun Fan, B. K. Minhas, Andrew Frauenglass, K. J. Malloy, and S. R. J. Brueck*

Center for High Technology Materials, University of New Mexico, Albuquerque, NM 87106, USA

(Received 6 April 2004; published 26 January 2005)

We experimentally demonstrate the first midinfrared (mid-IR) resonant magnetic nanostructures exhibiting a strong magnetic response corresponding to a negative permeability. This result is an important step toward the achievement of a negative refractive index in the IR. The possibility of extending negative permeability to higher frequencies is discussed; a structure with a negative effective permeability at a near-IR resonance frequency of 230 THz (1.3 μm) is proposed.

DOI: 10.1103/PhysRevLett.94.037402 PACS numbers: 78.20.Ci, 42.25.Bs, 42.79.Dj

In 1968, Veselago proposed the concept of a negative refractive index or left-handed material (LHM) [1] with both a negative permittivity (ε) and a negative permeability (μ). Many interesting properties are associated with negative materials including a negative index of refraction, backward phase propagation, a reversed Doppler effect, and backward emission of Cherenkov radiation. Recently, interest in LHMs has increased substantially with the theoretical prediction that a planar slab of LHM functions as a perfect lens without any diffractive loss of resolution [1,2] and with the first demonstrations of LHMs in the rf and THz electromagnetic regions [3–6].

While metals provide a negative permittivity at frequencies below the plasma frequency, naturally occurring materials with negative permeability are not available. Composite electromagnetic materials with resonant structures with sizes much less than the wavelength can act as an effective homogeneous media with a negative permeability. Pendry *et al.* proposed a split ring structure that responds to the magnetic field of incident radiation [7]. The split ring structure (SRS) can be viewed as an equivalent inductor-capacitor (LC) tank circuit. In the presence of a time-varying magnetic field, the magnetic field generated by the current induced in the ring opposes the external magnetic field. At frequencies in the vicinity of the resonance, a negative effective permeability can be realized. While most experimental work has been done at microwave and, more recently, THz frequencies, extending the phenomena to infrared (IR) and visible frequencies will greatly increase the range of applications. Based on the previous modeling work [7] and on the fabrication difficulties associated with scaling the SRS to higher frequency, there has been general pessimism about the prospects of extending these properties to optical frequencies in metallic structures.

In this Letter, we describe the fabrication, characterization, and modeling of arrays of a new nanostructure design with resonances in the mid-IR region and properties that demonstrate strong magnetic activity indicative of negative permeability. To our knowledge, this is the first experimental work on negative permeability reported in the IR. Importantly, the scaling of these results to even higher frequency is investigated, and a structure exhibiting negative permeability in the technologically important near-IR range (1.3 μm) is proposed.

The structure (Fig. 1, top) consists of an array of gold ''staples'' each with two outwardly splayed footings, separated from a thick continuous gold film by a ZnS dielectric layer. Each staple is a LC circuit with the structure part of the inductor associated with the upper loop of the staple and with two capacitors formed between the gold staple footings—the dielectric layer and the bottom gold layer. As a result of the thick Au film, light can only be reflected from or absorbed within the structure; there is no transmission in the IR. For analysis, it is simpler to view this structure in transmission as the staple and its image mir-

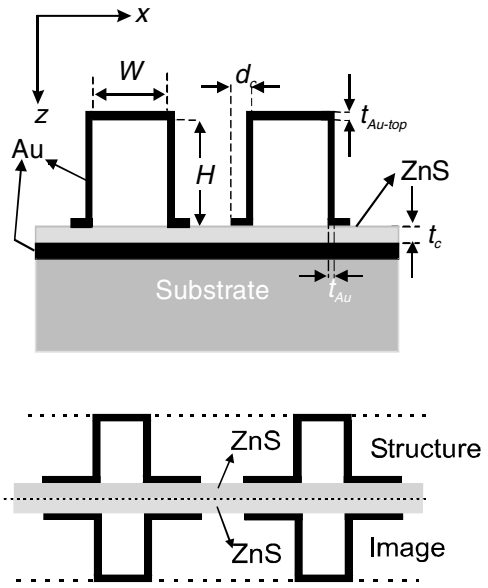


FIG. 1. Top: schematic of the nanostructures. The light is incident from the top; for TM polarization, H is in the y direction, perpendicular to the loop. Bottom: the equivalent LC circuit formed between the top staple structure and its image in the metal. The dashed lines are the reference planes for the effective permeability calculation.

 © 2005 The American Physical Society

rored in the metal surface, as shown in the bottom panel of Fig. 1. The period of the array is 600 nm, much smaller than the experimentally observed resonant wavelengths of between 4 and 7 μm. The area of the open gold ring loop and the size of the staple footings along with the thickness of the ZnS dielectric determine the inductance and capacitance, respectively, in the equivalent LC circuit. Three 1D samples (magnetic structures extending $\gg\lambda$ into the plane of Fig. 1) (A, B, C) and one 2D structure (as in Fig. 2, extent of each individual structure $\ll\lambda$) [sample (D)] were fabricated with different geometric parameters as shown in Table I. Figure 2 shows SEM micrographs of a 2D structure (D) and a 1D structure (B). Samples A and C have different footing sized (d_c) while samples A and B have different staple heights (H), with all other parameters held constant. Sample D has the same parameters as B except for the periodicity in the y direction and the removal of the nitride inside the loop.

The processing steps for the 1D samples are as follows. First, a 200 nm thick gold film is deposited on a silicon wafer by metal evaporation, followed by dielectric evaporation of a layer of ZnS ($n \sim 2.26$). Zinc sulfide is chosen because it is transparent throughout the $\sim$2 to 10 μm IR region. Next, a layer of Si_3N_4 is grown on the sample by plasma enhanced chemical vapor deposition (PECVD) ($n \sim 1.7$), followed by spun-on layers of a bottom antireflection coating (BARC) and an i-line photoresist (PR). A 1D 600 nm pitch grating is fabricated by interferometric lithography (IL) using a 355 nm, third-harmonic (YAG) laser source [8] [Fig. 3(a) and 3(b)]. The pattern is then transferred into the Si_3N_4 layer by reactive ion etching (RIE) and the residual PR and ARC are removed [Fig. 3(c)]. The etching is selective and does not affect the ZnS layer. Shadow evaporation is used to deposit Au on the top and on one sidewall of the Si_3N_4 gratings, and to form the footing on one side [Fig. 3(d)] of the staples. A second evaporation is carried out at the opposite tilt to form the completed ''staple'' structure. The angle of the metal evaporation controls the footing size. For the 2D samples, the same procedures are followed up to Fig. 3(c), then ARC and PR are spun on the wafer again, and another grating of the same period but perpendicular to the Si_3N_4 grating is fabricated, and further etching and lift-off processes are performed. For these 2D structures, the nitride is removed by plasma ashing after the metal has been deposited.

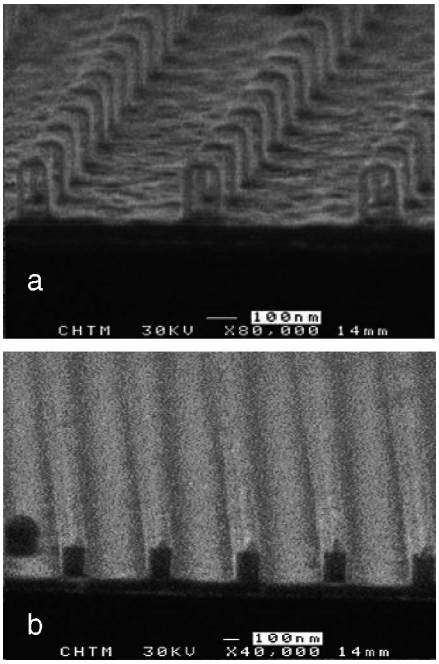


FIG. 2. SEM pictures of the arrays of 2D [sample D, shown in (a)] and 1D [sample B, shown in (b)] staple-shaped nanostructures.

Because of the anisotropic nature of the staple structure, we expect to couple to the resonant structure only when the magnetic field of the incoming light is perpendicular to the loop of the staple (TM polarization). Fourier transform infrared spectroscopy (FTIR) reflectance measurements were performed on these four samples for both TE and TM polarizations. The IR transmission is negligible due to the thick bottom metal, so the nanostructure optical properties are solely characterized by the resonance absorption. As expected, the TE measurements on all the samples did not show any structure [to avoid clutter in the figure, the TE reflectance is shown only for sample C in Fig. 4 (top)]. The TM measurements show reflectance dips that vary according to the sample geometry directly confirming the existence of structure-dependent magnetic activity for these staple arrays. The strong resonance, with as much as 50% absorption in a single layer of nanostructures, suggests that very high fields are generated in the nanostructures that will be of interest for future nonlinear optics experiments. As shown in Fig. 4 (top), sample A has the largest inductance and capacitance, so the resonance occurs at the lowest frequency. Sample C has the same inductance as sample A but a smaller capacitance, giving a higher frequency resonance (shorter wavelength). Samples D and B have the same capacitance as sample A and a smaller inductance, also giving a higher frequency resonance. Comparison between the 1D sample and 2D sample (samples B and D) with the same geometric parameters (not shown) was made; there is no significant difference either in the location of the resonance dip or the depth of the resonance. Experimentally, both samples B and D have relatively smaller reflection dips; the reason for this variation is under investigation.

TABLE I. Sample geometrical parameters (all dimensions in nm).

Sample	W	H	d_c	t_{Au}	$t_{\mathrm{Au\,top}}$	t_c
A	130	280	190	$\sim$15	$\sim$30	80
$B(D)$	130	180	190	$\sim$15	$\sim$30	80
C	130	280	90	$\sim$15	$\sim$30	80
Near-IR	80	50	20	10	20	20

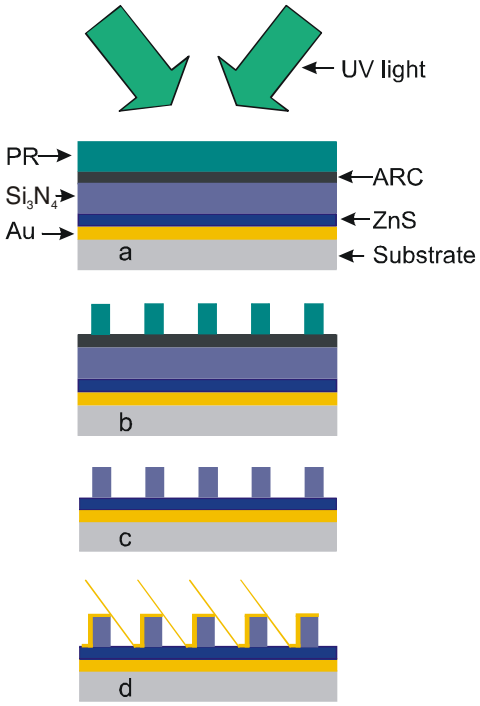


FIG. 3 (color online). Schematic diagrams of fabrication steps. (a) Interferometric lithography with two coherent UV beams. (b) The grating in PR after development. (c) The grating in the nitride layer after RIE etching. (d) Au deposition at an angle from one side.

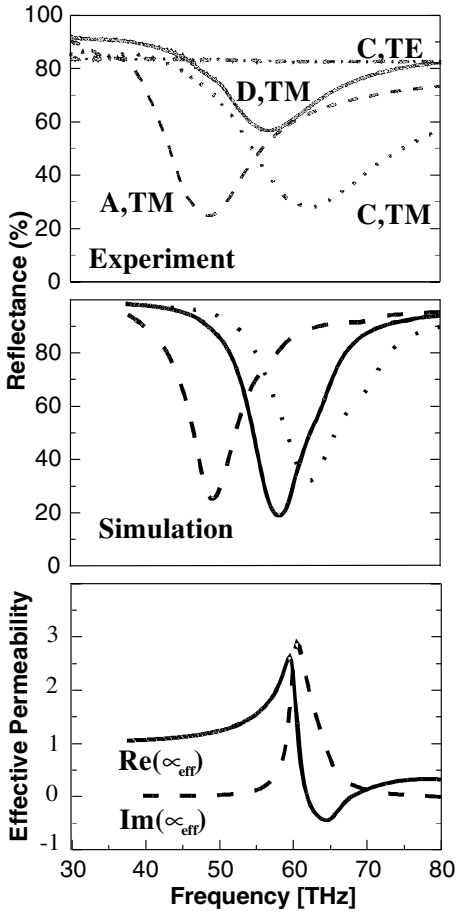


FIG. 4. Experimental (top) and simulation (middle) results for the reflectivity as a function of frequency for three different structures (cf. Table I). The resonance frequency is a function of the nanostructure parameters and is independent of array period, which is the same for all three structures. The resonance is observed only for TM polarization, i.e., incident magnetic field coupled into the inductive structure. The bottom panel shows the effective permeability extracted from the model for sample *C*.

Rigorous coupled wave analysis (RCWA) results for the three 1D structures are shown in Fig. 4 (middle). The RCWA is commonly used for analysis of scattering of electromagnetic waves from periodic structures [9]. Bulk values of the wavelength dependence of the Au optical properties were used [10]. The modeling explains the experimental data well except that the width of absorption dip of the experimental data is somewhat larger than that of the model. The broadening is probably due to the nonuniformity of the structures across the ~1 cm measured spot as well as to the nonideal material parameters.

To further investigate the strength of this magnetic activity, we extracted the effective permeability of sample *C* from the equivalent *LC* circuit (Fig. 1, bottom) [11,12]. The real part of the permeability shows a very strong modulation and is negative over a restricted frequency range; the imaginary part shows a complementary Lorentzian-like peak, features characteristic of a strong magnetic resonance. Extending the modeling across a range of structure parameters shows that a more negative permeability can be obtained by increasing the width *W* of the staple structures, i.e., increasing the fill factor.

Direct scaling of all dimensions of the present structure to extend these results to higher frequency was investigated. The results are shown in Fig. 5. The resonance frequency does not scale linearly with the inverse of the dimensional scaling factor; when the dimensions are reduced by a factor of 4, the resonance only shifts from 70 to 100 THz. At the same time, the minimum permeability becomes positive, evidencing weaker magnetic effects. This is consistent with previous results. Two terms contribute to the inductance: a structural term L_g related to the

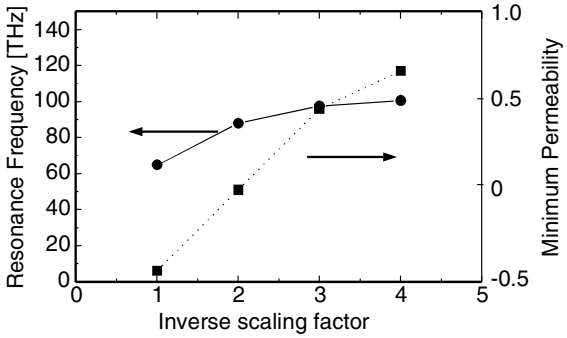


FIG. 5. Scaling of the present structure does not extend negative permeability to significantly higher frequencies as a result of inertial contributions to the structure inductance.

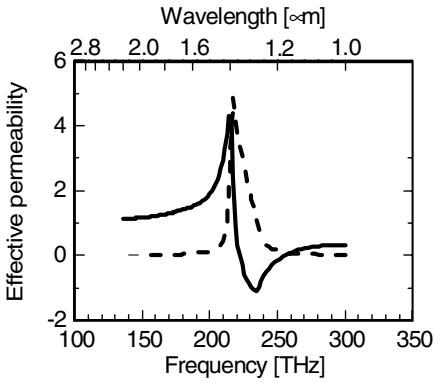


FIG. 6. Simulation of the effective permeability for an optimized structure extending the regime of negative permeability to near-IR wavelengths.

area of the loop, and an intertial term L_i associated with the finite electron mass. L_i is proportional to the circumference of the loop and inversely proportional to the thickness of the metal. The difficulty with simple scaling is the L_i is independent of the scale factor since the effects of reducing the loop perimeter and the metal thickness cancel. Consequently, the inductance and the resonance frequency vary only weakly with dimensional scaling.

The resonance properties of our experimental samples are not yet optimized: the ratio of the linewidth W (130 nm) to the pattern period (600 nm) results in a small fill factor; and the height of the loop H (280 nm) is much larger than W, so the ratio of the area to the perimeter is relatively small, increasing the relative importance of L_i. Increasing W, decreasing H so that $2H \sim W$, and only reducing the metal thickness slightly during the scaling improves the results dramatically. Figure 6 shows the simulated effective permeability for an optimized structure (with dimensions given in the last row of Table I). The pitch is reduced to 180 nm, again small compared to the resonance wavelength of 1.30 μm. The dielectric material was changed to SiO_2 to further decrease the capacitance. The real part of permeability has a strong modulation and a minimum as low as ~ -1 at a frequency of 230 THz (1.3 μm) in the center of the technologically important near-IR spectral region.

In summary, nanostructured magnetic resonators have been fabricated with resonances in the mid-IR spectral region. The measured reflectivity shows an absorption dip in good agreement with the simulation results. The simulation shows strong magnetic response and indicates negative magnetic permeability over a limited frequency range. Scaling to higher frequencies is investigated, and an optimized structure exhibiting strong magnetic resonance behavior at near-IR wavelengths is proposed. Combining this negative permeability structure with a negative permittivity structure will lead to metamaterials with negative refractivity in the IR. This can be accomplished by appropriately modifying the lower metal film, for example, by segmenting it in the orthogonal direction into an array of wires to decrease the effective permittivity [7]. This would also lift the present restriction to reflective operation. The independent control of the magnetic (staple) and electrical (wires) response based on simple geometrical changes will allow tuning of the electromagnetic response over a wide range.

Support for this work was provided by the ARO/MURI in "Deep Subwavelength Optical Nanolithography."

*E-mail address: brueck@chtm.unm.edu

[1] V. G. Veselago, Sov. Phys. Usp. **10**, 509 (1968).
[2] J. B. Pendry, Phys. Rev. Lett. **85**, 3966 (2000).
[3] D. R. Smith, Willier J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, Phys. Rev. Lett. **84**, 4184 (2000).
[4] R. Shelby, D. R. Smith, and S. Schultz, Science **292**, 77 (2001).
[5] Patanjali V. Parimi, Wentao T. Lu, Plarenta Vodo, and Srinivas Sridhar, Nature (London) **426**, 404 (2003).
[6] T. J. Yen, W. J. Padilla, N. Fang, D. C. Vier, D. R. Smith, J. B. Pendry, D. N. Basov, and X. Zhang, Science **303**, 1494 (2004).
[7] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, IEEE Trans. Microwave Theory Tech. **47**, 2075 (1999).
[8] X. Chen, S. H. Zaidi, S. R. J. Brueck, and D. J. Devine, J. Vac. Sci. Technol. B **14**, 3339 (1996).
[9] M. G. Moharam and T. K. Gaylord, J. Opt. Soc. Am. **72**, 1385 (1982).
[10] J. H. Weaver, C. Krafka, D. W. Lynch, and E. E. Koch, *Optical Properties of Metals*, Physics Data Vol. I and II (Fachinformationzentrum, Karlsruhe, Germany, 1981), Vol. 18-2.
[11] D. R. Smith and S. Schultz, Phys. Rev. B **65**, 195104 (2002).
[12] S. O'Brien and J. B. Pendry, J. Phys. Condens. Matter **14**, 6383 (2002).

Enhanced Infrared Transmission through Subwavelength Coaxial Metallic Arrays

Wenjun Fan, Shuang Zhang, Babar Minhas, Kevin J. Malloy, and S. R. J. Brueck*

Center for High Technology Materials and Electrical and Computer Engineering Department, University of New Mexico, Albuquerque, New Mexico 87106, USA

(Received 31 August 2004; published 24 January 2005)

We report the observation of enhanced near-infrared transmission through arrays of subwavelength coaxial metallic structures compared with that through comparable diameter hole arrays as a result of localized electromagnetic modes supported by the complex coaxial unit cell. Polarization and angle-dependent transmission measurements clearly demonstrate the coupling between this localized mode and delocalized surface plasmon modes. A generalized, multiple discrete states Fano line shape provides a good fit to the experimental results.

DOI: 10.1103/PhysRevLett.94.033902 PACS numbers: 42.25.Bs, 42.79.Dj, 78.66.–w

The observation of anomalously high transmission through periodic arrays of subwavelength holes in metal films has sparked considerable interest [1]. The small extent of these structures could lead to subwavelength-scale photonic circuits. Numerous studies of the transmission through hole arrays and through single apertures surrounded by periodic structures have been reported [2]. The concentration of the light in the apertures leads to an electric field enhancement that can be used to manipulate light-matter interactions and enhance nonlinear phenomena [3]. Surface plasma waves (SPWs), electromagnetic waves bound to each metal-dielectric interface (air and substrate), contribute to the enhanced transmission through metallic hole arrays. This association is often clear from simple momentum conservation arguments as features in the response as functions of angle and wavelength track the locations expected for SPW interactions, but the detailed origin of the enhanced transmission through such arrays is still under investigation [4–7].

Arrays of more complex unit cells offer additional degrees of freedom for optimizing the optical response. Baida [8] and Moreau [9] proposed that an array of metallic coaxial structures supporting attenuated guided electromagnetic modes would have substantially greater transmission and correspondingly larger field enhancements than simple hole arrays. In this Letter, we present the first experimental results for Au coaxial structures on a fused silica substrate and confirm this prediction. With an open annulus of width 182 nm and center radius of 194 nm (outer radius 376 nm) on a 1.03 μm period square array (individual open area of 0.33 μm^2; fractional open area of 31%), 79% transmission is observed at photon energy of 0.63 eV (1.97 μm). In comparison, a hole array with an outer diameter of 324 nm (same open area) shows a peak transmission of only 41% at the larger photon energy of 0.74 eV (1.68 μm) corresponding to a metal-substrate SPW coupling resonance. This very high transmission ($> 50\%$) for subwavelength apertures in metal films substantially increases the attractiveness of these materials for many applications. The enhanced transmission and peak wavelength shift are the result of the coupling between a localized mode (specifically, the TE_{11} mode) supported by the coaxial unit cell and the distributed surface plasma polaritons of the metal-dielectric interfaces. The transmission peak wavelength increases with increases of the coaxial gap and increases of the refractive index of fluids in the coaxial gap area, confirming the TE_{11} identification. The periodic array provides the necessary phase matching to allow coupling between all of these states and the radiation continuum. Fano line shapes, extended to describe coupling between the continuum and a multiplicity of discrete states, provide a good description of polarization-dependent and angle-dependent spectra.

We used a process based on interferometric lithography to fabricate the annular coaxial metal grids. Unlike the more common electron-beam lithography approach, there is no difficulty in extending the sample area for interferometric lithography, which is ideal for making large areas of periodic structures [10]. All the samples reported on here were uniformly fabricated over 2.5×2.5 cm^2 areas using a self-aligned process requiring only a single lithography step [11].

Three metallic annular coaxial grid samples and one simple hole-array sample were fabricated, all with 1.03 μm pitch and a metal (Au) thickness of 100 nm. For all of the coaxial samples, the inner-ring dot radius was 194 nm; the outer-ring radii were 298, 348, and 376 nm. The hole-array sample was fabricated with a radius of 324 nm. A scanning electron micrograph of the 348 nm coaxial array is shown in the inset to Fig. 1(a). Transmission spectra were recorded with a Fourier-transform infrared spectrometer at normal incidence with an incoherent light source, and are normalized to the transmission ($\sim 92\%$) of a bare substrate. The refractive index of the substrate, evaluated from the Fourier-transform infrared transmission curve, is 1.47. Figure 1(a) shows the transmission spectra with unpolarized incident light, equivalent to linearly polarized light for this symmetric geometry. All three of the coaxial array structures showed higher peak transmission than did the hole array. The peak transmission was weakly dependent on the width of the coaxial gap region, shifting to lower energy as the gap

 © 2005 The American Physical Society

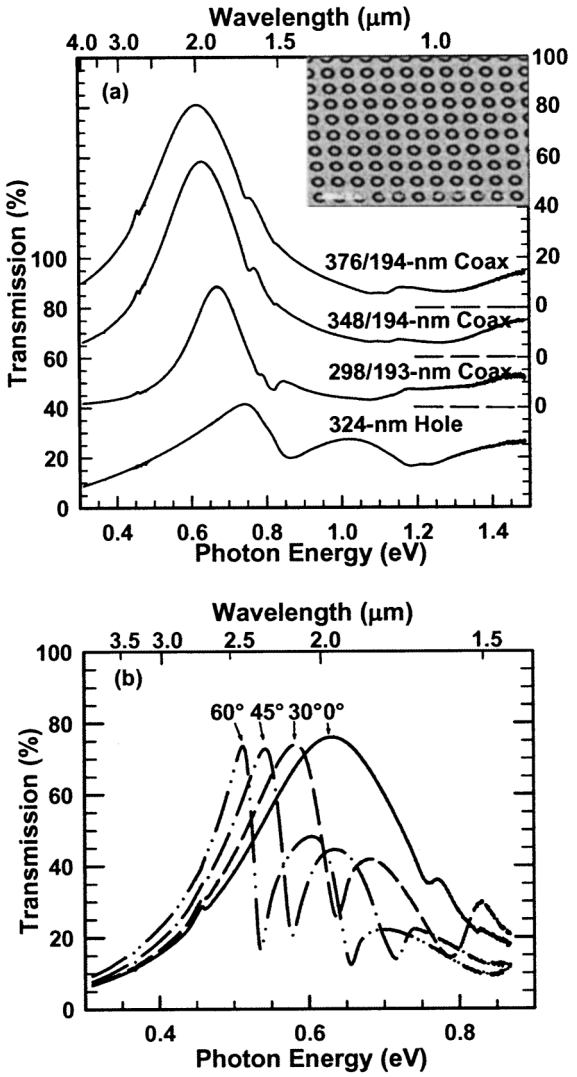


FIG. 1. (a) Measured transmission spectra for four samples. Inset: scanning electron microscopy of coaxial array with 1.03 μm pitch, 348 nm outer and 194 nm inner radii. (b) Angular dependence of TM transmission for 348/194 nm coaxial sample.

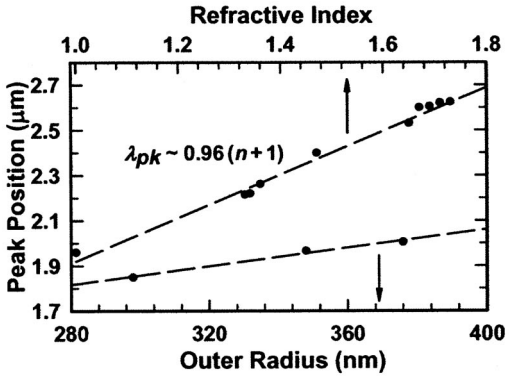


FIG. 2. Coaxial peak positions as a function of refractive index of fluids in the coaxial space (top curve for 348/194 nm sample) and as a function of the outer radius for the three experimental coaxial samples (bottom curve).

width was increased. As shown in Fig. 2, the peak transmission wavelength is linearly dependent on the annulus outer diameter. Nearly 80% peak transmission at 0.60 eV (2.05 μm) was observed for the largest gap width of 182 nm. In contrast, the simple hole array showed a peak transmission of only 41% for the same open area. The coaxial transmission peaks clearly reflect the coaxial geometry as well as the matching condition for coupling to SPWs based on the periodicity, with the peak transmission occurring at a lower energy than the SPW resonance. In contrast, the hole-array transmission peaks are coincident with the resonance predicted from the SPW dispersion relation. This gives additional degrees of flexibility for designing structures; the higher transmission through smaller holes also gives a stronger field amplification that may be useful for nonlinear-optics applications. The small feature on the spectra around 0.46 eV (2.7 μm) is instrumental, due to atmospheric CO_2 absorption in the measurement light path.

At normal incidence, there is no polarization dependence for this square symmetric pattern. However, off normal incidence, there is a strong difference between TM polarization, with an angle-dependent coupling to the various metal-substrate and metal-air SPW orders, and TE polarization for which the energies of the SPW coupling are independent of angle. Figure 1(b) shows several transmission spectra for TM polarization. Spectra were obtained for angles of incidence between 0° and 60° at 5° intervals. Because of limitations of the available wire-grid polarizer, results are only presented for energies less than 0.9 eV ($\lambda > 1.4$ μm). As the tilt angle is increased, the peak transmission shifts to lower energies and additional features resulting from SPW coupling (peaks) and Wood's anomalies (dips at the diffraction-order horizons) appear and progress to lower energies.

More insight into the origins of the enhanced transmission is obtained by measuring the transmission resonance as a function of the refractive index of the medium inside the gap. A series of measurements was carried out with different index matching fluids by sandwiching the 348/194 nm annular metal structure between two quartz plates. The peak transmission wavelength is proportional to the refractive index of the medium in the gap as shown in Fig. 2.

Two complementary approaches to modeling the results were pursued. Rigorous coupled wave analysis (RCWA) is used to obtain a detailed simulation. Because of the limitations of available computational capabilities, the RCWA code was only applied to normal-incidence geometries where symmetry significantly reduced the computational difficulty. A more intuitive, phenomenological, Fano line shape approach is used to extract an understanding of the multiple resonances and couplings involved in the problem [7,12]. This approach successfully demonstrates that the transmission is due to coherent interactions between SPW modes supported by the metal film and a localized reso-

nance, corresponding to the TE_{11} mode of the coaxial structures.

Normal-incidence, RCWA simulations [13] were carried out for all of the samples. In order to obtain convergence with a reasonable number of diffraction orders, the simulation was carried out for a square coaxial case keeping the same open area as the annular experimental geometry. The comparison of experiment and simulation for the high transmission case (376/194 nm) is shown in Fig. 3(a). The simulation intensity peak was a bit higher than the measurement, but overall the experimental results are very well matched, even reproducing the fine structure around 0.8 eV, which corresponds to coupling to the metal-glass SPW.

The analysis based on Fano line shapes, describing the interference between direct scattering into transmitted photons and coupling via discrete states such as the metal-air and the metal-glass SPWs as well as coaxial resonances, provides additional insight. The Fano line shape [14] extended to a number of discrete states is expressed as

$$I(\omega) \propto \left(1 + \sum_{\nu} \frac{q_\nu}{\Delta\varepsilon_\nu}\right)^2 \Big/ \left[1 + \left(\sum_{\nu} \frac{1}{\Delta\varepsilon_\nu}\right)^2\right], \qquad \Delta\varepsilon_\nu = \frac{\omega - \omega_\nu}{\Gamma_\nu/2}, \tag{1}$$

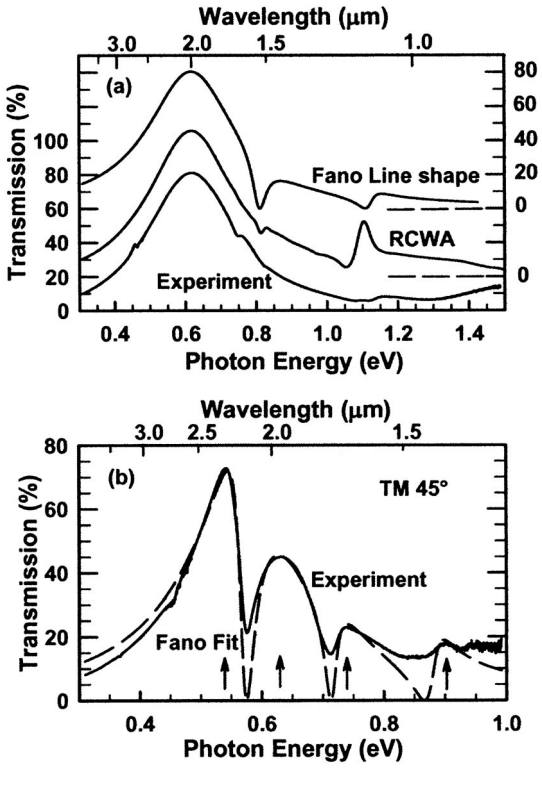


FIG. 3. (a) Experiment, RCWA, and Fano fit simulations for the 376/194 nm coaxial sample at normal incidence, offset for clarity. (b) Experimental transmission and Fano fit for 348/194 nm coaxial sample for TM polarization at an incident angle of 45°. The resonance positions for the Fano fit are indicated.

where ω_ν is the resonant state frequency, Γ_ν is the linewidth (FWHM) and q_ν is the Breit-Wigner-Fano [15] coupling coefficient for the νth discrete state. From Eq. (1), for real q_ν, the Fano curves necessarily extend to zero intensity. While the corresponding experimental features are not as pronounced, due to structural variations and bandwidth limitations, the asymmetric line shapes fit the experimental spectra very well, particularly away from normal incidence. The Fano fit for the 376/194 nm coaxial structure is shown in Fig. 3(a) along with the RCWA calculation and the experimental data. Overall the agreement is quite good, with the exception of the zero transmission discrepancy noted above. As a result of this difference, there is significant uncertainty in the fit around the SPW resonance for the normal-incidence case. At higher incident angles, the Fano fit is much better defined, the result for 45° incidence TM transmission spectrum, Fig. 3(b), shows excellent agreement between this phenomenological model and the experiment (with some deviations at the highest energies where the experimental peaks are least well-defined).

By carrying out this fitting procedure for all of the experimental spectra, the dispersion of the various discrete resonances contributing to the Fano line shape can be mapped out. These dispersion lines strongly suggest coupling [4,16,17] between the localized coaxial mode and the SPW modes on both the metal-substrate and metal-air interfaces. The energy eigenmode matrix can be written as

$$\begin{pmatrix} E_a & 0 & 0 \\ 0 & E_b & 0 \\ 0 & 0 & E_c \end{pmatrix} = \mathrm{diag}\begin{pmatrix} E_{\mathrm{coax}} & \kappa_{12} & \kappa_{13} \\ \kappa_{12}^* & E_{\mathrm{SP1}} & 0 \\ \kappa_{13}^* & 0 & E_{\mathrm{SP2}} \end{pmatrix}, \tag{2}$$

where κ_{12} and κ_{13} are coupling coefficients between the localized coaxial mode and the SPW modes, $E_{a,b,c}$ are the Fano fit resonant states, E_{coax} is the localized coaxial mode, and E_{SP} are the SPW modes, which are given approximately be a kinetic model (for a continuous metal film)

$$m\lambda = \Lambda\left(\mathrm{Re}\sqrt{\frac{\varepsilon_d\varepsilon_m}{\varepsilon_d + \varepsilon_m}} \pm \sin\theta\right), \tag{3}$$

where Λ is the pitch of the sample and ε_d, ε_m are the dielectric constants of the metal and dielectric (air and glass), respectively. For each angle, we can solve for the unknowns in Eq. (2), namely E_{coax}, κ_{12}, and κ_{13}. Physically, κ_{12} and κ_{13} should be relatively constant as the incident angle is varied; optimal results were obtained for $|\kappa_{12}| = 0.039$ and $|\kappa_{13}| = 0.014$. The coupling between the coaxial mode and the metal-substrate SPW is stronger than the coupling between the coaxial mode and the metal-air SPW. These fixed coupling strength numbers were used to evaluate the energy for the localized coaxial mode, as shown in Fig. 4(b) (hollow symbols). The average localized energy state from this coupling optimization is 0.588 eV, with less than a 2% variation across the fitted angular range.

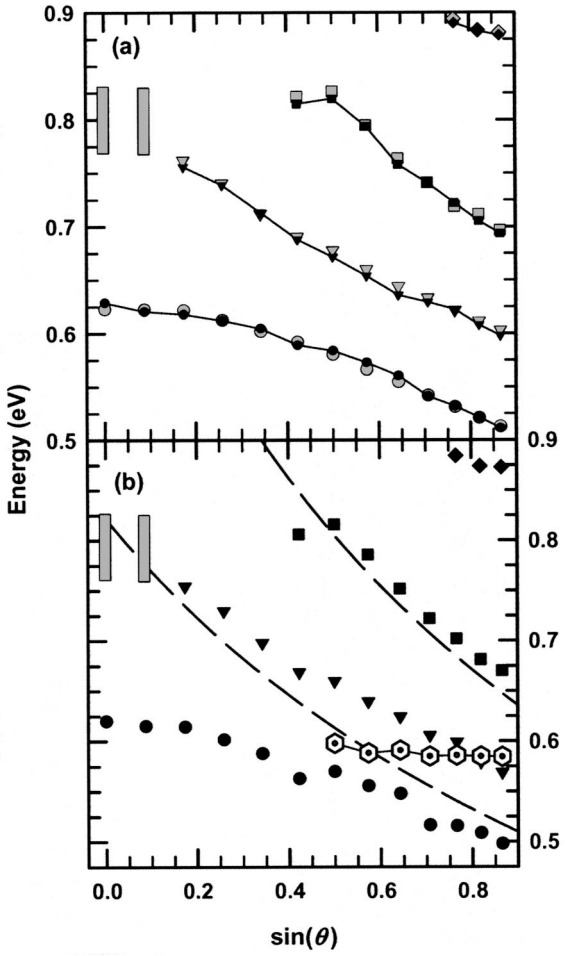


FIG. 4. (a) Peak positions from experimental spectra (gray) and from Fano fits (black). At low angles (0° and 5°) the features corresponding to the metal-glass SPW are not simple peaks and the entire region is denoted by the shaded rectangular bars. (b) Spectral resonances extracted from Fano fits (black symbols), SPW energies (dashed lines), and the localized coaxial energy from perturbation calculations (hollow symbols).

Recently, Baida *et al.* reported [18] extensive modeling of the enhanced transmission through coaxial arrays and concluded that the dominant transmission was due to the cutoff of the TE_{11} coaxial mode given, for a perfectly conducting metal, by

$$\lambda_{\text{cutoff,TE}_{1,1}} \approx \pi n(a + b) \tag{4}$$

where n is the refractive index of the medium filling the coaxial structure, and a and b are the inner and outer radii, respectively. The linear variation of the resonance wavelength as a function of both the outer radius b and the refractive index of the index matching fluid supports this identification (Fig. 2). Note that initially [8,9], it was proposed that the TEM_{00} mode, with no cutoff frequency, might play a large role in the enhanced transmission. By simple symmetry considerations, there is no coupling between the antisymmetric radial electric field of the TEM_{00} mode and a uniform plane wave excitation and, therefore, the TEM_{00} mode plays no role in the enhanced transmission.

In conclusion, we demonstrated the use of interferometric lithography to make large area annular coaxial metal arrays. Enhanced transmission compared to hole arrays was shown experimentally and theoretically with a peak transmission of 80% realized. Coupling between multiple localized guided modes, specifically the TE_{11} of the coax and the SPW resonances of the metal sheet and the radiation continuum was demonstrated. Further studies are underway to explore the applications of these nanophotonic structures in nonlinear-optics and infrared sensing.

This work was supported by U.S. Army Research Office under the MURI program in Deep Subwavelength Optical Nanolithography.

*Electronic address: brueck@chtm.unm.edu

[1] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, Nature (London) **391**, 667 (1998).
[2] H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal and T. W. Ebbesen, Science **297**, 820 (2002).
[3] W. L. Barnes, Alain Deteux and T. W. Ebbesen, Nature (London) **424**, 824 (2003).
[4] L. Martin-Moreno *et al.*, Phys. Rev. Lett. **86**, 1114 (2001).
[5] Q. Cao and P. Lalanne, Phys. Rev. Lett. **88**, 057403 (2002).
[6] D. S. Kim, S. C. Hohng, V. Malyarchuk, Y. C. Yoon, Y. H. Ahn, K. J. Yee, J. W. Park, J. Kim, Q. H. Park and C. Lienau, Phys. Rev. Lett. **91**, 143901 (2003).
[7] M. Sarrazin, J.-P. Vigneron and J.-M, Vigoureux, Phys. Rev. B **67**, 085415 (2003).
[8] F. I. Baida and D. Van Labeke, Opt. Commun. **209**, 17 (2002).
[9] A. Moreau, G. Granet, F. I. Baida, and D. Van Labeke, Opt. Express **11**, 1131 (2003).
[10] S. R. J. Brueck, Microelectron. Eng. **42**, 145 (1998).
[11] Process flow: deposit blanket SiN layer; define photoresist (PR) hole array by interferometric lithography; anistropically etch SiN in holes; isotropically etch SiN to generate an undercut defining the gap; deposit and lift off 1st Au layer to form central dots; spin a thick layer of PR and etch back to expose SiN; etch off SiN, leaving PR filler; deposit and lift off 2nd Au film to form outer mesh.
[12] C. Genet, M. P. Van Exter and J. P. Woerdman, Opt. Commun. **225**, 331 (2003).
[13] B. K. Minhas, W. Fan, K. Agi, S. R. J. Brueck and K. J. Malloy, J. Opt. Soc. Am. A **19**, 1352 (2002).
[14] U. Fano, Phys. Rev. **124**, 1866 (1961).
[15] J. F. Scott, Rev. Mod. Phys. **46**, 83 (1974).
[16] A. Christ, S. G. Tikhodeev, N. A. Gippius, J. Kuhl and H. Giessen, Phys. Rev. Lett. **91**, 183901 (2003).
[17] J. Bellessa, C. Bonnand, J. C. Plenet and J. Mugnier, Phys. Rev. Lett. **93**, 036404 (2004).
[18] F. I. Baida, D. Van Labeke, G. Granet, A. Moreau, and A. Belkhir, Appl. Phys. B **79**, 1 (2004).

Vortex Configurations, Matching, and Domain Structure in Large Arrays of Artificial Pinning Centers

S. B. Field,[1] S. S. James,[1,*] J. Barentine,[1] V. Metlushko,[2,†] G. Crabtree,[2] H. Shtrikman,[3] B. Ilic,[4] and S. R. J. Brueck[4]

[1]*Department of Physics, Colorado State University, Fort Collins, Colorado 80523*
[2]*MSD, Argonne National Laboratory, Argonne, Illinois 60439-4845*
[3]*Weizmann Institute of Science, Rehovot, Israel*
[4]*Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87131*

(Received 7 February 2000; published 28 January 2002)

High-resolution scanning Hall probe microscopy has been used to image vortex configurations in very large periodic arrays of artificial pinning sites. Strong matching effects are seen at fields where either one or two vortices can sit at a site; with three vortices per site, however, no clear matching is observed. Matching effects have also been observed at several fractional multiples of the matching field, including 1/5, 1/4, 1/3, 1/2, and 3/4. These fractional values are characterized by striking domain structure and grain boundaries.

DOI: 10.1103/PhysRevLett.88.067003 PACS numbers: 74.60.Ge, 74.25.Ha, 74.76.Db

The behavior of superconducting vortices in the presence of a periodic array of holes reveals rich and unexpected static and dynamic phenomena. In such arrays, flux can be quantized either "in" the holes (as a circulating current around their peripheries), or in the interstitial regions between the holes as ordinary Abrikosov vortices. At low fields, the holes act as very strong pinning centers for flux, and it is thus possible for holes to contain *multiquantum* vortices. However, the energy of such multiquantum vortices increases rapidly with the number of quanta n [1]. This leads to a *saturation number* n_s, beyond which it is energetically favorable for any additional flux to enter the interstitial regions between the holes as Abrikosov vortices. Transport [2–4] and magnetization [5–9] studies show distinctive features at matching fields where the vortex structure is commensurate with the hole array. Interesting configurations also arise at fractional matching fields, where the occupation number of each hole or the number of interstitial vortices forms a superstructure locked to the basic hole array [5,9].

Although magnetization and transport studies have elucidated much of the basic phenomenology of the array/vortex system, they can measure only its *global* properties and cannot deduce details of the *local configurational* vortex state. Recently, Lorentz [10] and scanning Hall probe [11] microscopies have been used to image vortex configurations, but only over relatively small areas of $\approx 10 \times 10$ pinning sites. In this Letter, we present large-area scanning Hall probe microscope (SHM) images of vortex configurations in arrays containing $\approx 10^6$ holes. Our images span some 5000 holes, and so yield important information on the *large-scale* structure of the vortex configurations. These studies reveal striking multiquantum and interstitial vortex patterns in a square-periodic hole array. At fractional matching fields, we resolve distinctive domains of phase-slip related vortex superstructures that are separated by domain walls with characteristic internal structure.

The Hall probe's active area coupled with its scan height above the surface yield an effective FWHM resolution of 3.4 μm. The sample investigated was a 100-nm-thick niobium film with 0.3-μm-diameter holes on a square lattice. The lattice constant a was 1.870 μm, yielding an expected first matching field $H_m = \Phi_0/a^2 = 5.913$ G. The array was produced using a lithographic technique based on the interference of light [8], which yields very large and uniform arrays (here ≈ 2 mm $\times$ 3 mm). Our data were taken far from the edges of the array. Takezawa and Fukushima [12] have calculated the pinning energy of a hole vortex relative to one in the bulk. For parameters similar to ours, they find a very deep pinning potential of ≈ 100 eV. Comparing this to typical bulk Nb pinning energies of ≈ 50 meV [13], it is clear that pinning from the hole array dominates pinning from defects in the Nb film.

SHM images near several matching fields are shown in Fig. 1. For each image, the samples were field cooled from above $T_c = 8.37$ K to a base temperature below 3 K where the images were taken. The top row of Fig. 1 shows the progression in the vortex configurations as the applied field H is varied around $\overline{H} \equiv H/H_m = 0$. At $\overline{H} \approx 0$ we see several isolated vortices (dark spots). The actual vortex diameter is much smaller than these spots, whose size is determined by the Hall probe resolution. A somewhat darker spot, just to the right and below center, appears in all scans and is presumably associated with a physical defect in the hole lattice, which allows extra flux to sit at that point. As the field is increased from zero, vortices begin to enter the sample in larger numbers, as seen in the $\overline{H} = 0.075$ and 0.15 images. A similar progression is observed for *negative* field increments away from zero, as shown at $\overline{H} = -0.075$ and -0.15. Here the vortices appear as light spots, indicating that they are oppositely directed to those seen in positive fields.

As we continue to increase the field towards what we will be able to identify as the first matching field H_m

 © 2002 The American Physical Society

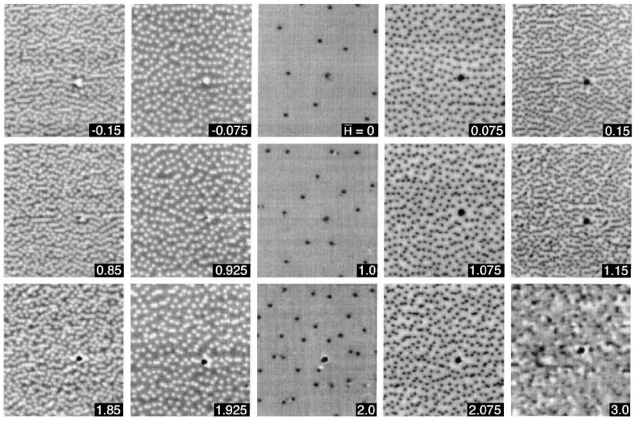


FIG. 1. SHM images obtained at applied magnetic fields H near the matching fields $\overline{H} \equiv H/H_m = 1$, 2, and 3. The scans are 124 μm $\times$ 138 μm, contain some 5000 holes, and have a full scale of 0.73 G. Note the striking similarity between the vortex configurations near $\overline{H} = 1$ and 2 and those near $\overline{H} = 0$. The smooth backgrounds at $\overline{H} = 1$ (2) consist of holes uniformly filled with 1 (2) vortices. At $\overline{H} = 3$, however, the vortex configuration is highly disordered because of weakly pinned interstitials competing with vortices in holes. The dark black spot is a local defect in the sample.

($\overline{H} = 1$), we begin to see a very remarkable progression (Fig. 1, second row). For instance, at an applied field of $\overline{H} = 0.85 = 1 - 0.15$, the image looks statistically identical to the image at $\overline{H} = -0.15$. And, remarkably, the image at $\overline{H} = 1$ is essentially indistinguishable from that at $\overline{H} = 0$. It is important to note here that in each image the average value of the field has been subtracted out. This is why the $\overline{H} = 1$ image at about 6 G and the $\overline{H} = 0$ image appear to have about the same overall gray level. We can then understand the appearance of the $\overline{H} = 1$ scan as follows. The field increment from an exactly $H = 0$ image (no spots) to the next no-spot image is 5.929 G, very close to the first matching field of 5.913 G deduced from the lattice geometry. Thus, we identify this later spot-free image at 5.929 G with the first matching field or $\overline{H} \equiv 1$. The smooth gray background near $\overline{H} = 1$ can then be explained in the following way. Exactly at $\overline{H} = 1$, each hole contains exactly one vortex, leading to a dense and uniform configuration of vortices. The spatial resolution of our Hall probe is not good enough to image individual vortices when they are this close together *and perfectly ordered* [14], so this regular array of vortices appears as a smooth gray background. The few black spots in the $\overline{H} \approx 1$ image are thus (easily visible) *extra* vortices.

The third row of Fig. 1 shows the vortex configurations as the field is further increased through the second matching field ($\overline{H} = 2$). A very similar progression is observed. At $\overline{H} \approx 2$, there is again a smooth background populated by a few black spots. Again, we interpret this background as the highly ordered state with now *two* vortices sitting in each hole; black spots for $\overline{H} > 2$ are again extra vortices, and white spots for $\overline{H} < 2$ now represent a hole with only *one* vortex. At the *third* matching field ($\overline{H} = 3$), however, the appearance is radically different (Fig. 1, bottom right). Instead of the smooth gray background we would expect if three vortices sat in each hole, we see a rather muddled image with no discernible structure. We believe this different appearance results from the sudden presence of *interstitial* vortices above $\overline{H} = 2$.

To probe this issue more directly, we have taken close-up scans just below and above $\overline{H} = 2$. Figure 2(a) shows a 25 μm $\times$ 28 μm scan at a field $\overline{H} = 2 - 0.084$, and Fig. 2(b) shows $\overline{H} = 2 + 0.084$. Also shown is the least-squares-fit positions of the holes as determined from an image taken at $\overline{H} = 1/2$ where the hole positions are unambiguous [see, e.g., Fig. 3(b)]. We also have used an absolute position sensor [15] mounted on the scanner head to compensate for any possible drifts between images. In Fig. 2(a) each white spot (i.e., a "missing" vortex) clearly sits directly on a hole, indicating as well that all *vortices* also sit on the holes. However, in Fig. 2(b) the black spots, which are now extra vortices, sit variously on holes *or* on interstitial sites.

Large holes "spread out" the fields of the hole vortices, lowering their energies and leading to stronger pinning. Thus the maximum number of vortices n_s that can occupy a hole depends on the hole radius R [1,16]. With the sputtering/lift-off lithography used here [8], there will be inevitable fluctuations in the hole diameters, perhaps yielding $n_s = 2$ for some holes and $n_s = 3$ for others. Vortices nucleating near $n_s = 2$ holes would end up as interstitials, while those near $n_s = 3$ sites could sit in a hole. This random-appearing admixture of vortices on holes and in interstitial sites would, by $\overline{H} = 3$, lead to the very disordered configuration observed at that field (Fig. 1).

It is also possible to have matching effects at noninteger multiples of H_m. Indeed, distinct features at $\overline{H} = 1/4$, $1/2$, and possibly $1/5$ (or $3/16$), $1/8$, and $1/16$ have

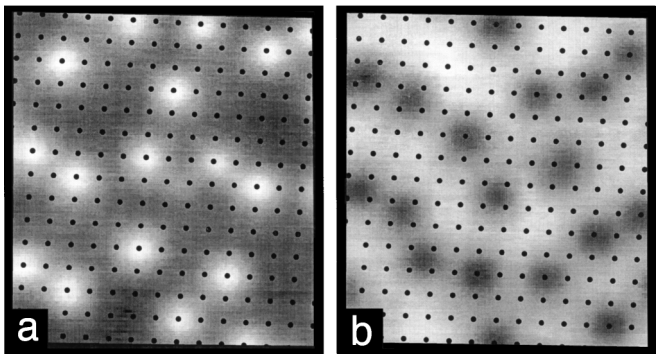


FIG. 2. Vortex configurations (a) just below ($\overline{H} = 1.916$) and (b) just above ($\overline{H} = 2.084$) the second matching field. The small dark circles are the positions of the holes. Below $\overline{H} = 2$, the vacancies (white spots) all sit directly on holes; thus all vortices must as well. Above $\overline{H} = 2$, the extra vortices (dark spots) sit on both hole *and* interstitial sites. 25 μm $\times$ 27 μm field is shown.

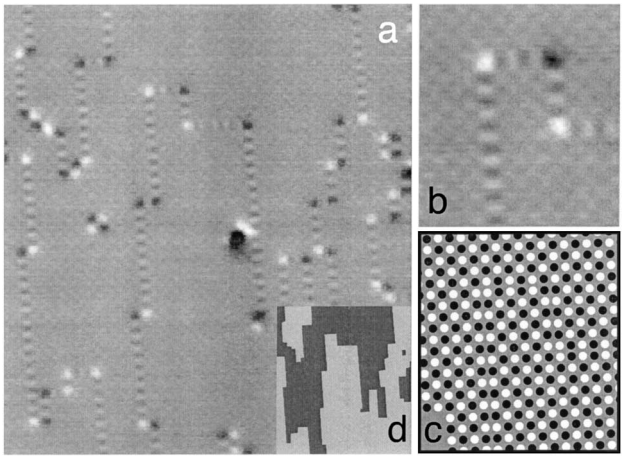


FIG. 3. (a) 122 μm $\times$ 138 μm image of the vortex configuration at $\overline{H} = 1/2$. There are large areas of vortices arranged in a checkerboard pattern; domains [outlined in (d)] of opposite checkerboard polarity are separated by stripelike grain boundaries. A close-up of this configuration is shown in (b) and schematically in (c).

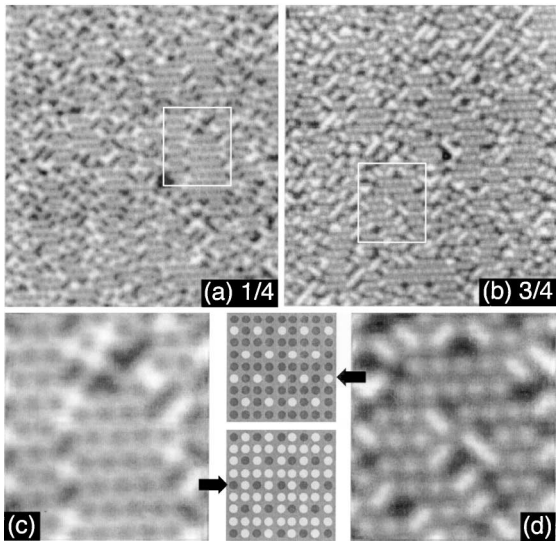


FIG. 4. 122 μm $\times$ 136 μm images of vortex configurations at $\overline{H} = 1/4$ (a) and 3/4 (b). Selected regions are expanded in (c) and (d). The configurations consist of relatively small well-ordered regions surrounded by complex disordered boundaries. Also shown schematically are the inferred vortex arrangements in the ordered regions. Light circles represent empty holes, and dark circles occupied holes.

been observed in magnetization studies [5,17], and Harada *et al.* [10] have imaged noninteger matching at 1/4, 1/2, 3/2, and 5/2. In our imaging experiments, we also observe matching effects at several of these fractional fields, as well as some not previously observed. We find that vortex configurations at fractional matching fields are characterized by striking domain structure and associated grain boundaries.

By far the strongest submatching effects occur at $\overline{H} = 1/2$. Figure 3(a) shows that the configuration consists of large areas of well-matched vortices separated by curious stripelike boundaries. The close-up in Fig. 3(b) reveals that the smooth regions consist of vortices occupying every other hole in a checkerboardlike fashion. Regions of different "polarity"—with the vortices occupying either the "red" or "black" squares of the checkerboard—are separated by striped grain boundaries [18]. Figure 3(c) schematically shows how a polarity shift results in rows or columns of alternating *pairs* of vortices and vacancies, which appear in the images as the striking boundary features. At the junctions between north-south and east-west boundaries, there is always a bright or dark spot, whose origin is again clear from Fig. 3(c). From considerations of the boundaries, we have mapped out the domains of differing polarity [inset, Fig. 3(a)]. One curious feature of these grain boundaries is that they run predominately north-south, indicating a breaking of the square symmetry of the lattice in these two directions. Our images demonstrate that the simple domain wall structure seen on small scales [11] yields unexpectedly complex patterns on large scales.

Weaker but still striking matching effects occur at $\overline{H} = 1/4$ as well as its *complimentary* [19] field of $\overline{H} = 1 - 1/4 = 3/4$ (Fig. 4). Zooming in on the boundary of such a patch [outlined in Figs. 4(a) and 4(b)] allows us to explore the vortex configuration in detail [Figs. 4(c) and 4(d)]. The vortex configuration in the ordered regions is shown to scale schematically in Fig. 4. Notice that the configuration may be viewed in terms of alternating empty and half-filled rows [10,20]; this again indicates the presence of a symmetry-breaking field which selects this particular grain orientation. One possible source of this asymmetry could be a small difference in the interhole spacing a in the horizontal and vertical directions. However, no such asymmetry is visible at the $\approx$0.1% accuracy of our optical diffraction measurements of a.

The origin of the intricate vortex patterns at submatching fields can be understood from their configurational energies. In our samples, the vortex spacing is large compared to the penetration depth and the energy of a configuration can be estimated from nearest-neighbor interactions only. At 1/2 filling, the observed vertical and horizontal domain walls contain nearest neighbors on half the lattice sites, whereas a 45° boundary, for example, would have nearest neighbors on all the lattice sites along the boundary and thus represent a much higher energy. The energy of the domain wall is considerably higher than that of the ordered phases where neighboring lattice sites are never occupied. This favors few boundaries and large ordered regions, as observed. In contrast, at 1/4 filling the vortices are substantially farther apart, lowering their repulsive energy. The energy differences between domain walls and ordered phases are less pronounced, and the possible domain wall configurations are also closer in energy. Therefore more domain wall configurations appear with greater frequency than at 1/2.

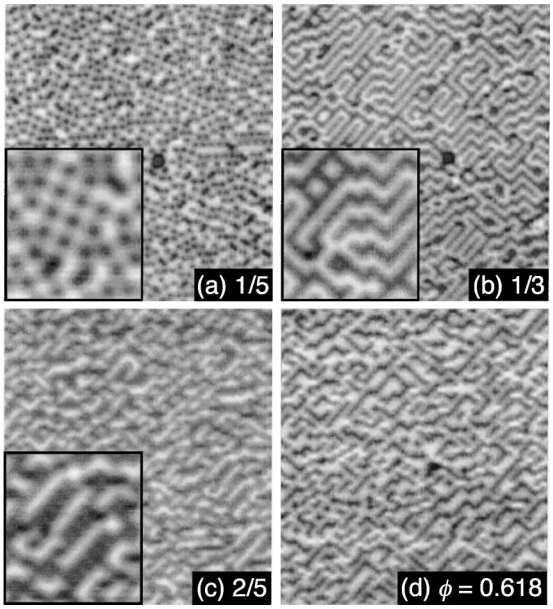


FIG. 5. 122 μm $\times$ 137 μm images of vortex configurations at several fractional values of $\overline{H}$. Also shown is the "irrational" field $\phi = (\sqrt{5} - 1)/2$.

Intuitive analysis is less obvious at other fractions (Fig. 5), but one can turn to simulations for guidance. At $1/3$ filling, Reichhardt and Grønbech-Jensen [21] find a disordered state containing many filled 45° diagonal rows separated by two empty diagonals, as seen in Fig. 5(b). An ordered structure of this type has also been predicted by Watson [20] for a "zero range" repulsive potential and by Teitel and Jayaprakash [22] in superconducting wire networks. At $1/5$ filling, filled diagonals at $\arctan(1/2) = 26.6°$ separated by four empty diagonals are the lowest energy configuration [5,20,21]. Again, since the vortex density is low, other configurations have only slightly higher energies and occur frequently. The same kinds of imperfect diagonal structures of filled and empty rows can be seen at $2/5$ filling [Fig. 5(c)], and at the irrational field of $(\sqrt{5} - 1)/2 \approx 0.618$ [Fig. 5(d)], indicating their relatively low configurational energies. The frequent occurrence of configurational disorder obscures any long-range distinction between rational and irrational filling in Fig. 5.

Vortex configurations in square-periodic hole arrays thus reveal remarkably complex patterns reflecting the interplay between the pinning energy of the hole array and the interaction energy of the multiquantum and interstitial flux structures. Our images show clear multiquantum occupation of the holes up to saturation, followed by the appearance of interstitial Abrikosov vortices at higher fields. At fractional filling factors, domains of well-ordered vortices are separated by a network of domain boundaries displaying remarkably simple order for half filling and increasing complexity at other fillings. There appears to be a nontrivial set of ordered fractional states, a complete inventory of which awaits a general theory correctly incorporating the several competing interactions present. Similar to other collective vortex phenomena, these competing ground states can be mapped onto a variety of statistical models that describe the basic phenomena of many other condensed matter systems [23].

We would like to acknowledge helpful discussions with C. Reichhardt and C. Olson. This work was supported by NSF Grant No. DMR-9701532 (S. B. F., S. S. J., J. B.) and the U.S. Department of Energy, Office of Basic Energy Sciences-Materials Sciences, under Contract No. W-31-ENG-38 (V. M., G. W. C.).

*Present address: IRC in Superconductivity, Madingley Road, Cambridge CB3 0HE, United Kingdom.
†Present address: Department of Electrical Engineering and Computer Science, University of Illinois at Chicago, Chicago, IL 60607-0024.

[1] G. S. Mkrtchyan and V. V. Shmidt, Sov. Phys. JETP **34**, 195 (1972).
[2] A. T. Fiory, A. F. Hebard, and S. Somekh, Appl. Phys. Lett. **32**, 73 (1978).
[3] L. Van Look *et al.,* Phys. Rev. B **60**, R6998 (1999).
[4] T. Puig *et al.,* Phys. Rev. B **58**, 5744 (1998).
[5] M. Baert *et al.,* Europhys. Lett. **29**, 157 (1995).
[6] V. V. Moshchalkov, Phys. Rev. B **54**, 7385 (1996).
[7] V. V. Moshchalkov *et al.,* Phys. Rev. B **57**, 3615 (1998).
[8] V. Metlushko *et al.,* Phys. Rev. B **59**, 603 (1999).
[9] V. Metlushko *et al.,* Phys. Rev. B **60**, R12 585 (1999).
[10] K. Harada *et al.,* Science **274**, 1167 (1996).
[11] A. N. Grigorenko *et al.,* Phys. Rev. B **63**, 052504 (2001).
[12] N. Takezawa and K. Fukushima, Physica (Amsterdam) **228C**, 149 (1994).
[13] C.-H. Sow, K. Harada, A. Tonomura, G. Crabtree, and D. G. Grier, Phys. Rev. Lett. **80**, 2693 (1998).
[14] D. Davidovic *et al.,* Phys. Rev. B **55**, 6518 (1996).
[15] S. B. Field and J. Barentine, Rev. Sci. Instrum. **71**, 2603 (2000).
[16] H. Nordborg and V. M. Vinokur, Phys. Rev. B **62**, 12 408 (2000).
[17] V. Metlushko *et al.,* Solid State Commun. **91**, 331 (1994).
[18] C. Reichhardt *et al.,* Phys. Rev. B **54**, 16 108 (1996).
[19] Indeed, for *all* configurations observed at fractional field $\overline{H} < 1$, essentially identical configurations (with vortices and vacancies exchanged) are observed at a field $1 - \overline{H}$.
[20] G. I. Watson, Physica (Amsterdam) **246A**, 253 (1997).
[21] C. Reichhardt and N. Grønbech-Jensen, Phys. Rev. B **63**, 054510 (2001).
[22] S. Teitel and C. Jayaprakash, Phys. Rev. B **27**, 598 (1983).
[23] G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Rev. Mod. Phys. **66**, 1125 (1994).

PHYSICAL REVIEW B VOLUME 60, NUMBER 18 1 NOVEMBER 1999-II

Interstitial flux phases in a superconducting niobium film with a square lattice of artificial pinning centers

V. Metlushko, U. Welp, G. W. Crabtree, R. Osgood, and S. D. Bader
MSD, ANL, Argonne, Illinois 60439-4845

L. E. DeLong
University of Kentucky, Lexington, Kentucky 40506-0055

Zhao Zhang and S. R. J. Brueck
University of New Mexico, Albuquerque, New Mexico 87131

B. Ilic, K. Chung, and P. J. Hesketh
EECS, University of Illinois at Chicago, Chicago, Illinois 60607

(Received 17 June 1999)

We demonstrate that a square lattice of artificial pinning centers in a superconducting Nb film induces the formation of highly ordered interstitial vortex phases with different symmetries for external magnetic fields as high as the eighth matching field. These ''supermatching'' phases are identified by distinct differences in the behavior of their critical currents, magnetoresistivity, and magnetization. Our results are consistent with predictions of supermatching lattice symmetries by recent numerical simulations. [S0163-1829(99)51142-9]

It is well known that vortex-vortex interactions in the mixed state of a type II superconductor generally lead to the formation of a triangular flux-line lattice (FLL) that is distorted in the presence of random pinning. The introduction of a regular lattice of artificial pinning centers (APC's) into a superconducting thin film provides additional strong pinning interactions that induce prominent anomalies in the temperature and magnetic-field dependences of the magnetization, electrical resistance, and critical current I_c.[1–5] The anomalies are particularly pronounced at ''matching fields'' H_n, where the density of flux lines (FL's) coincides wit the density of the APC, and essentially all of the magnetic field inside the film resides in quantized fluxoids centered on the APC. However, there is a maximum number $n_s = D/4\xi(T)$ of flux quanta [where D is the APC characteristic size and $\xi(T)$ is a temperature-dependent coherence length] that can be trapped by the APC.[6] Further increases in the applied field above $H_n(n_s)$ lead to the formation of Abrikosov vortices within the interstices of the APC. Although these ''interstitial FL's'' (IFL's) are relatively mobile, it has been discovered recently that they can organize into stable, ''supermatching'' flux-line lattices (SL's).[7,8]

The SL symmetry in the presence of a square lattice of antidots (circular holes) was studied in numerical simulations by Reichhardt *et al.*[9] for the case $n_s = 1$. They predict that the interstitial FL undergoes a transition from a lattice of corner-sharing squares to one of triangular coordination when the number of interstitial vortices changes from 2 to 3. This transition was directly observed in a thin, perforated Nb foil by Lorentz microscopy.[7] The existence of these SL was also inferred, and their relative stability studied, in ac magnetization measurements of an amorphous $W_{0.67}Ge_{0.33}$ film with a square lattice of antidots at various ac drive levels and angular orientations of the film plane with respect to the applied dc field.[8]

However, up to now, the existence and symmetry of the SL for higher fields has remained uncertain. Calculations[9] show that for the fifth matching field (MF) the IFL again forms a square SL, and that low-symmetry SL states are formed at the sixth and seventh MF. A transition to a triangular SL at the eighth MF was also predicted. In this paper, we show that highly ordered SL at the fifth and eighth MF are clearly visible in magnetization and transport measurements on thin Nb film with a square antidot lattice, and that formation of low-symmetry SL at the sixth and seventh MF leads to a broad maximum in $I_c(H)$ which probably reflects a smooth transition between the SL for $n = 6$ and 7.

Vortex pinning by a lattice of APC competes with a background of the usual mechanisms of pinning, including thickness modulations, nonsuperconducting inclusions, grain boundaries, surface roughness, etc. In order to clearly observe matching phenomena, pinning by the APC should be stronger than other pinning interactions. This imposes strict requirements on the quality of sample films with APC. In order to dominate the pinning, the APC must have a size comparable to the superconducting coherence length $\xi(T) = \xi_0/(1-T/T_c)^{1/2}$, which is estimated to be smaller than 100 nm for Nb at temperature relevant to this study. This small size scale restricts the techniques that can be used to fabricate such pinning centers. To date, electron-beam lithography has been used nearly exclusively to perform this task.[1–5] The main drawback of this technique is that it is expensive and too slow to pattern samples having convenient dimensions of order 0.1–1.0 mm. An effective, alternate method for producing large-area lattices of AP in superconducting thin films is based on the use of laser interferometric lithography (IL).[10–15] IL and photoresist liftoff were employed to produce a lattice of antidots of period $d = 1$ μm

 ©1999 The American Physical Society

and diameter $D \approx 0.3$ μm imbedded in a 1000 Å Nb film deposited by magnetron sputtering on a patterned Si/SiO_2 substrate. Details of the patterning and thin-film preparation can be found elsewhere.[16] The film used in the present study had a relatively low superconducting transition temperature $T_c \approx 6.7$ K due to a residual oxygen pressure in the growth chamber.[17]

Transport measurements were accomplished by attaching Au wires directly to the Nb film using a silver epoxy and postbaking. ac susceptibility and transport measurements were performed using a Quantum Design MPMS superconducting quantum interference device (SQUID) magnetometer with the applied dc magnetic field oriented perpendicular to the film plane. We concentrated our measurements on the temperature interval near T_c where the depth (Bean) profile of the magnetic field in the sample is nearly flat.[2] In this case uncertainties concerning the penetration profile of the FL are essentially eliminated, and the observed sample behavior can be understood by considering only a few ''unit cells'' of the antidot lattice.

The critical current $I_c(H)$ was defined using a voltage criterion of 0.2 μV, and the results for the field dependence of $I_c(H)$ and $\rho(H)$ at various temperature are summarized in Fig. 1. Sharp anomalies in $I_c(H)$ and $\rho(H)$ are observed at applied fields equal to multiples of the MF, $H_m = nH_1$, where $H_1 = \Phi/d^2 \approx 20.7$ G, Φ_0 is the flux quantum and n is an integer.

It was shown earlier[2,18] that the number of observed MF varies with temperature, with more anomalies visible at lower T, which was attributed to the change in the maximum number $n_s = D/4\xi(T)$ of flux quanta trapped by a single antidot due to the strong temperature dependence of $\xi(T)$ close to T_c. Therefore, changes in the collective pinning interactions between the IFL, and the IFL and the fluxoid currents, with increases in the number of interstitial vortices were not considered in detail prior to the initial discovery of SL for $n > n_s$.[7,8] In contrast, in the temperature interval $\Delta T = T_c - T = 0.3$ K emphasized in our measurement, we estimate $n_s = 1$ ($D = 0.3$ μm, $T/T_c > 0.955$, and $\xi_0 \approx 160$ Å), which means that only a single Φ_0 can be trapped at each antidot. Therefore, all matching anomalies observed at applied dc fields above H_1 must be caused by the existence of relatively stable Si with different symmetries. Thermal fluctuations have less influence on the stability of SL at lower temperature, which explains why the number of well defined MF anomalies visible in Fig. 1 changes from 2 at $T = 6.68$ K $= 0.997 T_c$ to 8 at $T = 6.55$ K $= 0.978 T_c$. Moreover, there was an absence of the usual type of MF anomaly near H_6 and H_7, which was replaced by a broad anomaly in the magnetoresistance and transport critical current data for $T \leqslant 6.6$ K.

The above observations were confirmed using more sensitive measurements of complex ac magnetic susceptibility $\chi = \chi' + i\chi''$, taking care that the ac amplitude h_0 was sufficient to attain full penetration of the ac field to the sample center. The results in Fig. 2 show a broad anomaly in $\chi(H)$ for fields spanning $H_m = 6H_1$ and $7H_1$ (which cannot be separately resolved), followed by a weak anomaly at $H = H_m = 8H_1$. Taken together, the data of Figs. 1 and 2 provide strong evidence that the anomalies at $H > H_5$ reflect the intrinsic properties of the SL phases.

Numerical simulations have pointed out that dissipation

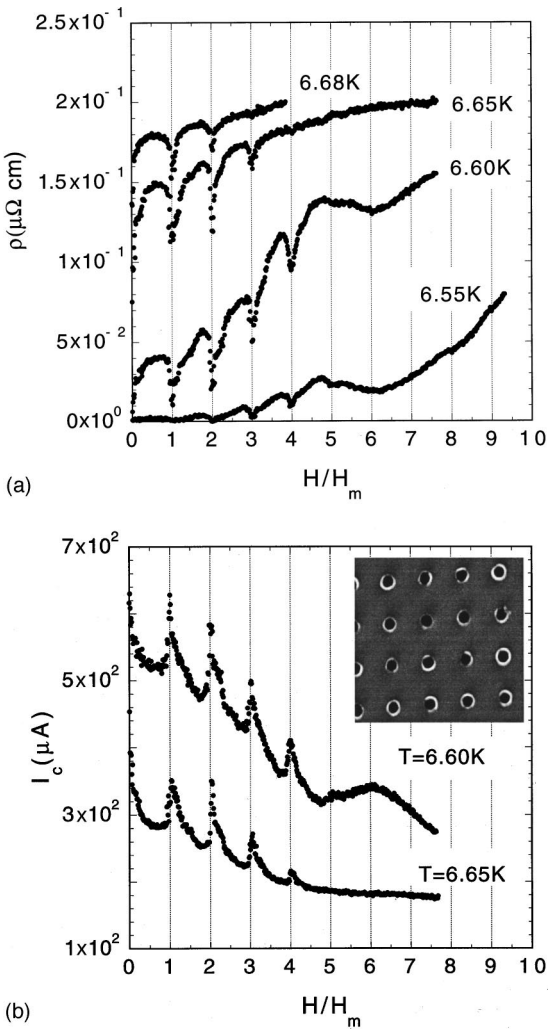


FIG. 1. (a) Magnetoresistivity (a) $\rho(H)$ vs applied dc magnetic field H for different temperatures and an applied ac current $I = 1067$ μA driven at a frequency of 23 Hz. (b) The critical current $I_c(H)$ vs H corresponding to the data of (a). Inset: SEM image of a square lattice of antidots for the sample Nb film.

by IFL is mainly due to the motion of IFL chains along the $\langle 100 \rangle$ or $\langle 010 \rangle$ directions of the antidot lattice.[9] The chart of possible IFL given in Fig. 3 indicates that for $H_5 \leqslant H \leqslant H_7$ ther are no long, straight chains of IFL which could ''slip'' along the $\langle 100 \rangle$ or $\langle 010 \rangle$ directions. This effectively places a restriction on IFL motion that would explain the suppression of dissipation which is reflected in the minima in $\rho(H)$ [Fig. 1(a)], the enhancement of the critical current $I_c(H)$ [Fig. 1(b)], and the corresponding anomalies clearly visible in $\chi(H)$ (Fig. 2) for $H_5 \leqslant H \leqslant H_7$.

The nonmonotonic dependence of $I_c(H_m)$ leads us to the conclusion that at lower-symmetry SL may yield a relatively high critical current, possibly through the suppression of mobile FL chains. The reemergence of relatively strong pinning at H_8 corresponds to the transition of the IFL to a more ideal hexagonal coordination of the IFL about the antidot sites. These data are in good qualitative agreement with numerical simulations[9] that produce matching anomalies to the second and fifth MF (corresponding to a square coordination of the IFL about the antidot sites), and the fourth and eighth MF (corresponding to roughly hexagonal coordination of the IFL

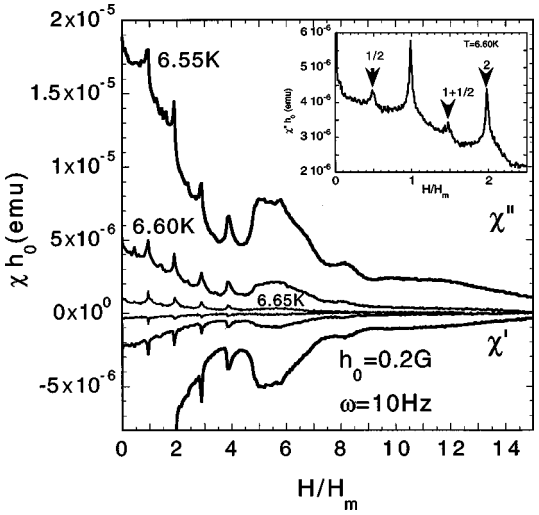


FIG. 2. Field dependences of the real (χ') and imaginary (χ'') parts of the ac magnetic susceptibility χ times the ac drive amplitude h_0 for a Nb film containing a square lattice of antidots at different temperatures T, $h_0=0.2$ G, and measuring frequency $\omega=10$ Hz. The inset: Fractional matching peaks in χ'' at $T=6.60$ K.

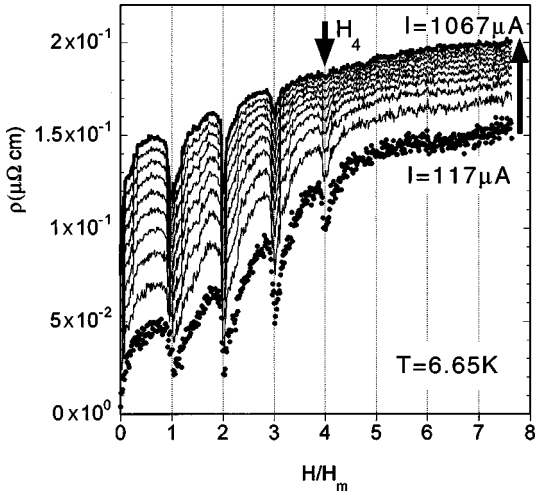


FIG. 4. Magnetoresistivity $\rho(H)$ vs applied dc magnetic field H at temperature $T=6.65$ K for different transport currents $I=117$, 223, 330, 435, 540, 645, 752, 856, 963, and 1067 μA, driven at a frequency of 23 Hz. The arrow indicates the direction of the transport current.

about the antidot site). These stable SL configurations can be realized only in the case of strong vortex-vortex interactions extending over a few intervortex separations. Indeed, it was shown that at relatively high temperatures close to T_c the effective thin-film penetration depth Λ is much higher than intervortex separation d, and one can expect collective vortex behavior.[2,3,8] The presence of long-range vortex-vortex interactions is also supported by our observation of fractional matching field ($H_{1/2}$, $H_{3/2}$, etc) for $H<H_3$ in the ac susceptibility data of Fig. 2. These MF indicate the expected existence of superstructures or lattices derived from vacancy ordering in the IFL when the vortex-vortex interactions extend beyond the antidot separation.[2,7,19]

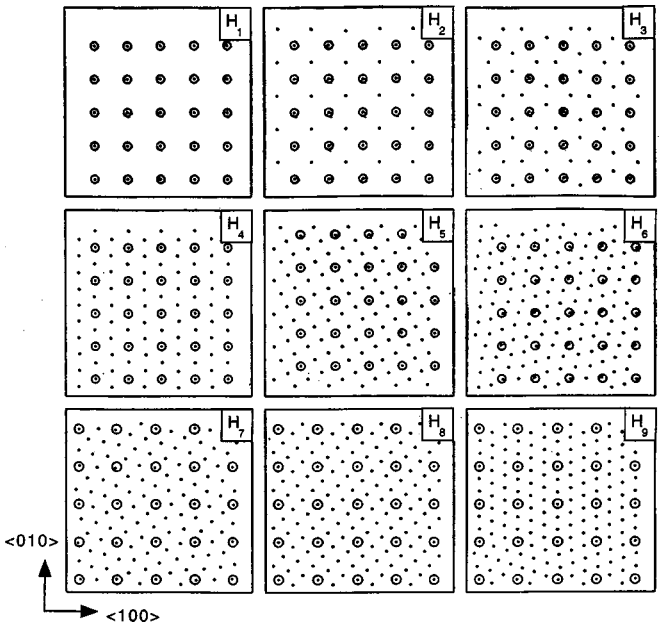


FIG. 3. Supermatching flux-line lattices SL for a square lattice of antidots, as predicted in the numerical simulations of Ref. 9.

The $\rho(H)$ data for different transport currents and $T=6.65$ K are shown in Fig. 4, which demonstrates that the number of resolved MF in the electrical transport data is also dependent upon applied current. This is again consistent with simulation results[9] (see Fig. 3), which show that IFL tend to move in chains along linear channels perpendicular to the applied current, since an increase in current along the $\langle 100\rangle$ direction induces an increase in the Lorentz force acting on the IFL along the $\langle 010\rangle$ direction. This leads one to expect relatively high dissipation near $H=H_4$, where the SL consists of nearly straight lines along the $\langle 010\rangle$ and $\langle 100\rangle$ directions, in contrast to the case for $H=H_3$, where relatively low-density chains are broken into short segments consisting of two IFL.

In conclusion, we have demonstrated that relatively stable supermatching vortex lattice phases comprised of two different type of fluxoids—those strongly trapped at APC sites and Abrikosov vortices weakly pinned within the interstices of the APC lattice—can exist in patterned superconducting thin film in fields up to $8H_m$, much higher than the first matching field. We present evidence that rather subtle changes in the SL symmetry are responsible for the unexpected loss of distinct MF anomalies observed in our measurements. Even with a relatively simple square APC lattice, the commensurate IFL are surprisingly complex and varied. Our results show that this complexity has dramatic influence on the dynamics of the IFL, inducing large changes in the depinning threshold and in the dissipation of the moving vortices. The close connection between complexity and dynamics makes superconductors with APC lattices an attractive system for studying dynamic behavior. The complexity can be controlled easily with the applied field and the symmetry of the APC, and the depinning dynamics should be sensitive not only to the matching field but also to the Lorentz force direction. One may find, for example, a periodic variation in the critical current with angle as the Lorentz force sweeps through easy and hard depinning directions. As we have shown, the matching field behavior as seen in transport and

ac susceptibility point to an unexpectedly rich spectrum of static and dynamic behavior in superconductors with periodic artificial pinning centers.

We thank F. Nori and C. Reichhardt for stimulating discussion. This work was supported by the U.S. DOE, BES-Materials Sciences, under Contract No. W-31-109-ENG-38 (U.W., V.M., and G.W.C.) by DARPA (Z.Z. and S.R.J.B.), and by U.S. DOE BES-Materials Sciences, Grant No. DE-FG02-97ER45653 (L.E.D.).

[1] A. T. Fiory, A. F. Hebard, and S. Somekh, Appl. Phys. Lett. **32**, 73 (1978).

[2] V. V. Metlushko, M. Baert, R. Jonckheere, V. V. Moshchalkov, and Y. Bruynseraede, Solid State Commun. **91**, 331 (1994); M. Baert, V. V. Metlushko, R. Jonckheere, V. V. Moshchalkov, and Y. Bruynseraede, Phys. Rev. Lett. **74**, 3269 (1995); V. V. Moshchalkov, M. Baert, V. V. Metlushko, E. Rosseel, M. J. Van Bael, K. Temst, R. Jonckheere, and Y. Bruynseraede, Phys. Rev. B **54**, 7385 (1996).

[3] J. I. Martin, M. Velez, J. Nogues, and I. K. Schuller, Phys. Rev. Lett. **79**, 1929 (1997).

[4] D. J. Morgan and J. B. Ketterson, Phys. Rev. Lett. **80**, 3614 (1998).

[5] A. Bezryadin and B. Pannetier, J. Low Temp. Phys. **102**, 73 (1996); A. Bezryadin, Yu. N. Ovchinnikov, and B. Pannetier, Phys. Rev. B **53**, 8553 (1996).

[6] G. S. Mkrtchyan and V. V. Shmidt, Zh. Eksp. Teor. Fiz **61**, 367 (1971) [Sov. Phys. JETP **34**, 195 (1972)].

[7] K. Harada, O. Kamimura, H. Kasai, T. Matsuda, A. Tonomura, and V. V. Moshchalkov, Science **274**, 1167 (1996).

[8] V. V. Metlushko, L. E. DeLong, M. Baert, E. Rosseel, M. J. Van Bael, K. Temst, V. V. Moshchalkov, and Y. Bruynseraede, Europhys. Lett. **41**, 333 (1998).

[9] C. Reichhardt, C. J. Olson, and F. Nori, Phys. Rev. B **57**, 7937 (1998).

[10] H. Anderson, H. I. Smith, and M. L. Schattenberg, Appl. Phys. Lett. **43**, 874 (1983).

[11] S. H. Zaidi and S. R. J. Brueck, J. Vac. Sci. Technol. B **11**, 658 (1993); S. H. Zaidi, A.-S. Chu, and S. R. J. Brueck, J. Appl. Phys. **80**, 6997 (1996); X. Chen, S. H. Zaidi, S. R. J. Brueck, and D. J. Devine, J. Vac. Sci. Technol. B **14**, 3339 (1996).

[12] T. A. Savas, S. N. Shah, M. L. Schattenburg, J. M. Carter, and H. I. Smith, J. Vac. Sci. Technol. B **13**, 2732 (1996); A. Fernandez, J. Y. Decker, S. M. Herman, D. W. Phillion, D. W. Sweeney, and M. D. Perry, *ibid.* **15**, 2439 (1997).

[13] X. Chen, Z. Zhang, S. R. J. Brueck, R. A. Carpio, and J. S. Petersen, in *Emerging Lithographic Technologies*, edited by David E. Seeger, SPIE Proceedings Vol. 3048 (SPIE, Bellingham, WA, 1997), pp. 309–318.

[14] C. O. Bozler, C. T. Harris, S. Rabe, D. D. Rathman, M. A. Hollis, and H. I. Smith, J. Vac. Sci. Technol. B **12**, 629 (1994).

[15] A. Fernandez, H. T. Nguyen, J. A. Britten, R. D. Boyd, M. D. Perry, D. R. Kania, and A. M. Hawryluk, J. Vac. Sci. Technol. B **15**, 729 (1997).

[16] V. Metlushko, U. Welp, G. W. Crabtree, Zhao Zhang, S. R. J. Brueck, B. Watkins, L. E. DeLong, B. Ilic, K. Chung, and P. J. Hesketh, Phys. Rev. B **59**, 603 (1999).

[17] C. C. Koch, J. O. Scarbrough, and D. M. Kroeger, Phys. Rev. B **9**, 888 (1974); W. De Sorbo, Phys. Rev. **132**, 107 (1963).

[18] V. V. Moshchalkov, M. Baert, V. V. Metlushko, E. Rosseel, M. J. Van Bael, K. Temst, Y. Bruynseraede, and R. Jonckheere, Phys. Rev. B **57**, 3615 (1998).

[19] Y. Bruynseraede, G. Verbanck, M. J. Van Bael, K. Temst, and V. V. Moshchalkov, in *Superconducting Superlattices II: Native and Artificial*, edited by I. Bozovic and D. Paruna, SPIE Proceedings Vol. 3480 (SPIE, Bellingham, WA, 1998), pp. 90–100.

Fabrication of an integrated nanofluidic chip using interferometric lithography

M. J. O'Brien II
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

P. Bisong and L. K. Ista
Department of Chemical and Nuclear Engineering, University of New Mexico, Albuquerque, New Mexico 87131

E. M. Rabinovich
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

A. L. Garcia
Department of Chemical and Nuclear Engineering, University of New Mexico, Albuquerque, New Mexico 87131

S. S. Sibbett
Intel Corporation, Rio Rancho, New Mexico 87124

G. P. Lopez[a)]
Department of Chemical and Nuclear Engineering, University of New Mexico, Albuquerque, New Mexico 87131

S. R. J. Brueck[b)]
Center for High Technology Materials, University of New Mexico, Albuquerque, New Mexico 87106

(Received 25 June 2003; accepted 21 September 2003; published 9 December 2003)

The fabrication of nanoscale structures with dimensions approaching the scale of biological molecules offers approaches to the study of fluid dynamics and biomolecular transport. Ultimately, a parallel lithographic approach will be necessary if devices based on these nanofluidics are to achieve widespread availability and acceptance. We report on a flexible, all-optical lithography alternative that is amenable to large-scale production. We use interferometric lithography (IL) and anisotropic etching to produce large areas of parallel, nanofluidic channels with widths of ~100 nm and depths of up to 500 nm. We also use standard optical lithography to create interfacing microchannels, such that the range of spatial scales on one chip varies by 10^4 (from mm scale reservoirs to 100 nm nanochannels). We provide initial demonstrations of capillary action and electrophoretic motion of fluorescent dye solutions. © *2003 American Vacuum Society.*
[DOI: 10.1116/1.1625964]

I. INTRODUCTION

The study of molecular transport phenomena in fluidic channels of nanoscopic dimensions is a current frontier in experimental and theoretical fluid dynamics.[1,2] Nanofluidic systems have a variety of applications including molecular separations, manipulation and detection of individual biomolecules, and sensors.[3–6] The lack of convenient and readily available experimental systems has hindered the validation of theoretical and simulation studies that have predicted unique transport properties[7] and molecular dynamics[1,8] in such systems. Fabrication techniques of such systems need to be amenable to high throughput production, allow nanoscale patterning over large surface areas, facilitate integration of nanofluidic components to the micro- and macroscale components, permit flexibility in design, and employ materials that are compatible with biomolecules.

[a)]Electronic mail: gplopez@unm.edu
[b)]Author to whom correspondence should be addressed; electronic mail: brueck@chtm.unm.edu

A. Previous approaches

Electron beam (e-beam) lithography has been most widely used in producing nanofluidic systems with well defined and controllable feature sizes. e-beam lithography allows for complex patterns and exquisite resolution.[9,10] However, high-throughput fabrication is not practical with e-beam lithography due to its slow, serial processing nature that is not well suited for producing large areas of nanostructures.

Recently, nanoimprint lithography[5,11,12] has been investigated as an alternative approach. This technique requires a mold to be formed first, usually with e-beam lithography for nanoscaled features[12,13] and optical lithography for larger features, such as interfacing gradient structures.[12] The pattern is transferred to a thermoplastic polymer (e.g., polymethylmethacrylate) through heat and pressure[5,11,12] or into an ultraviolet (UV)-polymerizing liquid.[14] Nanoimprint lithography approaches are parallel, fast, and well suited for creating nanotextured patterns over macroscopic areas. However, a new mold must be created whenever a feature characteristic needs to be changed (e.g., pitch, channel size, gradient scale,

etc.). Additionally, nanoimprint techniques have some difficulty in accommodating wide ranges of feature size into a single pattern. The issue of stamp lifetimes has also not yet been fully explored and may pose a challenge to high-throughput fabrication.[15]

B. All-optical lithographic approach

We have based our fabrication efforts on optical lithography, which is well developed, reliable, and flexible. Interferometric lithography (IL), a maskless technique based on the interference of two or more coherent beams, allows one to inexpensively and quickly pattern nanoscopic features over large surface areas with easily varied feature dimensions (e.g., pitch size and channel width). It is well suited to high-throughput manufacturing.[16]

While traditional optical lithography has been used by others for the fabrication of step-shaped nanofluidic devices on silicon, these structures are nanoscopic in vertical dimension only (controlled by deposition or etching) with macro/microscopic transverse dimensions (controlled by lithographic pattern formation).[4] With IL two coherent light beams of wavelength λ are crossed at an angle 2θ, producing a regular interference pattern with $d=\lambda/(2\times\sin\theta)$ describing the period.[16] With an UV light source, one can easily obtain periods on the order of hundreds of nm and transverse pattern features in the sub-100 nm range, well beyond the scales available from traditional optical lithography approaches. Developments such as deeper ultraviolet sources and immersion techniques promise to extend these scales to sub-100 nm periods and $\sim$10 nm channel widths.[17]

With this approach, nanochannels can be etched into silicon, rather than being pressed into plastic. Silicon can be easily oxidized after etching, providing an inert, electrically insulating, and hydrophilic surface that can be chemically functionalized with silane chemistry. A variety of etching processes and wafer-bonding techniques are available for silicon. This is important because most nanofluidic devices require one to seal the tops of nanotextured surfaces to form nanoscopic tunnels rather than trenches.[4,5,10,12]

IL may be combined with traditional optical lithography to yield a wide range of characteristic sizes (mm to nm) on a single device. In this article, we describe the fabrication of two variations of nanofluidic devices. The first is a large-area nanochannel array chip consisting essentially of an oxidized nanoscale Si grating with an anodically bonded Pyrex roof with holes in it for convenient introduction of fluids. The second device contains a limited area of nanoscale features, integrated microchannels, and macroscopic reservoirs, using a convenient cross configuration that interfaces the nanofluidics to the macroscopic world and provides control mechanisms for fluid flow (others have also used microchannels as an interface but used simpler, linear configurations).[10] Such cross configurations have been successfully and widely used in study and manipulation in microfluidics. For example, the formation of a compositional plug or band in microchannels is facilitated by the cross configuration.[18]

II. FABRICATION

Silicon $\langle 100\rangle$ wafers were cleaved into $3\times 4\ \mathrm{cm}^2$ chips that were cleaned in piranha solution (1 part H_2O_2, 2 parts H_2SO_4 by volume), triple rinsed with de-ionized (DI) water, dipped in HF acid (to remove the native oxide layer and any remaining inorganic contaminants), and again triple rinsed with DI water. A 150 nm thick layer of XHRiC-16 (Brewer Science, Inc.) antireflective coating (ARC) was spin deposited (4000 rpm, 30 s) and hard baked at 175 °C for 3 min. This was followed by spin deposition (4000 rpm, 30 s) of a 200 nm layer of positive photoresist [SPR510a photoresist diluted by an equal amount of EC-11 solvent (Shipley, Inc.)] and soft baking at 95 °C for 3 min. The exposure source was the frequency-tripled ($\lambda=355$ nm) output of a YAG–Nd laser (Infinity 40–100, Coherent Inc.). The laser beam is expanded and illuminates a right-angle assembly containing a mirror and a vacuum chuck to hold the Si sample. Although we have focused on pitches of 500 nm ($\theta\sim 26°$), the reflector assembly can be rotated to produce a variety of grating pitches.

After exposure, each sample was soft baked at 110 °C for 1 min, developed using undiluted MF702 developer (Shipley, Inc.), and rinsed with water, leaving a photoresist grating. The developed chip was placed in an e-beam evaporator where a thin (35–40 nm) layer of Cr was deposited. We then lifted off the remaining photoresist (and the Cr on top of it) using an airbrush acetone spray, leaving a negative-tone Cr etch mask layer on top of the remaining ARC (which is impervious to the acetone). A field-emission scanning-electron microscope (FE–SEM) image of one such mask is shown in Fig. 1(A).

Additional processing steps were used to prepare the integrated microfluidics on the more complex samples before etching. To form the 200 μm wide microchannel interfaces and the 2 mm diameter reservoirs, we first spun (3000 rpm for 30 s) a 1.44 μm thick layer of AZ 5214-E (Shipley, Inc.) resist onto the samples with Cr etch masks. The chips were then exposed with a rectangular exposure mask, using a conventional proximity aligner, developed, and baked (95 °C for 10 min) to produce a protective layer over the area where we wanted to retain the nanochannels, as shown in Fig. 1(B). We placed the chips in CEP-200 Cr etchant (Microchrome Technology) to remove the unprotected regions of the Cr mask, then removed the protective photoresist layer with acetone leaving a small area ($1\times 10\ \mathrm{mm}^2$) on the chip with a Cr etch mask [Fig. 1(C)]. A 1.44 μm thick layer of AZ 5214-IR (Shipley, Inc.) resist was spun over the chips, then exposed with both the microchannel exposure mask [Fig. 1(D)] and a second rectangular exposure mask to remove the resist from the nanochannel area [Fig. 1(E)]. This results in a chip with an etch mask made of photoresist and a Cr grating as represented in Fig. 1(F).

The samples were reactive ion etched (RIE) using a mixture of O_2 and CHF_3. The etched silicon gratings were cleaned with piranha solution to remove the ARC, Cr, and residual polymer from the RIE process [Fig. 2(A)]. After cleaning, we placed the chips in a quartz tube furnace con-

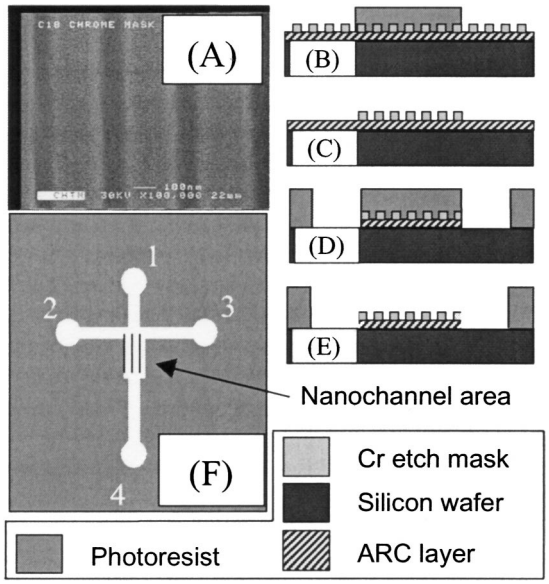


FIG. 1. Mask fabrication steps: (A) FE–SEM image of a Cr etch mask after lift-off; (B) photoresist is spun over the Cr grating and exposed (using an intensity mask) and developed, leaving a small area of the Cr etch mask protected by a layer of resist; (C) after Cr etch and acetone rinse a smaller area of Cr grating remains; (D) a second resist coat is spun, exposed with microchannel mask, and developed; (E) the chip is ready for etching after an additional exposure and develop process to clear photoresist away from the Cr etch mask; and (F) a top view of the final etch mask (consisting of a chrome etch mask for the nanochannel area and a photoresist etch mask) for an integrated chip with reservoir areas numbered for reference.

taining ultrahigh-purity grade O_2 at 1100 °C for 45–60 min to form an insulating oxide layer [Fig. 2(B)].

After etching and oxidation, all chips were capped with 1 mm thick Pyrex No. 7740 roofs using anodic bonding. These glass plates had a surface quality of 1.8 λ [measured across a 2.54 cm diameter circle with a laser interferometer (Zygo, Inc.) at 633 nm]. The roofs for the integrated chips were predrilled with four holes (~2 mm diameter), located above the circular reservoirs at the ends of the microchannels. The roofs for the large-area nanochannel chips had four holes (3 mm diameter) drilled in a row on one edge to provide ports for convenient loading of fluorophores and/or other solutions.

The bottom electrode of the bonding apparatus was the grounded metal surface of a hot plate that supported the oxidized silicon sample. The Pyrex roof was placed on top of the oxidized silicon chip. The upper electrode was a small aluminum block, which was placed on top of the Pyrex roof and connected to a high voltage dc power supply. We raised the temperature of the hot plate to 380 °C before charging the capacitor structure. We would typically increase the capacitor voltage slowly until the current reached a value of 1–2 mA, then allow the current to decay to ~0.5 mA (indicating the formation of a space charge blocking layer) before further increasing the voltage. We used as large a voltage as possible while avoiding arcing between the electrodes (typically 800–1000 V). We monitored the current flow through the system until it decayed to ~100 μA per chip before shutting off the heating element and allowing the sample to cool. The voltage was decreased to zero when the temperature dropped below 150 °C. A FE-SEM image of a bonded chip cross section is shown in Fig. 2(C).

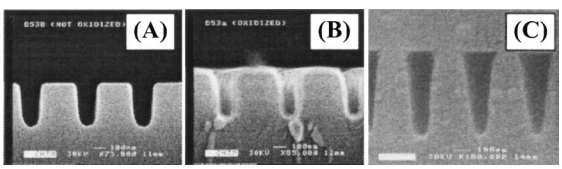


FIG. 2. FE–SEM images of etched samples: (A) silicon wafer after etching, (B) the same wafer after oxidation, and (C) an oxidized grating with a bonded Pyrex roof. All samples have been sputtered with gold prior to imaging to reduce charging effects. The scale bars indicate distances of 100 nm.

III. RESULTS

We prepared a solution of standard Tris/glycine electrophoresis buffer (0.24 mM Tris and 1.92 mM glycine, *p*H 8.8). This buffer was filtered through a 0.2 μm filter to remove particulate contaminants and then degassed under vacuum for at least 3 h to reduce outgassing during electrokinetic motion. A 5 mg/ml suspension of Alexa Fluor 532 C_5 maleimide (Molecular Probes) was prepared in 1.5 mM Tris HCl buffer (*p*H 8.8).

A. Fluid flow in the large-area nanochannel array chip

We mounted the large-area nanochannel array design into a Teflon chuck with reservoirs at each end of the chip. Platinum wire-mesh electrodes were inserted into the sides of both reservoirs (~3 cm from the chip edge). Poly(dimethylsiloxane) was used to secure the chip into the chuck and seal the system both for fluid flow and for electrical isolation. The assembly was imaged using an upright, laser scanning (543 nm) confocal microscope (Axioskop using an LSM5 scanning head, Zeiss, Inc.). The fluorescence output passed through a long-wavelength pass optical filter (560 nm cutoff). ~2 μl of Tris/glycine solution was added to each hole. After capillary action caused the buffer to move ~5 mm through the nanochannels, we added a few μl of Alexa 532 solution to one of the holes in the glass roof and imaged the flow of liquid due to capillary action. Figures 3(A) and 3(B) show the progression of the dye solution over 30 s. The average fluid velocity (average of several measured velocities at different points on the liquid front) was 12.2 $\pm$0.6 μm/s.

After the entire chip was filled via capillary action with Tris/glycine buffer, we filled the reservoirs with buffer. We placed 1 μl of Alexa 532 dye into one of the four holes on the top of the chip and biased the electrodes with 50 V. Figures 3(C) and 3(D) shows the progression (due to electrophoresis) of the negatively charged Alexa 532 dye towards the positive electrode over a period of 150 s. The average electrophoretic velocity (average of several measured velocities at different points on the liquid front) was 0.77 $\pm$0.03 μm/s.

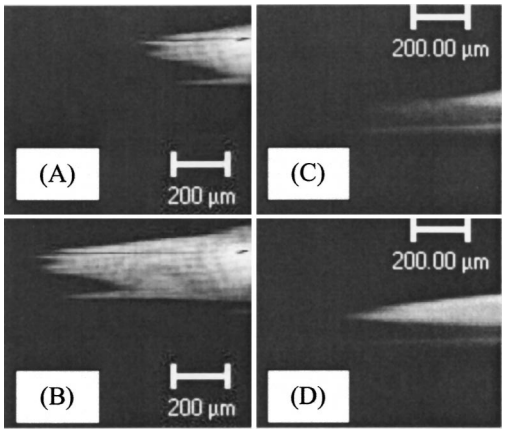


FIG. 3. Confocal images of fluid motion in the large-area grating design: the scale bars indicate distances of 200 μm, the reservoir to the left of the picture is biased at +50 V, and the reservoir to the right of the picture is grounded. (A) The bright area is Alexa 532 while the dark area corresponds to air-filled nanochannels. The ring like interference patterns are due to some surface damage in the glass roof from the bonding process. (B) 30 s later, capillary action has caused the Alexa solution to flow down the channels. (C) At a different location on the chip after the chip has been filled with buffer by capillary action: The bright area is Alexa 532 while the dark area is buffer in the nanochannels. (D) 150 s later, the negatively charged Alexa dye has clearly moved towards the positive electrode, indicating motion dominated by electrophoresis.

B. Electrokinetic flow through integrated chips

We loaded one of the integrated chips with liquid via capillary action. First, 50 μl of DI water was loaded into reservoir 4 [as numbered in Fig. 1(F)]. We monitored the flow of water through the microchannel and nanochannel areas by eye until the water had reached the end of the nanochannels. At that point, 50 μl of DI water was introduced into reservoir 3, filling the remaining three microchannels. After allowing the system to equilibrate for 30 min, 50 μl of Tris/glycine buffer was added to reservoirs 1 and 2 and reservoirs 3 and 4 were topped off with buffer. We inserted platinum wire electrodes into all four reservoirs and imaged the assembly with the confocal microscope. The reservoir 2 electrode was grounded. Three independent power supplies (common ground) were connected to the remaining three electrodes.

We introduced Alexa 532 dye into reservoir 3 and biased its electrode at −100 V. The remaining electrodes were at ground potential. The negatively charged dye moved towards the center of the chip, entering both the nanochannel area and the microchannel connected to reservoir 1, as seen in Fig. 4(A). Figure 4(B) shows the juncture 60 s after Fig. 4(A) was recorded. Since the dye entered the nanochannels closest to reservoir 3 first, it has progressed the farthest in those (the lowest) channels.

We then changed the bias for reservoirs 1 and 3 to −100 V while reservoir 2 remained at 0 V. This flushed the dye out of microchannel 1 and pinched off a slug of dye in the nanochannels across from microchannel 1, as seen in Fig. 4(C). Figure 4(D) shows the clear progression of the slug down the nanochannel area after 20 s. We calculate the velocity of this slug to be 26.5 ± 0.9 μm/s.

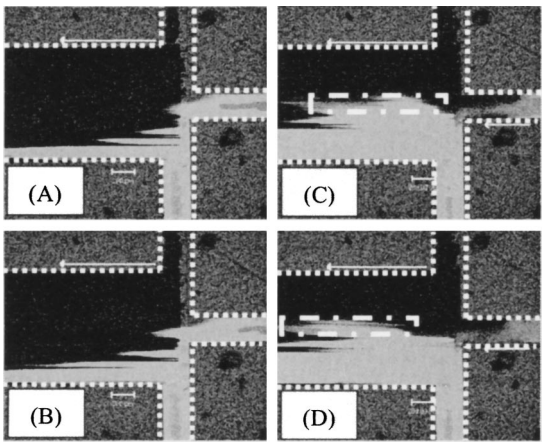


FIG. 4. Confocal microscopy of electrophoresis in integrated chip: the channels are outlined with dotted lines for clarity, the left hand side of each picture contains the nanochannel area, and microchannels are on the upper, lower, and right hand sides of each picture. The scale bars indicate 200 μm distances. (A) The electrode for the lower channel is biased at −100 V while the other three remain grounded. Electrophoresis causes the dye to flow through the lower microchannel towards the top part of the picture. It begins to flow down the nanochannels and the microchannel on the right as it reaches them. (B) 60 s later dye has progressed down the nanochannels. (C) This picture was taken after the electrode for the channel on the right hand side of the picture is biased at −100 V (along with the lower channel). The dye has been flushed out of the right hand side microchannel and pinched off a slug of dye (outlined with broken lines) in the nanochannels directly opposite the right hand side microchannel. (D) 20 s later the slug has progressed down the nanochannel area.

C. Discussion

We have demonstrated that our designs are suitable for experiments studying electrokinetic motion in nanoscale channels. It is interesting to note that we were able to achieve much higher electrophoretic velocities in the integrated chip. This is due, in part, to the difference in nanochannel lengths: The length of the nanochannel area of the integrated chip is about 1/3 the length of the sealed nanochannels in the large-area chip. The microchannels are less than 2/3 of the length of the sealed channels, and have a lower resistance (due to the larger cross-sectional area), resulting in a larger electric field across the nanochannels for a given electrode bias voltage. This, coupled with a twofold higher electrode bias, could account for a factor of 6 difference, less than the factor of 32 observed. Another contributing factor may the difference in buffer concentrations between the two solutions. It is also interesting to note that the fluid velocities due to capillary action in the large-area chip are also quite fast (approximately half that of the electrophoretic velocities in the integrated chips). Further investigation is necessary to fully characterize the fluid-flow characteristics in these nanochannel samples.

Application of 100 V to the large-area nanochannel array chips quickly resulted in a drop of current across the chip and an eventual cessation of electrokinetic motion, which is caused by outgassing of the solution. Indeed, rapid generation of microscopic bubbles could be seen at the edge of the

glass on these chips. These effects were not seen in the integrated chips, possibly because of the much better defined macro–micro–nano interface hierarchy.

IV. CONCLUSIONS

We have created functional nanofluidic chips based on all-optical lithographic processes with feature sizes ranging from <100 nm to 2 mm on a single chip. IL is the basis of the nanochannel fabrication, and allows for flexible nanopatterning of silicon chips over large surface areas. Traditional optical lithography provides convenient microfluidic structures for interfacing with the nanochannels and controlling fluid flow as we have demonstrated with cross-type microfluidics. Our techniques are suited to high-throughput manufacturing, provide flexible nanotexturing over large areas, and can produce a broad range of feature sizes on a single chip. They allow one to use inert and hydrophilic nanotextured surfaces (oxidized silicon with glass roofs) that are compatible with electrokinetic studies.

Future fabrication work will involve the addition of integrated microelectrodes to the chips, smaller nanochannel pitches, two-dimensional patterning of the nanochannels (e.g., introducing gradients in channel widths), and adding additional switching capabilities to form more complex micro/nanofluidic arrangements. Future experimental work will include detailed parametric studies of electrokinetic motion and flow rates of biomolecular species, and investigations of separations of biomolecular species within the nanochannel devices.

ACKNOWLEDGMENTS

The authors would like to thank David Burckel and Andrew Frauenglass for advice and help with IL, and Beth Fuchs for fabrication advice across a broad variety of topics. Support for this work was provided by the ARO/MURI in Deep Subwavelength Optical Nanolithography, the Mission Research Corporation in a subcontract from the Air Force Research Laboratory, Sandia National Laboratories, the National Science Foundation (Grant No. DGE 00114319, CTS 0304237), and by Intel Corporation. Confocal images in this article were generated in the Fluorescence Microscopy Facility at the UNM Cancer Center.

[1]R. Qiao and N. R. Aluru, Model. Simul. Microsystems **2002**, 28 (2002).
[2]A. Acrivos, B. Khusid, J. Koplik, and G. Drazer, Model. Simul. Microsystems **2002**, 66 (2002).
[3]B. B. Haab and R. A. Mathies, Anal. Chem. **71**, 5137 (1999).
[4]J. Han and H. C. Craighead, Anal. Chem. **74**, 394 (2002).
[5]H. Cao, Z. Yu, J. Wang, J. O. Tegenfeldt, R. H. Austin, E. Chen, W. Wu, and S. Y. Chou, Appl. Phys. Lett. **81**, 174 (2002).
[6]H. G. Craighead, Science **290**, 1532 (2000).
[7]C. L. Rice and R. Whitehead, J. Phys. Chem. **69**, 4017 (1965).
[8]X. J. Fan, N. Phan-Thien, N. T. Yong, and X. Diao, Phys. Fluids **14**, 1146 (2002).
[9]C. K. Harnett, G. W. Coates, and H. G. Craighead, J. Vac. Sci. Technol. B **19**, 2842 (2001).
[10]A. Hibara, T. Saito, H.-B. Kim, M. Tokeshi, T. Ooi, M. Nakao, and T. Kitamori, Anal. Chem. **74**, 6170 (2002).
[11]H. Cao, J. O. Tegenfeldt, R. H. Austin, and S. Y. Chou, Appl. Phys. Lett. **81**, 3058 (2002).
[12]V. Studer, A. Pepin, and Y. Chen, Appl. Phys. Lett. **80**, 3614 (2002).
[13]S. Y. Chou, P. R. Krauss, and P. J. Renstrom, Appl. Phys. Lett. **67**, 3114 (1995).
[14]M. Colburn, A. Grot, B. J. Choi, M. Amistoso, T. Bailey, S. V. Sreenivasan, J. G. Ekerdt, and C. G. Willson, J. Vac. Sci. Technol. B **19**, 2162 (2001).
[15]S. Zankovych, T. Hoffmann, J. Seekamp, J.-U. Bruch, and C. M. S. Torres, Nanotechnology **12**, 91 (2001).
[16]S. H. Zaidi, A.-S. Chy, and S. R. J. Brueck, J. Appl. Phys. **80**, 6697 (1996).
[17]A. K. Raub and S. R. J. Brueck, Proc. SPIE **5040**, 667 (2003).
[18]S. C. Jacobson and J. M. Ramsey, in *High Performance Capillary Electrophoresis*, edited by M. G. Khaledi, Chemical Analysis Series Vol. 146 (Wiley, New York, 1998), p. 613.

NANO LETTERS

2004
Vol. 4, No. 7
1295−1299

A Facile Approach to Directed Assembly of Patterns of Nanoparticles Using Interference Lithography and Spin Coating

Deying Xia and S. R. J. Brueck*

Center for High Technology Materials, University of New Mexico, 1313 Goddard, SE, Albuquerque, New Mexico 87106

Received April 30, 2004; Revised Manuscript Received May 27, 2004

ABSTRACT

We present a facile approach to the directed assembly of silica nanoparticles (≲ 100-nm diameter) into periodic arrays on flat surfaces using interference lithography and spin coating. Periodic photoresist patterns, used as templates for the spin-coating process, were prepared using interference lithography. Silica nanoparticle dispersions were spin coated on these patterned surfaces, and the resist was removed, leaving periodic nanoparticle patterns on flat surfaces. Parallel arrays and continuous cross networks of particles were fabricated successfully with periods between hundreds of nanometers and several microns over areas of several square centimeters. The morphology (period, layer width and thickness, etc.) of the particle patterns can be controlled by varying the process conditions.

The fabrication of patterned arrays of nanoparticles exhibiting various properties on surfaces and having a desired morphology is an essential step toward the creation of novel nanoscale materials and devices. The assembly and fabrication of one-, or two-dimensional (1D/2D) micro/nanoscale colloidal structures are very promising for this purpose. These micro/nanoscale colloidal arrays may find applications in photonic materials,[1,2] optoelectronic devices,[3,4] biochemical sensors,[5] template processing (nanosphere lithography,[6,7] colloidal stamps,[8,9−10] ordered porous materials[11]), etc. Development of a versatile manufacturing technology for producing these ordered arrays is a critical requirement for the success of these applications. Several approaches have been developed to precisely position nanoparticles in patterned arrays, including sedimentation,[12] electrostatic-induced crystallization,[13,14−15] convective self-assembly,[16,17] and physical confinement.[18,19] Most of these studies have focused on the creation of patterned particle arrays with ordered structures onto well-defined confined geometries[18,19] or chemically modified surfaces.[14,15] For most of these reports, the patterned array periodicity has been at length scales on the order of micrometers or larger, and the particles used for patterned arrays have been larger than ∼100-nm diameter[11,14,18] (up to micrometers[20]), although some experiments using ≲ 20-nm diameter particles have been reported.[21,22] Achieving one- and two-dimensional (1D/2D) patterned arrays with 10- to 100-nm diameter nanoparticles having controlled morphologies on a flat substrate remains an important challenge that is receiving a great deal of attention for establishing nanofabrication techniques.

Even though electron-beam lithography enables fabrication of small structures on micro/nanoscale ranges, its throughput is too low for the manufacturing[3] of large-area nanoscale patterns and it requires sophisticated and expensive instrumentation.[23] Interferometric lithography (IL) is a very promising, low-cost, reliable, and scalable technology for the fabrication of nanoscale periodic patterns over large areas.[24] For IL, two (or a few) coherent UV beams are used to produce a periodic interference pattern on a photoresist (PR) film with the spatial period $P = \lambda/2\sin\theta$ where λ is the wavelength and θ is the half angle between the two beams. With the addition of immersion techniques,[25] the limits of interferometric lithography extend to periods of $\lambda/2n$, where n is the immersion liquid refractive index.

In this work, we used IL to make large areas of periodic photoresist structures on a Si wafer for use as templates for patterning spin-coated silica nanoparticles. Following particle deposition, the PR was removed by heating or by chemical treatment, leaving ordered silica particle arrays on a flat substrate.

Immersion methods for patterning colloid particles are relatively slow (∼hours) and require a large volume of colloidal solution. Uniform filling of open, continuous channels or networks with colloidal crystals can prove difficult because of channel plugging problems during capillary-driven infiltration of suspensions.[26] Spin coating is widely employed in semiconductor manufacturing, is simple and convenient, uses only a small volume of fluid, and is rapid (∼1 min)

* Corresponding author. E-mail: brueck@chtm.unm.edu.

10.1021/nl049355x CCC: $27.50 © 2004 American Chemical Society
Published on Web 06/18/2004

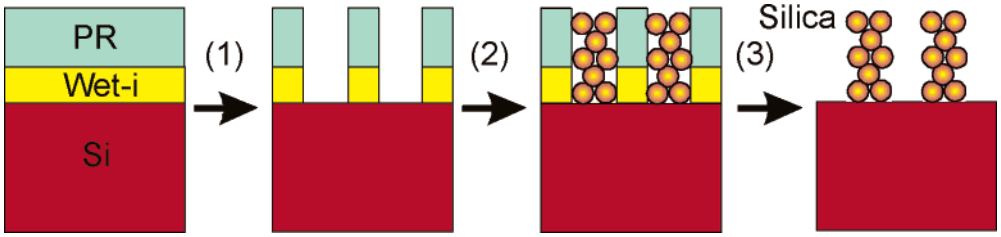


Figure 1. Schematic illustration of deposition of patterned silica particles on planar surface using developing-soluble ARC (wet-i) layer: (1) expose using interference lithography and develop; (2) deposit silica particle using spin coating; (3) strip off PR/wet-i.

and highly reproducible. Spin-coating assembly allows the formation of ordered patterns of nanoparticles with controlled thicknesses and transverse patterns within confined PR templates on an otherwise flat substrate.

Figure 1 illustrates the scheme for directed assembly of silica nanoparticles into patterned arrays on flat surfaces using interference lithography and spin coating. A bottom antireflecting coating (BARC) layer (wet-i, wet developable BARC for I-line lithography) was spun on the precleaned sample to minimize reflection of the IL beams from the substrate, eliminating vertical standing wave patterns that otherwise cause difficulties with the PR exposure and development processes. A positive-tone photoresist (Shipley SPR510A) diluted to produce a ~200-nm thick film was used in these experiments. IL with a 355-nm tripled YAG laser source was used to produce the periodic patterns on this PR/wet-I film stack. The angle between the two beams determined the period of the pattern, while the exposure fluence and development parameters controlled the final pattern line widths. Parallel PR/wet-i lines (1D) were formed on the Si wafer surface after exposing, baking, and developing. For 2D arrays of nanoscale posts, a double exposure with a 90° rotation of the sample was performed, and all other steps were the same as those for 1D patterns.

After the PR patterning, silica nanoparticle colloidal dispersions were spin coated onto these samples. In general, the spin coating was 30 s at 4000 rpm with acceleration 100 rpm/s^2. Suspensions of two size distributions of silica nanoparticles were utilized for directed assembly of 1D and 2D periodic particle structures supported on flat surfaces. The suspensions were Snowtex colloid silica from Nissan Chemical Industries. Ltd. The concentrations of both large silica particles (ZL, 78 ± 16 nm diameter) and small silica particles (OL, 50 ± 13 nm diameter) were 5 wt %. The suspension was agitated for 5 min in an ultrasonic bath before spin coating. There are two reasons for utilizing silica particles instead of polystyrene beads: (1) the silica particles are likely to match the Si wafer surface and to be compatible with semiconductor processing; and (2) the silica particles are readily applicable in spin coating due to their relatively high mass density (2.13 gm cm^{-3}) compared to polystyrene (1.05 gm cm^{-3}), which is close to the mass density of the water-based solvent.

The final step was to strip the PR/wet-i patterns, leaving 1D or 2D ordered patterns of silica nanoparticles on the flat surface of the Si wafer. Two methods, immersion in piranha solution and calcination at high temperature, were successful in completely removing the PR/wet-i while not disturbing

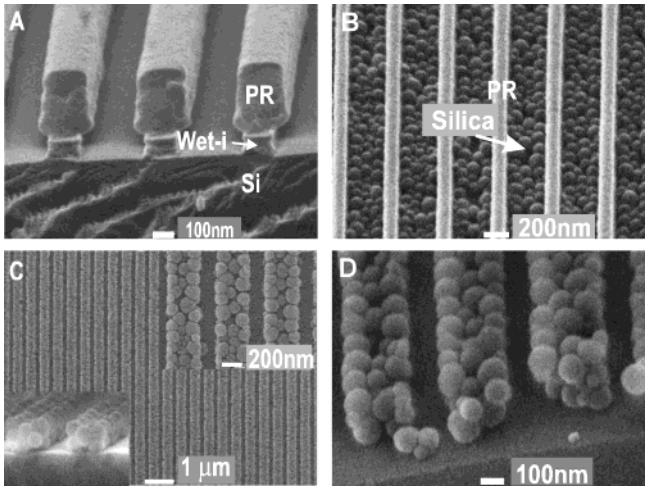


Figure 2. FE-SEM images of 1D patterns with 500 nm period. (A) PR/wet-i patterns at a cleaved edge. (B) ZL (~78 nm diameter) silica particles confined between PR/wet-i walls. (C) Top-view of typical region over a large area after removal of the PR template. Top right: magnified top view. Bottom left: cross-sectional views. (D) Tilted 45° view at a cleaved edge.

the silica particle arrays. Low volumetric ratio (1:1, 1:1.5, 1:2) piranha compositions (30% H_2O_2 : 98% H_2SO_4) worked well in removing the organic material while retaining the silica particle patterns. Higher volumetric ratios (e.g., 1:4) of piranha had violent reactions with the PR/wet-i, producing many bubbles that collapsed the silica particle patterns to some degree. In general, we soaked the as-patterned samples in 1:1 piranha solution for 15 min. Calcinations were then carried out at 800 °C for 1.5−2 h. The mechanical stability of particle patterns on flat surfaces was enhanced as a result of partial sintering between the silica particles and the Si surface at this temperature.[27] The resulting structures were characterized by field-emission scanning electron microscopy (FE-SEM) at 30 kV with a 10-nm thick Au film sputtered on the sample surface. Some top views and cross-section views of images were also obtained to check the structure of the ordered patterns.

Figure 2A shows an FE-SEM image of typical PR/wet-i lined patterns with ~500 nm period on a Si surface. Standing wave patterns on the PR sidewalls were minimized after use of the wet-i layer. The post-develop width of the wet-i layer (70−80 nm thick) was controlled by preexposure bake and development parameters; it is somewhat over-developed in Figure 2A. Figure 2B shows another example of FE-SEM images for ordered 1D patterns after spin coating with suspension of ZL silica particles. The ZL silica particles

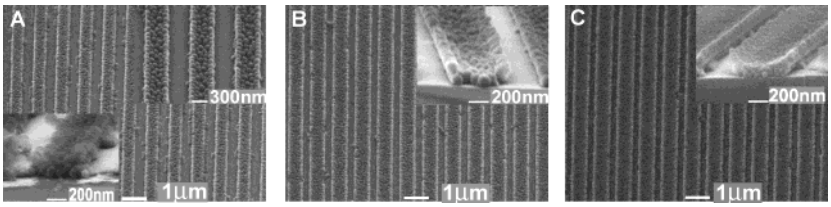


Figure 3. FE-SEM images (tilted 45°) of 1D patterns with 1000 nm period. (A) ZL (~78 nm) particle arrays at a 1:1 line/space ratio. Top right inset: magnified tilted 45° top view. Bottom left inset: cross-sectional view. (B) ZL (~78 nm) particle arrays with a 2:1 line/space ratio. Top right inset: magnified cross-sectional view. (C) OL (~50 nm) particle arrays at a 3:1 line/space ratio. Top right inset: magnified cross-sectional view.

densely filled the confined spaces between PR/wet-i walls through spin coating of silica suspension. Some isolated point defects existed on 1-D particle lines with a density of less than 5%. Well-ordered patterns of close-packed silica particles were achieved by controlling the processing conditions. The thicknesses of the silica particle layers packed inside the confined spaces could be increased by increasing the number of spin coating cycles. 1D ZL silica particle arrays were produced after 1:1 piranha treatment to remove the organic material, as shown in Figure 2C and 2D. Parallel arrays of particle lines with 500-nm period were produced over a large area (Figure 2C). Three-wide columns of ZL nanoparticles were assembled on flat surfaces (Figure 2D and top-right inset in Figure 2C). Two layers of particles were close-packed on bare and flat surfaces of Si wafer in this case (Figure 2D and bottom-left inset in Figure 2C). The width ratio between particle-occupied line and unoccupied was roughly 3:2 for this sample. The width ratios of these two regions could be tuned by adjusting exposure time or power density.

One-dimensional, two-column particle arrays were also prepared on flat surfaces with wide PR/wet-i wall patterns (data not shown). The ordered colloidal crystal parallel lines on flat surfaces are determined by template geometry (pitch, width ratio), colloidal properties (diameter, concentration, etc.), and spin-coating parameters.[26] The forces driving the silica particles into these close-packed formations are a combination of capillary, gravitational, centrifugal, and electrostatic forces, as well as the physical confinement at the PR line edges.[18] Together, these forces cause the silica particles to settle and form close-packed arrays within the confined spaces of the PR pattern. The particle arrays remain after piranha treatments or calcinations to remove the PR. The choice of approach for removal of organic templates depends on the potential applications and the subsequent processing requirements. For example, the piranha treatment is preferable if the silica arrays are to be employed as an intermediate template for a subsequent synthesis or fabrication, and the silica arrays are then removed; the calcinations are mostly likely to be a good choice if the micro/nanostructure is fabricated based on these particle arrays.

Figure 3 demonstrates the flexibility of this approach. The period of the patterns is easily adjusted by changing the incidence angles in the interferometric exposure. Figure 3 gives examples of 1-μm period patterns. The uniformity of the particle lines is good over a large area. In these large-scale, periodic patterns, the particles assembled into a thin

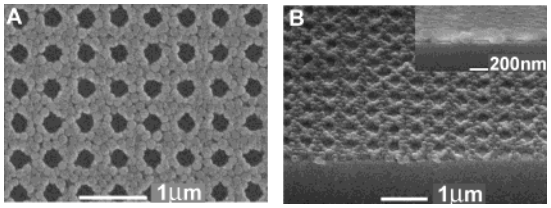


Figure 4. The FE-SEM images of 2D cross network patterns of ZL (~78 nm) particles with a 500-nm period. (A) Top-view of typical region. (B) Tilted 45° view at a cleaved edge. Top right inset: magnified cross-sectional view.

colloidal crystal as expected. As noted from Figure 3A and 3B, the morphology of the particle lines was impacted by the pattern line/space ratio for the same period and under the same experimental deposition conditions. Approximately two-layer-thick films of close-packed particles were formed for the sample with the 1:1 line/space ratio while only a single-particle-thick close-packed layer was obtained for the 2:1 line/space ratio sample. As seen in the magnified inset to Figure 3A, some particles extended under the PR as a result of the undercut of the wet-i and formed a single layer pattern. The defects were mainly grain boundaries due to stacking faults and few points of vacancy from these small silica particles relative to the pattern periods.[11,28] Another interesting phenomenon is that there was an increase in the film thickness at both line edges as shown in the inset images in Figure 3. We attribute this phenomenon to the high spin speed and relatively small ratio between particle diameter and width of the lines. This edge-high effect is minimum for parallel arrays with fewer than three columns as shown in Figure 2C and 2D, no matter how many layers of particles are deposited. Figure 3C shows ordered particle lines fabricated with smaller OL silica particles (~50-nm diameter) for a line/space ratio 3:1 sample. Close-packed, multiparticle thick silica particle lines were observed for these small silica OL particles. Moreover, the enhanced thickness at the edges of the line was more pronounced for this case.

Two-dimensional periodic patterns of particles on a flat surface were also fabricated using this lithography/deposition approach. To create the 2D periodic structures shown in Figure 4, we produced a periodic array of PR/wet-i posts on a flat surface through double exposures with positive PR, and then followed the same procedures as for 1D patterns. Periodic, continuous cross-network patterns at a 500 nm period were fabricated with ZL. The empty circular hole was about 220- to 250-nm diameter. One layer of close-packed 2D patterns was formed in this case (Figure 4B). The period

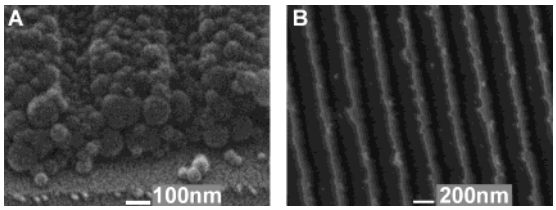


Figure 5. The FE-SEM images (tilted 45° view) of 1D patterns with 500 nm period. (A) Binary particle arrays with ZL (~78 nm) and OL (~50 nm) silica particles at a cleaved edge. (B) Dented Au lines using silica particle array template.

and hole diameter of 2D networks are tunable through adjusting parameters in the exposure step, while the multilayer deposition and particle size are controllable in the spin-coating step. We can also produce other kinds of 2D patterns (e.g., elliptical hole, or hexagonal symmetry holes) in the exposure step, thus more complex 2D particle patterns can be fabricated through this flexible approach.

The combination of large-area, nanoscale lithography and spin coating provides a very useful method for engineering complex nanoscale building blocks. The fabrication of a nanoscale patterned surface is the first step or prerequisite which may be relevant for a broad range of technological applications such as controlling crystallization, fabricating biological and chemical sensors, and engineering microelectronic and optoelectronic devices. For 1D ordered particle patterns, the patterned particle arrays with two-particle or binary particle assemblages could be fabricated further for applications in photonic band gap materials and optical devices.[11] Further, ternary particle arrays could be fabricated for potential applications in the display devices and 2D photonic crystals. As for 2D ordered particle patterns, they could be used as square templates for the growth of large colloidal particles with the (100) planes parallel to the surface of the substrate.[19,29] More complicated nanostructures with silica nanoparticles can be fabricated using these initial structures as building blocks. In addition, ordered silica particle patterns on Si wafers could serve as the patterned SiO_2 layer for epitaxial masks.[30] The process for fabrication of patterned SiO_2 layers with silica particles instead of a thermally grown SiO_2 layer was simple, rapid, and did not require additional processing: etching, metallization, etc. To our knowledge, this is the first demonstration of fabrication of periodic patterns with silica particles in the range of 50- to 100-nm diameters on a bare Si planar surface.

Figure 5 shows two typical examples of further fabrication based on 1D ordered particle patterns. Binary particle arrays were formed through a second spin coating of a smaller particle OL suspension onto an already ordered ZL particle array (Figure 5A). The small silica particle OL completely filled the gap between large silica particle (ZL) walls. Some OL particles also filled the voids between the large ZL particles on the top surface of the ZL particle walls. These silica particle patterns could also be used as masks in subsequent processing steps, such as metallization. Metal lines were deposited through a 1D array of ZL particles (Figure 5B). The Au lines were deposited by e-beam evaporation on a 1D ZL particle patterned sample prepared by immersion in piranha solution for removal of the organic layer. After the Au deposition, the silica pattern was easily removed through ultrasonic treatment and a large area of Au lines was formed on a Si flat surface. The patterned Au thin film could be used as a catalyst for aligned growth of ZnO nanowires that have potential applications in optoelectronics and field emission.[31,32] This demonstrates a new, alternative processing approach for nanoscale patterning which might be beneficial to some applications.

In summary, we have presented a simple, inexpensive, and reliable approach to directed assembly of $\lesssim$ 100-nm diameter silica nanoparticles into 1D and 2D periodic patterns on flat surfaces using interference lithography and conventional spin-coating processes. Parallel arrays of lines and continuous cross networked particle arrays with empty cylindrical holes were successfully fabricated through this approach with periods between hundreds of nanometers and several micrometers over areas of several square centimeters. The number of particle layers, the pattern period, and the line/space ratio are controllable by varying the process conditions. Interferometric lithography is capable of periods to $\lesssim$ 70 nm ($\lambda/2n$ for $\lambda = 193$ nm and $n(H_2O) = 1.44$) and spaces of ~35 nm, these small dimensions will require extension of this work to smaller particle diameters. This method provides a flexible and versatile route to the fabrication of particle arrays on a flat surface, which have potential applications in optics, electronics, sensing, and as a masking step for templating more complex nanostructures. More monodisperse particle-size distributions should result in increased local order within the particle arrays.

Acknowledgment. We thank A. Biswas, Dong Li, and A. Frauenglass for advice and technical help. Support for this work was provided by the ARO/MURI in Deep Subwavelength Optical Nanolithography and by the Mission Research Corporation under a subcontract from the Air Force Research Laboratory.

References

(1) Wijnhoven, J. E. G. J.; Vos, W. L. *Science* **1998**, *281*, 802.
(2) Yang, S. M.; Miguez, H.; Ozin, G. A. *Adv. Funct. Mater.* **2002**, *12*, 425.
(3) Veinot, J. G. C.; Yan, H.; Smith, S. M.; Cui, J.; Huang, Q.; Marks, T. J. *Nano Lett.* **2002**, *2*, 333.
(4) Yamasaki, T.; Tsutsui, T. *Appl. Phys. Lett.* **1998**, *72*, 1957.
(5) Moller, R.; Csaki, A.; Kohler, J. M.; Fritzsche, W. *Langmuir* **2001**, *17*, 5426−5430.
(6) Denis, F. A.; Hanarp, P.; Sutherland, D. S.; Dufrene, Y. F. *Nano Lett.* **2002**, *2*, 1419.
(7) Mikrajuddin; Iskandar, F.; Okuyama, K. *Adv. Mater.* **2002**, *14*, 930.
(8) Chen, X.; Chen, Z.; Fu, N.; Lu, G.; Yang, B. *Adv. Mater.* **2003**, *15*, 1413.
(9) Sprenger, M.; Walheim, S.; Schafle, C.; Steiner, U. *Adv. Mater.* **2003**, *15*, 703.
(10) Kuo, C. W.; Shiu, J. Y.; Cho, Y. H.; Chen, P. *Adv. Mater.* **2003**, *15*, 1065.
(11) Choi, D.; Yu, H. K.; Jang, S. G.; Yang, S. M. *Chem. Mater.* **2003**, *15*, 4169.
(12) Blaaderen, A. V.; Ruel, R.; Wiltzius, P. *Nature* **1997**, *385*, 321.
(13) Decher, G. *Science* **1997**, *277*, 1232.
(14) Zheng, H.; Lee, I.; Rubner, M. F.; Hammond, P. T. *Adv. Mater.* **2002**, *14*, 569.
(15) Hua, F.; Shi, J.; Lvov, Y.; Cui, T. *Nano Lett.* **2002**, *2*, 1219.

(16) Ye, Y. H.; LeBlanc, F.; Hache, A.; Truong, V. *Appl. Phys. Lett.* **2001**, *78*, 52.
(17) Zhang, J.; Alsayed, A.; Lin, K. H.; Sanyal, S.; Zhang, F.; Pao, W. J.; Balagurusamy, V. S.; Heiney, P. A.; Yoah, A. G. *Appl. Phys. Lett.* **2002**, *81*, 3176.
(18) Yin, Y.; Lu, Y.; Gates, B.; Xia, Y. *J. Am. Chem. Soc.* **2001**, *123*, 8718.
(19) Yang, S. M.; Ozin, G. A. *Chem. Commun.* **2000**, 2507.
(20) Lee, I.; Zheng, H.; Rubner, M. F.; Hammond, P. T. *Adv. Mater.* **2002**, *14*, 572.
(21) Hamann, H. F.; Wodds, S. I.; Sun, S. *Nano Lett.* **2003**, *3*, 1643.
(22) Lahav, M.; Sehayek, T.; Vaskevich, A.; Rubinstein, I. *Angew. Chem., Int. Ed.* **2003**, *42*, 5576.
(23) Ross, C. A.; Haratani, S.; Castano, F. J.; Hao, Y.; Hwang, M.; Shima, M.; Cheng, J. Y.; Vogeli, B.; Farthoud, M.; Walsh, M.; Smith, Henry I. *J. Appl. Phys.* **2002**, *91*, 6848.
(24) Brueck, S. R. J.; Zaidi, S. H.; Chen, X.; Zhang, Z. *Microelectron. Eng.* **1998**, *41/42*, 145.
(25) Raub, A. K.; Brueck, S. R. J. *Proc. SPIE* **2003**, *5040*, 667.
(26) Ozin, G. A.; Yang, S. M. *Adv. Funct. Mater.* **2001**, *11*, 95.
(27) Kuai, S.; Badilesca, S.; Bader, G.; Bruning, R.; Hu, X.; Truong, V. *Adv. Mater.* **2003**, *15*, 73.
(28) Kitaev, V.; Ozin, G. A. *Adv. Mater.* **2003**, *15*, 75.
(29) Yin, Y.; Xia, Y. *Adv. Mater.* **2001**, *13*, 267.
(30) Lee, S. C.; Malloy, K. J.; Brueck, S. R. J. *J. Appl. Phys.* **2001**, *90*, 4163.
(31) Huang, M. H.; Mao, S.; Feick, H.; Yan, H.; Wu, Y.; Kind, H.; Weber, E.; Russo, R.; Yang, P. *Science* **2001**, *292*, 1897.
(32) Zhao, Q. X.; Willander, M.; Morjan, R. R.; Hu, Q. H.; Campbell, E. E. B. *Appl. Phys. Lett.* **2003**, *83*, 165.

NL049355X

[7] A. Piattelli, A. Scarano, C. Mangano, *Biomaterials* **1996**, *17*, 1767.
[8] S. V. Dorozhkin, M. Epple, *Angew. Chem. Int. Ed.* **2002**, *41*, 3130.
[9] H. Fleich, *Bisphosphonates in Bone Disease. From the Laboratory to the Patients*, Academic Press, San Francisco, CA **2000**.
[10] See for example: *Drug Therapeutics Bull.* **2001**, *39*, 68.
[11] S. A. Doggrell, *Expert Opin. Pharmacother.* **2002**, *3*, 1007.
[12] a) H. Denissen, E. Van Beek, C. Löwik, S. Papapoulos, A. Van der Hooff, *Bone Miner*. **1994**, *25*, 123. b) H. Denissen, E. Van Beek, T. Van den Bos, J. De Blieck, C. Klein, A. Van den Hooff, *J. Bone Miner. Res.* **1997**, *12*, 290. c) H. Denissen, E. Van Beek, R. Martinetti, C. Klein, E. Van der Zee, A. Ravaglioli, *J. Periodont. Res.* **1997**, *32*, 40. d) H. Denissen, R. Martinetti, A. Van Lingen, A. Van den Hooff, *J. Periodont.* **2000**, *71*, 272. e) H. Denissen, C. Montanari, R. Martinetti, A. Van Lingen, A. Van den Hooff, *J. Periodont.* **2000**, *71*, 279.
[13] R. Z. LeGeros, in *Monographs in Oral Science* (Ed: H. M. Myers), Karger, Basel, Switzerland **1991**.
[14] B. Bourgeois, O. Laboux, O. Gauthier, L. Obadia, E. Aguado, G. Daculsi, J. M. Bouler, *J. Biomed. Mater. Res.* **2003**, *65A*, 402.
[15] L. Wilder, K. A. Jaeggi, M. Glatt, K. Müller, R. Bachmann, M. Bisping, A. Born, R. Cortesi, G. Guiglia, H. Jeker, R. Klein, U. Ramseier, J. Schmid, G. Schreiber, Y. Seltenmeyer, J. R. Green, *J. Med. Chem.* **2002**, *45*, 3721.
[16] B. N. Ames, in *Methods in Enzymology*, Vol. 8 (Eds: S. P. Colowick, N. O. Kaplan), Academic Press, Orlando, FL **1966**, p. 115.
[17] H. Freundlich, in *Colloid and Capillary Chemistry* (Ed: H. S. Hatfield), Methuen, London **1926**.
[18] J. Guicheux, D. Heymann, A. V. Rousselle, F. Gouin, P. Pilet, S. Yamada, G. Daculsi, *Bone* **1998**, *22*, 25.
[19] For a review, see: D. Heymann, J. Guicheux, F. Gouin, N. Passuti, G. Daculsi, *Cytokine* **1998**, *10*, 155.
[20] J. P. David, L. Neff, Y. Chen, M. Rincon, W. C. Horne, R. Baron, *J. Bone Miner. Res.* **1998**, *13*, 1730.
[21] K. Tezuka, T. Sato, H. Kamioka, P. J. Nijweide, K. Tanaka, T. Matsuo, M. Ohta, N. Kurihara, Y. Hakeda, M. Kumegawa, *Biochem. Biophys. Res. Commun.* **1992**, *186*, 911.
[22] B. Bujoli, S. Josse, J. Guicheux, P. Janvier, J. M. Bouler, G. Daculsi, World Patent *WO 03074098*, **2003**.
[23] Consensus Development Conference, "Diagnosis, Prophylaxis, and Treatment of Osteoporosis" *Am. J. Med.* **1993**, *94*, 645.
[24] J. M. Bouler, R. Z. LeGeros, G. Daculsi, *J. Biomed. Mater. Res.* **2000**, *51*, 680.

Directed Self-Assembly of Silica Nanoparticles into Nanometer-Scale Patterned Surfaces Using Spin-Coating**

By *Deying Xia*, *Abani Biswas*, *Dong Li*, and *Steven R. J. Brueck**

The fabrication of nanoparticle ensembles with both short- and long-range order across a hierarchy of spatial scales remains a major challenge for nanoscience. Patterned nanoparticles on different substrates have been the focus of many investigations due to their unique potential for many applications including: nanoelectronics, nanophotonics, magnetoelectronics, and biochemical sensing.[1–5] Colloidal crystallization using silica and polystyrene particles can lead to optical-wavelength photonic crystals.[6–8] Generally, conventional lithographic approaches are nearing fundamental limits and it seems clear that directed self-assembly—an integration of lithography and self-assembly—is destined to play an important role across a wide range of nanotechnology applications. This will require a sophisticated understanding of the chemical and physical interactions between lithographically defined patterns and nanoparticles. This paper is a step towards that understanding.

Fabrication approaches for nanopatterning of surfaces include: electron-beam lithography,[6] optical and interference lithography,[3,9] imprint lithography[10,11] soft lithography,[7] and colloidal[12] and self-assembly lithography.[8] These techniques can create topographical (grooves, holes, elliptical "tubs", etc.) and material variations on substrate surfaces. Even though electron-beam lithography enables fabrication of small structures with high flexibility, its serial nature makes it too slow to generate large numbers of large-area nanometer-scale patterns.[3] For self-assembly, the long-range order is limited in many cases by the formation of multiple domains or orientations with attendant grain boundaries. Imprint lithography and interference lithography (IL) are available to pattern large-area samples with nanometer-scale features. IL is very promising as a low-cost, reliable, and scalable technology for establishing sub-100 nm features across large areas.[13]

A rich variety of methods have been demonstrated for depositing colloid particles on patterned surfaces into highly ordered one-dimensional (1D), two-dimensional (2D), or

[*] Prof. S. R. J. Brueck, D. Xia, Dr. A. Biswas, D. Li
Center for High Technology Materials
University of New Mexico
1313 Goddard SE, Albuquerque, NM 87106 (USA)
E-mail: brueck@chtm.unm.edu

[**] This work was supported by the ARO/MURI in "Deep Subwavelength Optical Nanolithography", and by the Mission Research Corp. under a subcontract from the Air Force Research Laboratory. We thank A. Frauenglass for advice and technical assistance, and Jim Lyke (AFRL), Jeff Brinker (UNM), Frank von Swol (UNM), and Vakhtang Poutkradze (UNM) for useful discussions.

 DOI: 10.1002/adma.200400095 © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ADVANCED MATERIALS

three-dimensional (3D) arrays for many applications.[8,10,14,15] Typical approaches include gravity sedimentation,[6] convective self-assembly,[10,16] electrostatically induced crystallization,[8,17] and self-assembly under physical confinement.[7,18,19] Immersion methods for depositing colloidal particles are relatively slow and require a large volume of colloidal solution. Spin-coating is widely employed in semiconductor manufacturing facilities, particularly in lithography processes for the deposition of uniform photoresist layers. It is simple, convenient, and sophisticated automated tooling is available. Spinning requires only a small volume of solution, offers high reproducibility, is very fast, and builds upon a large installed manufacturing base.

In this communication, we present an approach that makes use of traditional spin-coating to assemble silica nanoparticles (< 100 nm diameter) in nanometer-scale grooves and cylindrical holes. This work is an extension of previous reports of spin-coating-directed deposition of silica particles from the micrometer regime well into the nanometer regime.[20] Silica particles were chosen for this study because of their chemical compatibility with the silica channels; their relatively high mass density (compared with latex spheres); and their chemical and thermal stability. In future studies, functionalization of the silica nanoparticles will be explored.

A field-emission scanning electron microscope (FE-SEM) was used to characterize the patterns. The influences of factors including the solution concentration and pH value, and the spin speed, etc. were investigated in order to optimize the patterns of nanoparticles in grooves on Si samples.

The procedure for the fabrication of nanometer-scale groove structures on Si wafer is described in the Experimental section, and is depicted schematically in Figure 1. The procedure for the fabrication of nanometer-scale cylindrical holes in a silicon dioxide layer on a Si substrate is similar to that for nanometer-scale groove structure on Si wafer except for double exposures (rotated by 90°) at the lithography step. Silica nanoparticles were spun onto the patterned Si samples.

For the preliminary investigation of spin-coating-directed self-assembly of silica nanoparticles into nanometer-scale grooves, the influence of different pH values on the final patterns was observed. The spin speed was 3000 rpm and ~80 nm diameter silica particles (ZL, see Experimental section) were used. For a strongly acidic solution, the ZL particles agglomerate in solution and the spin-coating does not produce an ordered result. At medium acidic solutions (~pH 4), some silica particles remain on the mesa surface, although the selectivity of deposition into the grooves is much improved. This behavior is likely caused by interactions and bonding between silica particles in the solution before spinning. The extent of inter-particle bonding depends on the silica surface chemistry, which varies with the solution pH.[21,22] At > pH 7, the adhesion and interparticle bonding is dramatically decreased so that the silica particles stay separate in the colloidal suspension.[22] The surface of as-synthesized silica particles is always terminated with a silanol group (–Si–OH), which ionizes to give a negatively charged interface at > pH 7.[17,21] Since the patterned samples used in our experiments were treated with oxygen plasma, their surface should also have had negative charges at > pH 7. As a result, both the repulsive interactions between particles and those between the particles and the surface prevent clustering in the liquid and serve to maintain the particles in the liquid as long as possible during the spin-coating for > pH 7.[17] Our experimental results at > pH 7 show that particles were deposited selectively into the grooves during the spin-coating. Despite the fact that the grooves were not fully filled in some cases, a relatively uniform distribution into these grooves was nevertheless obtained over the entire sample. In the following results, a ZL colloidal solution of pH 8.8 was used.

The concentration of silica nanoparticles also has an effect on the formation of the particle pattern found on the grooves. If an excessively high concentration is used, a thick deposit of silica particles that fully covers the wafer surface is formed without any impact of the lithographic patterning. At very low concentrations, only a sparsely populated distribution of particles forms within the patterned features. Based on initial screening experiments, a 1 wt.-% suspension was found to be the most suitable for directed deposition of dense, patterned features using 80 nm and 50 nm silica particles at spin speeds of between 2000 and 5000 rpm.

The forces contributing to depositing particles inside grooves include electrostatics, capillarity and surface tension, physical confinement, spinning forces, gravitational force, and

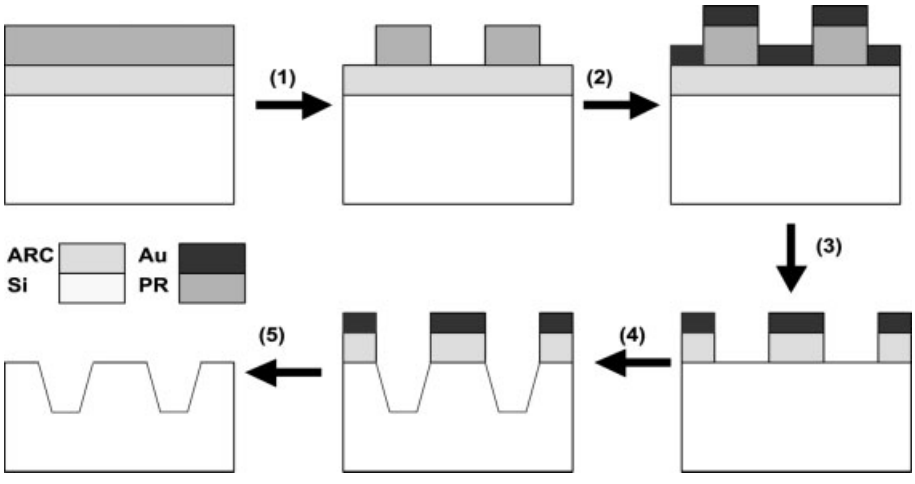


Figure 1. Schematic illustration of the fabrication of periodic arrays of grooves using interference lithography (IL): 1) expose and develop; 2) deposit Au; 3) lift-off and etch ARC; 4) etch Si; 5) remove ARC and Au.

 © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

particle–wall and particle–particle binding.[18] At the initial stages of directed self-assembly using spin-coating, the silica particle suspension spreads across the wafer without much impact from the patterning. At later stages, the combination of gravity-driven sedimentation, evaporation-induced capillary forces (dominant), centripetal forces, and blocking forces from physical confinement at the pattern edges leads to the directed self-assembly of silica particles into the confined spaces. From an experimental perspective, the important factors influencing the deposition pattern through this approach include: spin speed; particle properties (mass density, concentration, particle shape, etc.); fluid properties (pH, solvent, viscosity, evaporation, surface tension, etc.); and confinement geometry. Deposition by spin-coating is a complex nonequilibrium process.[23]

The influence of spin speed on nanoparticle deposition into grooves is an important issue.[20] For these experiments, the grooves were about 200–220 nm wide and 150–200 nm deep. ZL silica colloidal solution (pH 8.8) was used. The FE-SEM images for spin speeds of between 2000, 2500, and 3500 rpm are shown in Figure 2. There are no silica particles on the mesa tops for any of these cases. At 2000 rpm (Fig. 2A), a discontinuous column of particles is observed on the right-hand edge of several grooves. The rotation direction of the spinner determines which groove edge these particles settle onto. The centrifugal force for upper column particles at 2000 rpm is too low to overcome the physical confinement due to the groove wall. Inside the grooves, below these edge particles, there are two close-packed columns of particles. As the spin speed was increased to 2500 rpm (Fig. 2B), a loosely packed chain (one column) or a zigzag chain (one and a half columns) are the main particle-distribution patterns. At this higher spin speed, the excess particles on top of the two columns of particles were removed by the spinning forces. The larger forces at higher velocities also caused the particles inside the grooves to re-distribute and some weakly bound particles escaped leaving gaps in the particle arrays. However, at 3500 rpm (Fig. 2C), most of the particles were condensed into two parallel discontinuous close-packed columns.

An interesting phenomenon is that there is often a larger particle at the end of the deposition region followed by a gap with no particles. At this high spin speed, the redistribution of particles results in close-packed columns with larger particles terminating a region of smaller particles. This phenomenon was also observed for smaller, ~50 nm diameter silica particles (OL) in small grooves (data not shown). The cross-sectioned images (Fig. 2D) show the two columns of particles and the dimensions of grooves for spin speed of 3500 rpm. One layer of ZL silica particles resides inside the grooves due to the relatively deep groove profile. The discontinuous close-packed columns at this high spin speed are also likely to be impacted by the cross-sectional groove profiles. The groove side-wall slope was a result of the reactive-ion etching (RIE) of Si. The relatively deep groove structures may have had insufficient physical confinement in this high velocity spin-coating process. To test this, shallow groove structures (~100 nm deep) were employed for depositing ZL at 4000 rpm. A large area of close-packed, two-column patterns was obtained over the entire sample (Fig. 2E). The magnified inset images show that almost perfect two parallel columns of ZL silica particles were formed into grooves except for rare point defects and overfilled defects (indicated by the white arrows). Linear and zigzag chains of particles in these wide, shallow grooves were also observed in limited areas (Figs. 2F,G). The particle pattern in Figure 2F often appeared at the transition region between perfect two-column-particle region (Fig. 2E) and pseu-

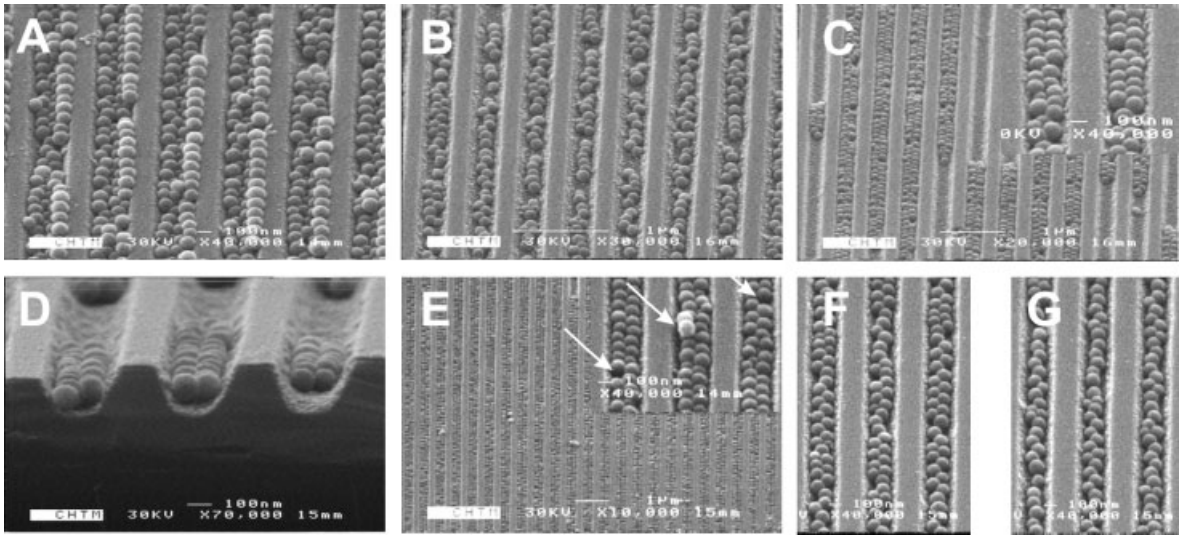


Figure 2. Field-emission scanning electron microscope (FE-SEM) images of ~80 nm diameter silica particles on grooved samples for different spin speeds: A–D) relatively deep groove structure, A) 2000 rpm; B) 2500 rpm; C) 3500 rpm. The inset is a magnified image of a close-packed pattern; D) 3500 rpm edge view; E–G) shallow groove structure, E) 4000 rpm. The inset is a magnified image of close-packed pattern with isolated defects (white arrows); F) 4000 rpm zigzag chain region; G) 4000 rpm pseudo-linear chain region.

 © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ADVANCED MATERIALS

do-one-column-particle region (Fig. 2G). Typical sample sizes were ~2 cm^2; consistent and reproducible deposition was observed over more than 80 % of the sample area exclusive of edge areas where the fluid flow was complicated by geometrical factors. Spin-coating is widely used in industry for much larger samples (300 mm diameter wafers) and we expect that these results will readily scale to much larger areas.

The silica particle size, groove width, groove profile, and roughness also have an influence on particle distribution. A spin speed of 4000 rpm was used for this set of experiments. If the grooves were narrowed to ~80–120 nm wide, only a single column or a zigzag column (~1.5 particles wide) of ZL silica particles was formed. Overfilled particles appear in some regions. Figures 3A,B show that one column of ZL is found for 80 nm wide grooves; however, zigzag-shaped particle arrangements appear for the overfilled case. Figure 3C shows an image of an overfilled two-column particle pattern for ~100 nm wide grooves. It is likely that the combination of spinning force and physical confinement at the groove side wall lead to the formation of these tightly packed, overfilled particle patterns. When a smaller OL particle suspension (50 nm diameter) was used with the same samples, a larger number of parallel columns of nanoparticles were found in the same grooves (Figs. 3D,E). Two or three columns of OL silica particles were assembled into ~120 nm wide grooves (Fig. 3D) while four or five particle wide columns of OL silica particles were deposited into ~200 nm wide grooves (Fig. 3E). Isolated OL silica particles were sparsely adhered on to the mesa tops. This was possibly as a result of the relatively low pH value (~4.8) of the OL solution. Figures 3D,E show that the smaller OL silica particles demonstrated a lower order. This was partially because the smaller silica nanoparticles have a larger polydispersity and potentially because they were more comparable in scale to the groove roughness; additionally this may be due to the fact that the effect of confinement on packing order is expected to become weaker with an increase in ratio of confinement size (groove width) to the silica diameter.[24]

The spin-coating-directed self-assembly of silica nanoparticles into 2D nanometer-scale patterns was investigated as well. ZL silica nanoparticles were spun into nanometer-scale cylindrical holes in an SiO_2 layer on top of a Si substrate with a spin speed of 3000 rpm. Two patterned samples with different hole diameters were tested for holding silica particles. Figures 4A,B show the images of the large-hole patterned sample before spin-coating. The diameter of the holes was 200–230 nm with good pattern coverage on almost the entire sample area (Fig. 4A). The hole sidewalls were vertical and the depth (thickness of the SiO_2 layer) was ~60 nm (Fig. 4B). After spin-coating, the large holes held one, two, three, or four ZL particles (Fig. 4C–F). The typical polygonal aggregates on this large-hole patterned sample were the colloidal cluster of triangle (trimer) and square (regular quatramer) (Fig. 4C–E). This confirms the predictions from the calculation of colloidal cluster as a function of ratio (D/d) between the hole dimensions (D) and the diameter of silica particles (d).[17,18] Holes with two particles were seldom observed, and a hole with only one particle was rare (Fig. 4D). However, in some holes, the four or five particles would overfill the holes (Figs. 4C,E). Figure 4F shows typical images of self-assembled clusters inside holes and overfilled cases. The yield of clusters of particles in cylindrical holes was above 90 %. Several kinds of polygonal aggregates exist on the same sample because of the large variation of diameter (d) for silica nanoparticles.[24] If the more uniform nanoparticles were available, we believe that well-defined aggregates could be obtained in nanometer-scale patterned structures.

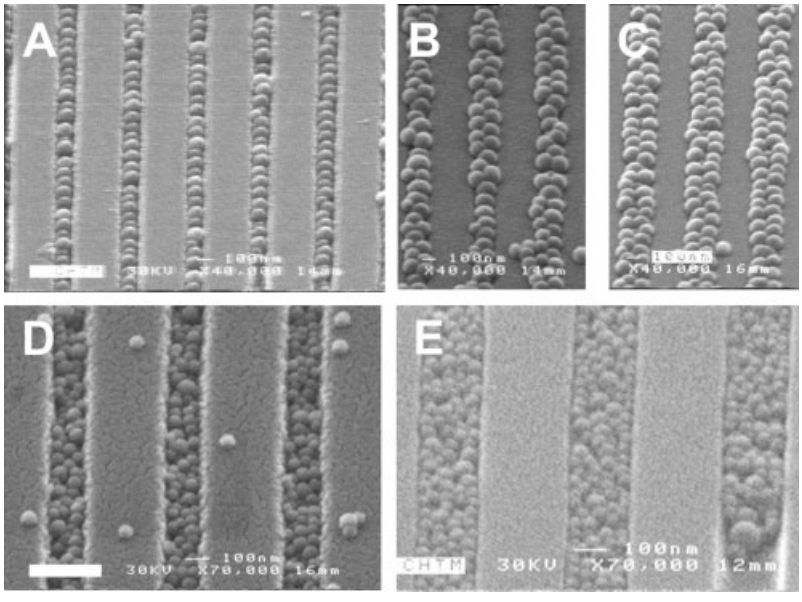


Figure 3. The FE-SEM images of 80 nm (ZL) and 50 nm (OL) silica particles on narrow-grooved samples: A) one column of ZL particles into 80 nm grooves; B) zigzag chain of overfilled ZL particles in 80 nm grooves; C) two columns of overfilled ZL particles in 100 nm grooves; D) OL particle in 120 nm grooves; E) OL particle in 200 nm grooves.

As the hole diameter was decreased to 90–120 nm, the number of particles in the cluster was decreased as expected (Fig. 5). In general, monomers and dimers were the typical clusters inside small holes (Figs. 5A,B). The maximum number was three with overfilled packing. The monomer dominates among these three types of clusters (Fig. 5C). Figure 5D shows typical images of self-assembled clusters inside holes and overfilled cases. Some small islands of close-packed particles associated with a hole position were also observed in some parts of surface (Figs. 4E, 5B,C).

In summary, we have presented a simple, inexpensive, rapid, and reliable approach to the assembly of silica nanoparticles (< 100 nm diameter) into nanometer-scale Si grooves and cylindrical holes using a conventional spin-coating process. The nanometer-scale-patterned samples were prepared using interfero-

 © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

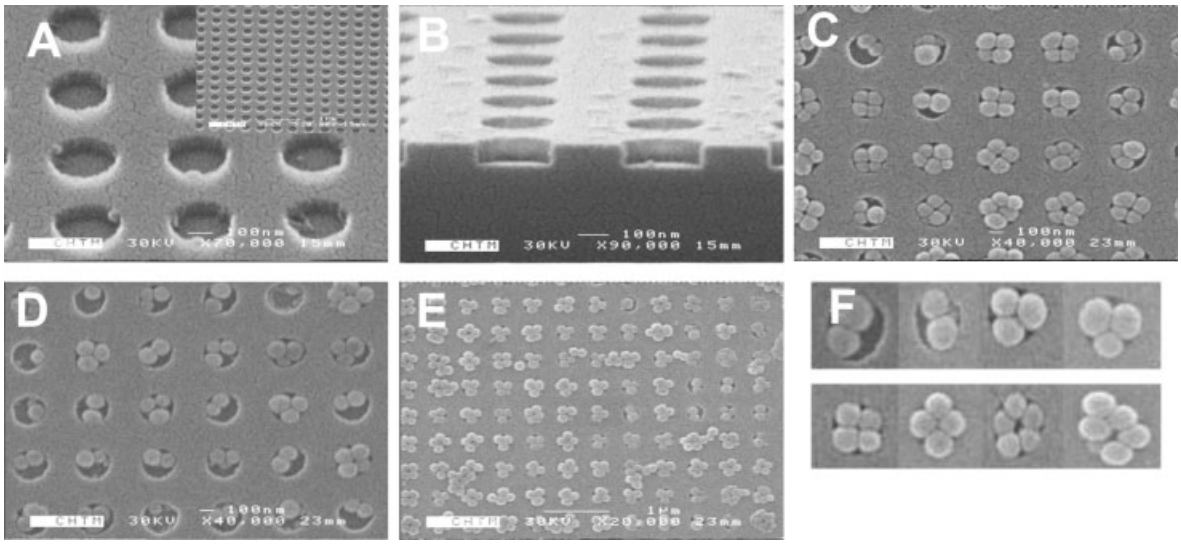


Figure 4. The FE-SEM images of 80 nm silica particles on large cylindrical-hole samples (dia. 200–230 nm): A) hole pattern before spin-coating. The inset image is a view of a large area of the pattern; B) side view of hole pattern before spin-coating; C,D) typical regions with particle clusters; E) low magnification of typical region; F) typical clusters of particles.

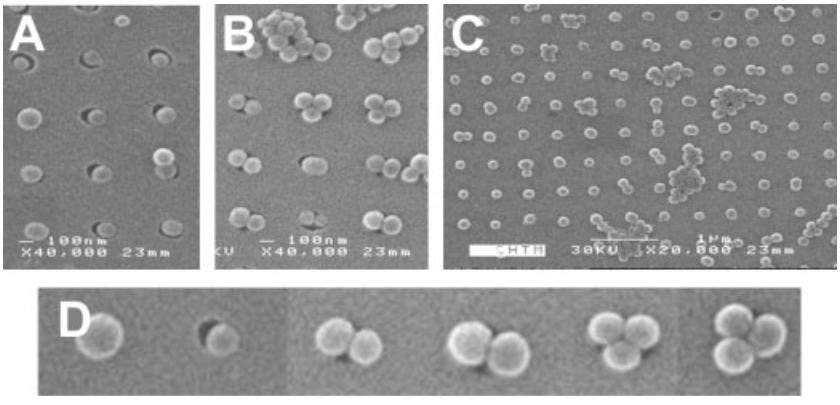


Figure 5. FE-SEM images of 80 nm diameter silica particles on small cylindrical hole samples (dia. 90–120 nm): A,B) typical regions with particle clusters; C) typical region; D) typical particle clusters.

metric lithography. The width of grooves, diameter of cylindrical holes, and the silica particle size could be less than 100 nm over large areas. We showed that the silica nanoparticle distributions depended on the spin speed, the pH value, the groove geometry (width, depth, sidewall slope) and geometry of cylindrical holes (diameter, depth), and the particle size. Linear (one particle wide), zigzag (1.5 particles wide) chains, and two columns of silica nanoparticles could be deposited into grooves by adjusting the above parameters. We also have demonstrated that long-range 2D arrays of silica nanoparticles including colloidal clusters of monomers, dimers, trimers, and regular quatramers on patterned samples could be controlled by adjustment of the ratio between the diameters of holes and particles. Well-defined clusters of 2D array of nanoparticles inside 2D patterns could be obtained using this approach if the uniform nanoparticles were employed. We believe that extensions of this work will lead to directed self-assembly of particles in grooves and 2D structures. This will be useful for applications in photonics and nanoelectronics.[18] A wide range of potential applications exist for the directed self-assembly of functional nanoparticles, especially biological moieties, into designed patterns using this approach. Directed self-assembly is an important method for large-scale patterned nanostructured materials with the hierarchical range of length scales required to translate nanometer-scale phenomena into macroscopic effects.

Experimental

The fabrication of periodic arrays of nanometer-scale groove patterns in silicon wafers includes using interferometric lithography (IL), lift-off, and reactive-ion etching (RIE). Initially, the silicon sample was put into piranha solution and BOE 20:1 solution in order to clean the sample surface. Following this, an antireflective (ARC) polymer

 © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

ADVANCED MATERIALS

layer (Brewer Science XHRIC-16) was spun on the sample at 4000 rpm for 30 s and oven prebaked at 175 °C for 3 min. Photoresist (PR; Shipley 510A) was spun on the ARC at 4000 rpm and oven prebaked for 3 min at 95 °C. IL with 355 nm third-harmonic YAG laser beam was used to produce periodic nanopatterns in the PR [9]. The angle between the two beams determined the 520 nm pattern period while the exposure dose and development time affected the detailed pattern shape. After exposure, the resist was post baked for 2 min at 110 °C and developed using a MF-702 developer for 45 s followed by a rinse with deionized (DI) water. Pattern transfer was effected via lift-off of a thin gold film (~30 nm) that was deposited by electron-beam evaporation to form a hard mask and RIE of the underlying ARC and the silicon. Acetone was then used to liftoff the PR leaving a hard Au mask atop the ARC layer. Oxygen plasma was used to etch through the ARC for 3 min with the power of 50 W while a mixture of O_2 and CHF_3 (50:130) was used to etch Si for 3–6 min at a power of 150 W for an etch depth ranging from 30 to 200 nm. Finally, the ARC/Au layer was removed in an ultrasonic bath using piranha. The fabrication of periodic arrays of nanometer-scale cylindrical hole was performed on Si wafer coated with a 60 nm SiO_2 layer. The SiO_2 was thermally grown in a dry-O_2 oxidation furnace. The processing for fabrication 2D patterns was similar to that for groove structures except for double exposures and inductively coupled plasma (ICP) etching of the SiO_2. After the samples were exposed with a first 1D grating, they were rotated 90° and exposed a second time to obtain 2D photoresist patterns (here posts for positive photoresist). After lift-off of a Au hard mask as above, an ICP system was next used to etch the SiO_2 in the regions where there was no Au hard mask in etching step. A mixture of Ar and CHF_3 (3:1) was used to etch SiO_2 for 2 min at the RF power of 500 W, bias of 150 W, and the pressure of 10 mtorr. The ICP etching could etch selectively SiO_2 with the mixture of Ar and CHF_3 in this case.

Two particle sizes were investigated. Both are from Nissan Chemical Industries. Ltd. The larger silica particles (1 wt.-%, ~pH 8.8) were prepared by diluting Snowtex ZL (40.9 wt.-%, ~pH 9.6, 78±16 nm diameter [25]) with DI water. The smaller silica nanoparticles (1 wt.-%, ~pH 4.8) were prepared by diluting Snowtex OL (20 wt.-%, ~pH 3, 50±13 nm diameter) with DI water. HCl or NH_4OH were used to adjust the pH value of the water-based colloidal solutions. Before spin-coating, the treatment of an oxygen plasma RIE was used in order to ensure that the sample surface was hydrophilic. Next, the silica nanoparticle colloidal solution was spun onto the patterned samples. Images were obtained using field emission scanning electron microscopy (FE-SEM) (tilted at 45°) at 30 kV with a 10 nm thick Au film to minimize charging.

Received: January 22, 2004
Final version: March 26, 2004

[1] J. E. G. J. Wijnhoven, W. L. Vos, *Science* **1998**, *281*, 802.
[2] B. Gates, Y. Xia, *Appl. Phys. Lett.* **2001**, *78*, 3178.
[3] C. A. Ross, S. Haratani, F. J. Castano, Y. Hao, M. Hwang, M. Shima, J. Y. Cheng, B. Vogeli, M. Farthoud, M. Walsh, H. I. Smith, *J. Appl. Phys.* **2002**, *91*, 6848.
[4] R. Moller, A. Csaki, J. M. Kohler, W. Fritzsche, *Langmuir* **2001**, *17*, 5426.
[5] F. J. Castan, Y. Hao, M. Hwang, C. A. Ross, B. Vogeli, H. I. Smith, S. Haratani, *Appl. Phys. Lett.* **2001**, *79*, 1504.
[6] A. V. Blaaderen, R. Ruel, P. Wiltzius, *Nature* **1997**, *385*, 321.
[7] S. M. Yang, G. A. Ozin, *Chem. Commun.* **2000**, 2507.
[8] H. Zheng, I. Lee, M. F. Rubner, P. T. Hammond, *Adv. Mater.* **2002**, *14*, 569.
[9] S. C. Lee, K. J. Malloy, S. R. J. Brueck, *J. Appl. Phys.* **2001**, *90*, 4163.
[10] J. Zhang, A. Alsayed, K. H. Lin, S. Sanyal, F. Zhang, W. J. Pao, V. S. Balagurusamy, P. A. Heiney, A. G. Yoah, *Appl. Phys. Lett.* **2002**, *81*, 3176.
[11] J. Aizenberg, P. V. Braun, P. Wilzius, *Phys. Rev. Lett.* **2000**, *84*, 2997.
[12] F. A. Denis, P. Hanarp, D. S. Sutherland, Y. F. Dufrene, *Nano Lett.* **2002**, *2*, 1419.
[13] S. R. J. Brueck, S. H. Zaidi, X. Chen, Z. Zhang, *Microelectron. Eng.* **1998**, *41–42*, 145.
[14] Y. Xia, B. Gates, Y. Yin, Y. Lu, *Adv. Mater.* **2000**, *12*, 693.
[15] Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, H. Yan, *Adv. Mater.* **2003**, *15*, 353.
[16] Y. H. Ye, F. LeBlanc, A. Hache, V. Truong, *Appl. Phys. Lett.* **2001**, *78*, 52.
[17] I. Lee, H. Zheng, M. F. Rubner, P. T. Hammond, *Adv. Mater.* **2002**, *14*, 572.
[18] Y. Yin, Y. Lu, B. Gates, Y. Xia, *J. Am. Chem. Soc.* **2001**, *123*, 8718.
[19] Y. Yin, Y. Xia, *Adv. Mater.* **2002**, *14*, 605.
[20] G. A. Ozin, S. M. Yang, *Adv. Funct. Mater.* **2001**, *11*, 95.
[21] R. K. Iler, *The Chemistry of Silica*, Wiley, New York **1979**.
[22] J. D. Batteas, M. K. Weldon, K. Raghavachan, in *Nanotribology: Critical Assessments and Future Research Need* (Eds: S. M. Hsu, Z. C. Ying), Kluwer Academic, Boston, MA **2000**, Ch. 27.
[23] D. Wang, H. Mohwald, *Adv. Mater.* **2004**, *16*, 244.
[24] V. Kitaev, G. A. Ozin, *Adv. Mater.* **2003**, *15*, 75.
[25] F. Hua, J. Shi, Y. Lvov, T. Cui, *Nano Lett.* **2002**, *2*, 1219.

Ordered, Nanostructured Tin-Based Oxides/Carbon Composite as the Negative-Electrode Material for Lithium-Ion Batteries**

By *Jie Fan*, *Tao Wang*, *Chengzhong Yu*, *Bo Tu*, *Zhiyu Jiang*, and *Dongyuan Zhao**

Nanostructured materials have received considerable attention because of their novel electronic, optical, magnetic, and catalytic properties. Supermolecules, such as surfactants and copolymers, are valuable structure-directing agents for preparing these novel materials with diverse framework compositions (including silicas, metals, metal oxides/phosphates/sulfides, and organosilicates).[1,2] However, this process encounters limitations when the formation of some inorganic networks cannot be efficiently controlled. Recently, inorganic templates (such as mesoporous silica, mesoporous carbon, nanocrystals) have been used to synthesize mesoporous materials based on a nanocasting process.[3–6]

Tin-based oxides have attracted much attention, as they offer a promising carbon alternative to negative-electrode mate-

[*] Prof. D. Zhao, Dr. J. Fan, Dr. T. Wang, Dr. C. Yu, Dr. B. Tu, Prof. Z. Jiang
Shanghai Key Laboratory of Molecular Catalysis and Innovative Materials, Department of Chemistry
Fudan University
Shanghai, 200433 (P.R. China)
E-mail: dyzhao@fudan.edu.cn

[**] The authors thank Dr. Q. Wu and Q. Zhou for helpful discussion. This work was supported by the National Science Foundation of China (20101002, 20173012 and 20233030), State Key Research Program (G2000048001 and 2002AA321010), Shanghai Sci. & Tech. Committee (0212nm043, 03DJ14004, 03527001). Supporting Information is available online at Wiley InterScience or from the author.

 © 2004 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim DOI: 10.1002/adma.200400106

Lithographically directed deposition of silica nanoparticles using spin coating

Deying Xia and S. R. J. Brueck[a)]
Center for High Technology Materials and Department of Electrical and Computer Engineering, University of New Mexico, 1313 Goddard SE, Albuquerque, New Mexico 87106

(Received 3 June 2004; accepted 27 September 2004; published 14 December 2004)

Directed self-assembly is an important direction for the extension of patterning to the nanoscale regime and below. Here, examples are given of both individual particle placement and composite structure formation involving assemblies of many nanometer scale particles, both on patterned surfaces and on planar surfaces where the photoresist is used as a sacrificial layer to define the pattern and is subsequently removed. The underlying technologies employed in these experiments include interferometric lithography to define large-area, nanometer-scale patterns and directed self-assembly by spin coating to control particle placement. Three sizes of silica nanoparticles (mean diameters: 78, 50, and 15 nm) were employed for spin-coating processes. Single linear silica particle chain patterns and isolated two-dimensional particle patterns were easily formed on patterned surfaces. Silica particle rows, cross networks, and isolated posts with controllable thickness could be formed on flat surfaces using this approach. Directed self-assembly using nanoscale lithography and spin coating is a facile approach to the extension of lithographic techniques to the nanoscale. © *2004 American Vacuum Society.* [DOI: 10.1116/1.1821582]

I. INTRODUCTION

Organized patterns of nanoscale particles have a wide range of potential applications in photonic crystals,[1,2] biosensors,[3,4] magneto-electronics,[5] nanofluidics,[6] and other emerging fields. For example, arrays of small magnetic elements have potential application to high-density data storage devices and magnetic random access memory devices. Isolated particles in periodic arrays can be initiators for the growth of colloidal crystals with long-range order[7] and patterned growth for some special materials using catalyst particles.[8] Top-down, lithographic surface patterning has been achieved by electron beam lithography,[9] photolithography, interferometric lithography,[10] and soft lithography while directed assembly of particles into the lithographically defined patterns has been accomplished by sedimentation,[9] parallel plate fluidic confinement,[1] spin-coating,[2,11] convective deposition (directed evaporation induced self-assembly),[12,13] and electrostatic (layer-by-layer) deposition.[14] Most of this work has been focused on particle diameters from above 100 nm to $\sim 1\ \mu$m or larger.

While significant advances continue to be made in conventional, top-down, lithographic capabilities, it is unquestionable that the extension of traditional techniques to the 10 nm scale and below is going to be increasingly difficult. At the same time, bottom-up self-assembly techniques have been very successful at nanometer scales and below, but tend to have increasing difficulty as the scale of the pattern is increased, for example with grain boundaries arising from independent nucleation of colloidal crystals at uncorrelated positions. The combination of these two approaches, directed self-assembly, is an interesting alternative that will allow us to maintain the exquisite hierarchical capabilities of traditional lithography at larger scales while extending to subnanometer regimes with the precision and accuracy that molecular forces provide. Directed self-assembly—an integration of lithography and self-assembly—is destined to play an important role across a wide range of nanotechnology applications.

Interferometric lithography (IL) can be used to produce highly ordered, periodic nanostructure arrays over macroscopic surface areas with easily varied feature dimensions. This process is fast and inexpensive. In this work, we use IL to define nanometer-scale patterns. Meanwhile, spin coating has been employed to rapidly form two-dimensional (2D) colloidal crystals. Outstanding advantages of the spin-coating process include its rapidity, the low volume of solute required, and the availability of highly capable instrumentation as a result of its widespread use in the semiconductor industry. The spin-coating approach is employed in this work to deposit silica nanoparticles on patterned surfaces with both hard-mask (e.g., SiO_2) and soft-mask (photoresist) defined patterns. For the soft-mask patterns, a "lost wax" process is used to remove the photoresist after the spin coating, leaving the particle patterns on a flat surface.[15] Continuous and isolated nanoparticle patterns are generated and may prove useful as templates for nanoscale patterned growth and nanosphere optics, and as building blocks for more complex nanostructures.

II. EXPERIMENT

A. Interferometric lithography (IL)

In interferometric lithography, two coherent laser beams with wavelength λ are crossed at an angle 2θ to produce a periodic interference pattern on a photoresist (PR) film with

[a)]Author to whom correspondence should be addressed; electronic mail: brueck@chtm.unm.edu

the spatial period $d=\lambda/(2\sin\theta)$.[16] The angle between the two beams determines the pattern period while the exposure time, power density, and development time impact the linewidth and the pattern morphology. With an ultraviolet light source ($\lambda=355$ nm, third-harmonic from a Nd:YAG laser), one can easily obtain periods on the order of hundreds of nanometers.

B. Silica nanoparticle suspensions

Three sizes of silica nanoparticles were employed in the spin-coating process. They are commercial products of Nissan Chemical America Corp. (Houston, TX). The larger silica particles were Snowtex ZL with 78 nm mean diameter. The medium silica nanoparticles were Snowtex OL with 50 nm mean diameter. The smaller silica nanoparticles were Snowtex C with 15 nm mean diameter. The desired silica nanoparticle concentrations were prepared by diluting corresponding Snowtex products with deionized (DI) water.

C. Directed deposition of silica nanoparticles on patterned surfaces

The first step is to fabricate the periodically patterned grooves and holes on a silicon wafer with a thin SiO_2 layer. Initially, the silicon sample was put into piranha solution (sulfuric acid and hydrogen peroxide with volume ratio 4:1) for 8 min, subsequently into buffered oxide etchant (BOE) 20:1 solution for 2 min to clean the sample surface. A thin SiO_2 layer ($\sim$60 nm) was prepared with dry thermal oxidation at 900 to 1000 °C. Then an antireflective coating (ARC) layer (XHRiC-16, Brewer Science, Inc., Rolla, MO) was spun on the sample at 4000 rpm for 30 s and oven-baked at 175 °C for 3 min yielding a final thickness of $\sim$150 nm. Diluted photoresist (SPR 510A, Shipley, Inc., Marlborough, MA) was spun atop the ARC at 4000 rpm followed by a similar oven bake for 3 min at 95 °C for a final thickness of $\sim$250 nm.

Subsequently, IL was used to produce the PR patterns. The exposure times were 20–45 s with power 50 or 150 mJ. A 2 min post-exposure bake at 110 °C was used. The sample, after exposure and post-bake, was developed (MF-702, Shipley, Inc.) for 45 s and rinsed with DI water. After definition of the PR pattern, an e-beam deposition of gold (Au) was carried out at a pressure of 10^{-6} mTorr with a deposition rate of $\sim$1 Å/s until a thickness of $\sim$300 Å was reached to provide a hard mask. Acetone was then used to dissolve and lift off the remaining regions of PR with Au capping layers. The sample was then put in the vacuum chamber of a reactive ion etching (RIE) system to etch the ARC and silicon oxide layers in the regions where there was no protecting Au hard mask. Oxygen was used to etch ARC for 3 min at a power of 50 W, a pressure of 50 mTorr while a mixture of O_2 and CHF_3 (pressure ratio: 50:130) was used to etch SiO_2 for 3 to $\sim$6 min at a power of 150 W. Finally, after groove patterns were produced in the SiO_2 layer by RIE, piranha was used in an ultrasonic bath to remove the ARC layer along with the top Au capping layer. Hole samples with 60-nm-thick SiO_2 were prepared similarly except for double-IL exposures with crossed 1D patterns and inductively coupled plasma (ICP) etching of the SiO_2. After the samples were exposed with a first 1D grating, they were rotated 90° and exposed a second time to obtain 2D positive photoresist patterns. After lift off of an Au hard mask as above, an ICP system was next used to etch the SiO_2 in the regions where there was no Au hard mask in etching step. A mixture of Ar and CHF_3 (3:1) was used to etch 60 nm SiO_2 for 2 min at the rf power of 500 W, bias of 150 W, and the pressure of 10 mTorr. The ICP process selectively etches SiO_2 with vertical sidewalls.

In the deposition step, an aqueous silica nanoparticle suspension (1 wt %) was spun on the patterned sample. Before spin coating, an oxygen plasma RIE (1 to $\sim$2 min) was used to ensure that the SiO_2 surface was hydrophilic. One drop of suspension solution was applied on the clean surface of the sample. The spin coating was 30 s at 3000 rpm with an acceleration of 100 rpm/s. After spin coating, the sample was baked at 90 °C for 5 min to drive off any residual solvent. The resulting structures were characterized by field-emission scanning-electron microscopy (FE-SEM) at 30 kV with a 10 nm thick Au film sputtered on the sample surface. Cross-section views of images were also obtained to check the structure of the ordered patterns.

D. Directed deposition of silica nanoparticles on planar surfaces using soft PR masks

In the process of the directed deposition of silica nanoparticles on planar surfaces, a wet developable BARC for I-line lithography (wet-i™ 10-7, Brewer Science, Inc., Rolla, MO) was spun on a precleaned sample at 3000 rpm for 30 s. The baking time for wet-i was 90 s at 200 °C. A positive-tone photoresist (SPR510A, Shipley, Inc.) diluted to produce a $\sim$200-nm-thick film was used in these experiments. IL was used to produce the periodic patterns on this PR/wet-i film stack. Parallel PR/wet-i lines (1D) were formed on the Si wafer surface after exposing, baking, and developing. For 2D arrays of nanoscale posts, a double exposure with a 90° rotation of the sample was performed, and all other steps were the same as those for 1D patterns. For 2D arrays of nanoscale holes, a negative-tone photoresist (NR7-500P, Futurrex, Inc., Franklin, NJ) was used in IL.

After the PR patterning, silica nanoparticle colloidal dispersions (5 wt %) were spun coated onto these samples. In general, the spin coating protocol was 30 s at 4000 rpm with an acceleration of 100 rpm/s^2. Suspensions of small silica nanoparticles (OL:50 nm, C:15 nm) were utilized for directed assembly of 1D and 2D periodic particle structures supported on flat surfaces. The suspension was agitated for 5 min in an ultrasonic bath before the coating step.

The final step was to strip off the PR/wet-i patterns leaving the 1D or 2D ordered patterns of silica nanoparticles on the flat surface of the Si wafer. Two methods—immersion in piranha solution and calcination at high temperature—were successful in completely removing the PR/wet-i while not disturbing the silica particle arrays. In general, we soaked the as-patterned samples in 1:1 piranha solution for 15 min for

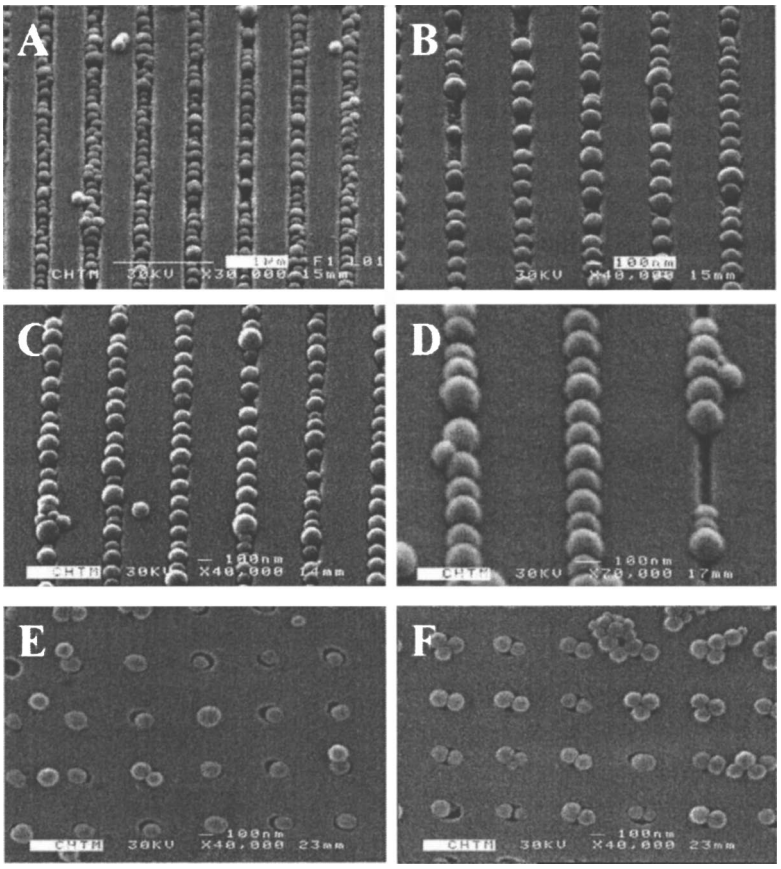


FIG. 1. Field emission scanning electron microscope (FE-SEM) images (tilted 45° view) of ~78-nm diameter silica particles on grooves (1D) and holes (2D): (a) ~100-nm-wide groove sample; (b) ~80-nm-wide groove sample; (c) ~60-nm-wide groove sample; (d) ≲50-nm-wide groove sample; and (e)–(f) ~100 nm hole sample.

the immersion method, while the as-patterned samples were put into furnace at 800 °C for 1.5–2 h for the calcination method. In the calcination approach, the mechanical stability of particle patterns on flat surfaces was enhanced as a result of partial sintering between the silica particles and the Si surface at these elevated temperatures.

III. RESULTS AND DISCUSSION

A. Directed deposition of silica nanoparticles on patterned surfaces

Several methods for deposition of silica particles on patterned surfaces were investigated before the spin-coating approach was adopted. These included convective deposition and dip coating. In convective deposition, the sample was mounted vertically in a beaker containing diluted aqueous solution of silica particles. The beaker was kept in a temperature-controlled oven. The temperature was always maintained at 55–60 °C and the fluid was gradually evaporated. In reverse dip coating (bottom-flow deposition), the sample was again mounted vertically in a beaker containing the silica solution. A drain port and valve at the bottom of the beaker allowed for controlled removal of the liquid leaving behind the deposited particles on the sample, essentially the inverse process of lifting the sample out of the liquid. Even though in some cases the results for dip coating and convective deposition are similar to those for spin coating, a large amount of silica particle suspension is consumed in each experiment and the deposition time is considerably longer than for spin coating. Moreover, for these processes the distribution of silica particles on patterned samples was relatively nonuniform across the sample, and alternating bands of dense and sparse particle deposition were observed. This phenomenon of multiple band patterns for evaporative depositions has been previously observed.[17] To fabricate a uniform particle distribution, it is required to reduce the evaporation rate and/or to tilt the sample at an angle greater than 10°. Based on these issues and its comparative success and ease of use, spin coating was adopted as the principle approach to the deposition of silica particles on patterned samples.

Several factors effect the spin-coating deposition of silica nanoparticles on patterned samples, including the spin speed, the solution pH value, and the surface pattern geometry.[10] We investigated the influence of these factors in preliminary experiments. Silica particles were deposited selectively into the grooves or holes during spin coating using silica suspensions with pH ≳ 7. This pH ensured that the particles remained negatively charged and separated in the solution. For more acidic solutions, the particles tended to agglomerate in solution and deposit in large clumps independent of the surface patterning. The spinning program was key to ensuring the deposition was exclusively in the grooves or holes rather than on the top surface of wafer. Our results show that a relatively high speed of spin coating was helpful for deposition of particle on patterned surfaces. The speed of 4000 rpm was used for the experiments reported here. The concentra-

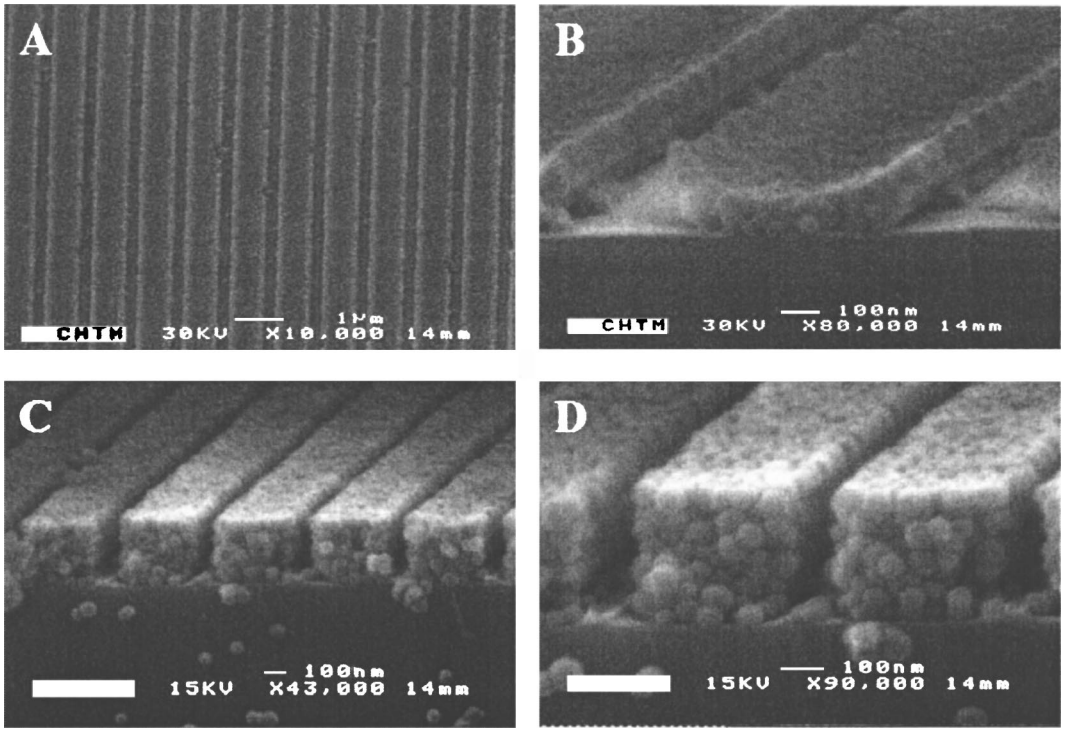


FIG. 2. FE-SEM images of 1D particle arrays with ~50 nm diameter silica particles on planar surfaces: (a) large area image (tilted 45° view) with 1000 nm period; (b) magnified cross-section image with 1000 nm period; (c) large area cross-section image with 500 nm period after two cycles of spin-coating; and (d) magnified cross-section image with 500 nm period after two cycles of spin-coating.

tion of silica particles also has an effect on the formation of the particle pattern found on the patterned surfaces. If an excessively high concentration is used, a thick deposition of silica particles that fully covers the wafer surface is formed without any impact of the lithographic patterning. At the other extreme of very low concentrations, only a sparsely populated distribution of particles forms within the patterned features. Based on our initial screening experiments, a 1 wt % suspension was found to be most suitable for directed deposition of ZL silica particles on patterned surfaces for these experiments. Due to accelerated evaporation during spin coating, factors such as humidity and ambient room temperature have little effect on the particle patterns.

The directed deposition of silica nanoparticles on narrow-groove samples was studied. These 1D patterned samples have $\lesssim$100-nm-wide grooves with a 500 nm period. A diluted suspension of 78±16-nm diameter silica nanoparticles (1 wt %, pH~8.8) was prepared by diluting Snowtex ZL (40.9% wt %, ~pH9.6) with DI water. The typical deposition patterns are shown in Fig. 1. In Fig. 1(a), the grooves are relatively wide (~100 nm) and deep, so the silica particles are almost completely contained within the grooves. Very few silica particles settle on the flat surface of the samples away from the grooves, while some voids exist along the particle chains. Due to the large distribution of silica particle sizes, zigzag chains of silica particles were observed in some parts of the grooves. In Figs. 1(b) and 1(c), the grooves are narrower (~60–80 nm) and half-buried particle chains are formed along the parallel grooves. In addition, void defects appear more often and exhibit longer gaps on these samples. In Fig. 1(d), the width of grooves is less than 50 nm. The bulk of the silica particles protrude above the surface of the sample. A much higher density of void defects exist on this sub-50 nm grooved sample.

The more important forces in this deposition process are the centripetal force associated with the spinning, the particle-wall blocking forces, and surface tension forces. As the width of grooves decreases, the blocking force from the sidewall of the groove becomes weaker compared to the centripetal forces. The silica particles are easier to dislodge from the narrow grooves. In order to obtain good coverage of particles into grooves, it is important to match the particle size to groove spacing.

The deposition of the 78-nm diameter silica particles on 2D patterned samples was investigated as well. The diameter of holes in the 2D pattern is ~100 nm at a period of 500 nm. In most holes, single particles (monomer) or particle pairs (dimers) are found [Figs. 1(e) and 1(f)]. For the monomer case, the individual particle is held inside a hole. Occasionally, some holes are empty without any particles while some particles appear on the surfaces. For the dimer case, the two particles overfill the holes. In rarer cases, trimers are found as shown in Fig. 1(f). Some clusters of particles also appear on some parts of surfaces. We attribute the multiple modes of particle clusters to the wide distribution of silica particle sizes and also to the variation of hole size. The types of particle clusters follow the predictions from the calculation of colloidal cluster coordination as function of ratio (D/d) between the hole dimensions (D) and the diameter of silica particles (d).[18]

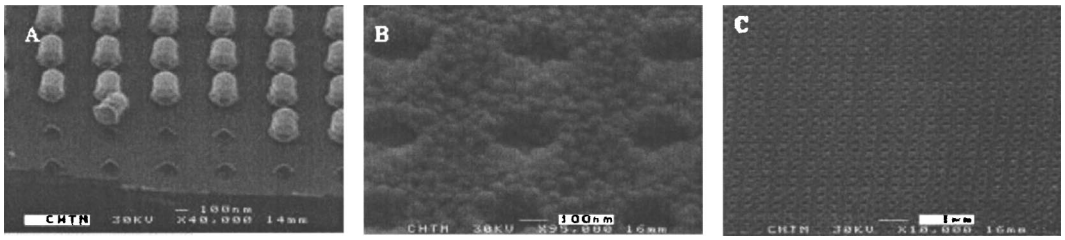


FIG. 3. FE-SEM images (tilted 45° view) of 2D, 50-nm-diameter particle arrays on planar surfaces with ~500 nm periods: (a) PR/wet-i posts for continuous cross network of particles before deposition of silica particles with spin-coating; (b) magnified image of continuous cross network with ~50 nm silica particle; and (c) large area image of continuous cross network with ~50 nm silica particle.

B. Directed deposition of silica nanoparticles on planar surfaces

Two types of small silica nanoparticles were used for directed deposition using soft PR masks on planar surfaces. They have the mean diameters of 50 ± 13 nm (OL) and 15 ± 5 nm (C), respectively. Approximately 5 wt % aqueous suspensions of OL and C were used. The pH values of suspensions have less influence on the final, multilayer, close-packed particle patterns on planar surfaces using small silica nanoparticles. Figure 2 shows 1D 50-nm-diameter particle patterns on planar surfaces after removal of the PR lines. In Figs. 2(a) and 2(b), the period is 1000 nm and a single cycle of spin coating was used. Parallel silica particle rows are observed over a large area [Fig. 2(a)]. The particle deposit shows a nonuniformity across the line with thicker deposits at both edges [Fig. 2(b)]. These are due to the high spin speed and relatively small ratio between particle diameter and width of the lines.[19] This "edge-high" effect was also observed for 1D patterns with a 500 nm period. Figures 2(c) and 2(d) show the 1D patterns with a 500 nm period after two cycles of spin coating. With the added deposition, both the thickness is increased and the bending of the deposition at the edges of the lines is reduced. The decreased blocking force from the relatively low PR walls in the second cycle of spin coating resulted in the reduced bending of the deposited particle film at the line edges.

Typical images of 2D patterns of silica nanoparticles deposited on flat surfaces are shown in Figs. 3 and 4. The period for each of these 2D samples was 500 nm. Figure 3(a) shows the PR/wet-i posts after double exposures and before deposition of silica particles. After the deposition of silica nanoparticles and removal of PR/wet-i posts, a continuous cross network of 50-nm-diameter silica particles with periodic voids is formed over a large area [Fig. 3(c)]. Two layers of silica particles were deposited to form this network [Fig. 3(b)]. Figure 4(a) shows the PR/wet-i holes after exposure and development and before the silica particle deposition. A negative PR with a thickness of 500 nm was used. Slight standing wave ridges were formed along hole walls even though an ARC layer was used. Before spin-coating deposition of silica particles, the PR/wet-i layer has a periodic array of ~200-nm-diameter empty holes. After a spin-coating deposition of the smaller ~15-nm-diameter silica nanoparticles C and removal of the photoresist, 2D isolated posts are formed (~15 nm) as shown in Fig. 4(b). (The resolution of the FE-SEM is not sufficient to distinguish the individual particles at this small size.) The posts, formed with a single deposition step, have concave top surface profiles, similar to the 1D deposition shown in Fig. 2(b), due to the high spin speed and relatively small ratio between particle and post diameters. In contrast, uniform isolated post patterns were not successfully obtained using the larger, 50 nm particles. We attribute this to the difficulty of filling these deep PR holes with comparatively "large" particles and the possibility of disturbing the isolated particle clusters during the immersion/calcination treatments.

From the above experiments, we can see that the particle size, pattern feature, and spin program are important factors influencing the final morphology of particle patterns on planar surfaces. On the one hand, it is easier to fill empty spaces

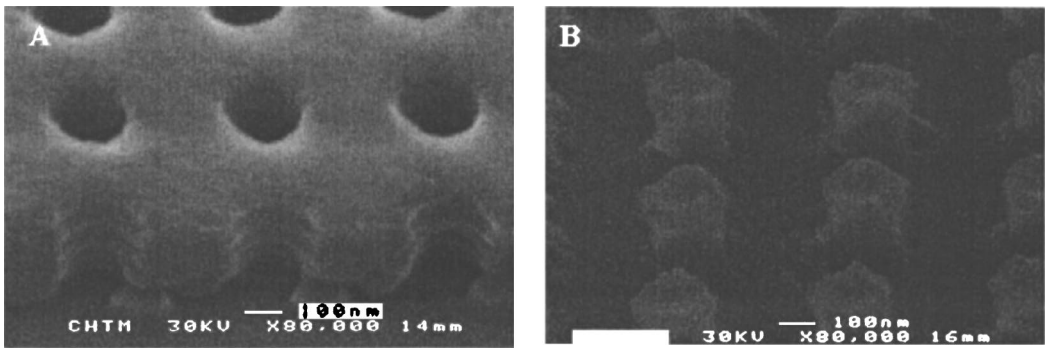


FIG. 4. FE-SEM images (tilted 45° view) of 2D particle arrays with ~15-nm-diameter silica particles on planar surfaces with ~500 nm periods: (a) PR/wet-i holes and (b) isolated particle posts.

in PR patterns with smaller particles. On the other hand, the smaller the ratio of particle diameter to feature pattern size, the more pronounced the "edge-high" effect. Other deposition methods such as layer-by-layer can avoid the "edge-high" effect but with the consumption of a large volume of suspension and need for a long deposition time. Continuous particle networks are easier to form and exhibit fewer defects compared to isolated particle patterns. By applying more cycles of spin coating, the "edge-high" effect can be minimized and complex structures can be obtained.

IV. CONCLUSIONS

1D and 2D patterns of silica nanoparticles were fabricated both on patterned surfaces and on planar surfaces using interferometric lithography and spin-coating. Single linear silica particle chain patterns and isolated 2D particle patterns were easily formed on the patterned surfaces with feature sizes around 100 nm. Silica particle rows, cross networks, and isolated posts with controllable thickness could be formed on flat surface with the smaller silica particles ($\lesssim$50 nm) using this approach. These initial experiments illustrate some of the possibilities of extending patterning by combining top-down and bottom-up processes. Future work will include extensions to smaller particles and smaller lithographic features, as well as the extension to functionalized nanoparticles such as metals and biological species.

ACKNOWLEDGMENTS

The authors would like to thank A. Biswas, Dong Li, Qiming Li, and A. Frauenglass for advice and experimental support. Support for this work was provided by the ARO/MURI in Deep Subwavelength Optical Nanolithography and by the Mission Research Corporation under a subcontract from the Air Force Research Laboratory.

[1]Y. Xia, B. Gates, Y. Yin, and Y. Lu, Adv. Mater. (Weinheim, Ger.) **12**, 693 (2000).
[2]G. A. Ozin and S. M. Yang, Adv. Funct. Mater. **11**, 95 (2001).
[3]M. Glazer, J. Fidanza, G. McGall, and C. Frank, Chem. Mater. **13**, 4773 (2001).
[4]R. Moller, A. Csaki, J. M. Kohler, and W. Fritzsche, Langmuir **17**, 5426 (2001).
[5]F. J. Castano, Y. Hao, M. Hwang, C. A. Ross, B. Vogeli, H. I. Smith, and S. Haratani, Appl. Phys. Lett. **79**, 1504 (2001).
[6]J. O'Brien II, P. Bisong, L. K. Ista, E. M. Rabinovich, A. L. Garcia, S. S. Sibbett, G. P. Lopez, and S. R. J. Brueck, J. Vac. Sci. Technol. B **21**, 2941 (2003).
[7]V. Kitaev and G. A. Ozin, Adv. Mater. (Weinheim, Ger.) **15**, 75 (2003).
[8]X. Wang, C. J. Summers, and Z. L. Wang, Nano Lett. **4**, 423 (2004).
[9]A. van Blaaderen, R. Ruel, and P. Wiltzius, Nature (London) **385**, 321 (1997).
[10]D. Xia, A. Biswas, D. Li, and S. R. J. Brueck, Adv. Mater. (Weinheim, Ger.) **16**, 1427 (2004).
[11]D. Wang and H. Mohwald, Adv. Mater. (Weinheim, Ger.) **16**, 244 (2004).
[12]J. Zhang, A. Alsayed, K. H. Lin, S. Sanyal, F. Zhang, W. J. Pao, V. S. Balagurusamy, P. A. Heiney, and A. G. Yoah, Appl. Phys. Lett. **81**, 3176 (2002).
[13]Y. Cui, M. T. Björk, J. A. Liddle, C. Sönnichsen, B. Boussert, and A. P. Alivisatos, Nano Lett. **4**, 1093 (2004).
[14]H. Zheng, I. Lee, M. F. Rubner, and P. T. Hammond, Adv. Mater. (Weinheim, Ger.) **14**, 569 (2002).
[15]P. Jiang, J. F. Bertone, and V. L. Colvin, Science **291**, 453 (2001).
[16]X. Chen, S. H. Zaidi, S. R. J. Brueck, and D. Devine, J. Vac. Sci. Technol. B **14**, 3339 (1996).
[17]S. H. Im, M. H. Kim, and O. O. Park, Chem. Mater. **15**, 1797 (2003).
[18]Y. Yin, Y. Lu, B. Gates, and Y. Xia, J. Am. Chem. Soc. **123**, 8718 (2001).
[19]D. Xia and S. R. J. Brueck, Nano Lett. **4**, 1295 (2004).