AD-A262 993

ANNUAL REPORT ON ELECTRONICS RESEARCH AT THE UNIVERSITY OF TEXAS AT AUSTIN

APPENDIX - JSEP SPONSORED PUBLICATIONS

NO. 47 - Appendix

For the period May 15, 1992 through February 14, 1993

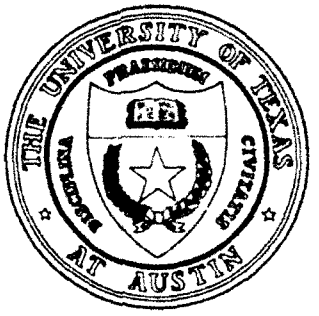
JOINT SERVICES ELECTRONICS PROGRAM

Research Contract AFOSR F49620-92-C-0027

February 14, 1993

APR 0 7 1993

Approved for public release;
Distribution Unlimited



ELECTRONICS RESEARCH CENTER

Bureau of Engineering Research
The University of Texas at Austin
Austin, Texas 78712-1084

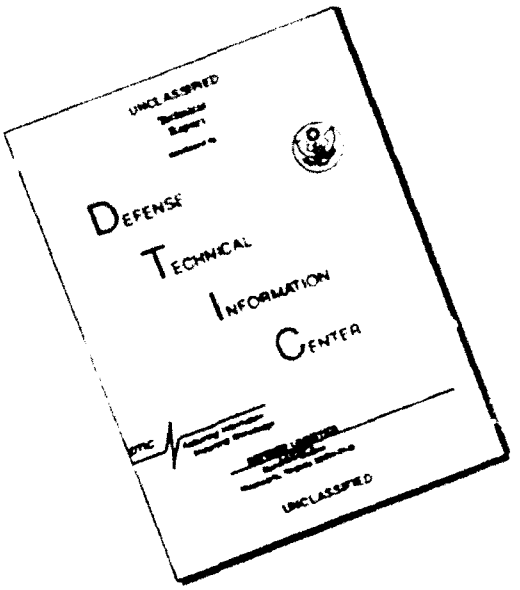
93-07180

The Electronics Research Center at The University of Texas at Austin consists of interdisciplinary laboratories in which graduate faculty members, and Master's and Ph.D. candidates from numerous academic disciplines conduct research. The disciplines represented in this report include information electronics, solid state electronics, and electromagnetics.

The research summarized in this report was supported by the Department of Defense's JOINT SERVICES ELECTRONICS PROGRAM (U. S. Army, U. S. Navy and the U. S. Air Force) through the Research Contract AFOSR F49620-92-C-0027. This program is monitored by the Department of Defense's JSEP Technical Coordinating Committee consisting of representatives from the U. S. Army Research Office, Office of Naval Research and the U. S. Air Force Office of Scientific Research.

Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

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 comments regarding this burden estimate 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
	February 14, 1993	Annual Report 15 May 1992-14 Feb. 1993

4. TITLE AND SUBTITLE
Annual Report on Electronics Research at
The University of Texas at Austin: Appendix - JSEP

5. FUNDING NUMBERS
Research Contract
AFOSR
F49620-92-C-0027

6. AUTHOR(S)
Edward J. Powers and eight faculty associated with the Joint Services Electronics Program at The University of Texas at Austin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas at Austin
Electronics Research Center
Austin, Texas 78712-1084

8. PERFORMING ORGANIZATION REPORT NUMBER
No. 47 - Appendix

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
Building No. 410
Bolling AFB, D.C. 20332

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

This Appendix to the Annual Report contains reprints of publications relating to basic research in information electronics, electromagnetics, and solid-state electronics. The research is described in Annual Report No. 46 and was carried out at The University of Texas at Austin under the auspices of the DoD Joint Services Electronics Program.

14. SUBJECT TERMS
Information electronics, electromagnetics, and solid-state electronics.

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

ANNUAL REPORT ON ELECTRONICS RESEARCH AT THE UNIVERSITY OF TEXAS AT AUSTIN

APPENDIX - JSEP SPONSORED PUBLICATIONS

No. 47 - Appendix

For the period May 15, 1992 through February 14, 1993

JOINT SERVICES ELECTRONICS PROGRAM

Research Contract F49620-92-C-0027

Submitted by Edward J. Powers

on behalf of the Faculty and Staff

of the Electronics Research Center

February 14, 1993

ELECTRONICS RESEARCH CENTER

Bureau of Engineering Research
The University of Texas at Austin
Austin, Texas 78712

Solid-State Electronics Vol. 35, No. 8, pp. 1139–1149, 1992
Printed in Great Britain. All rights reserved

0038-1101/92 $5.00 + 0.00
Copyright © 1992 Pergamon Press Ltd

THE COUPLED HOLE–PHONON SYSTEM AND MINORITY-ELECTRON TRANSPORT IN *p*-GaAs

K. SADRA and B. G. STREETMAN

Department of Electrical and Computer Engineering and Microelectronics Research Center, The University of Texas at Austin, Austin, TX 78712, U.S.A.

(Received 16 July 1991; in revised form 13 January 1992)

Abstract—We present a Monte-Carlo study of minority-electron transport in *p*-GaAs including dynamic screening and plasmon–phonon couping using a phenomenologically damped quasi-nonequilibrium two-band dielectric function. Calculated mode spectra contain the acoustic and optical plasmons of the two-component hole plasma, and suggest that despite the role of the acoustic plasmon, plasma oscillations do not extend to wave vectors as large as the inverse screening length. Furthermore, we find no evidence of the drag effect. For heavy dopings, calculated high-field minority-electron drift velocities are significantly larger than experimental values, indicating that further work is required to achieve a true understanding of minority-electron transport.

INTRODUCTION

During the last few years, minority electron transport has been the subject of a number of experimental and theoretical investigations, not only because of its relevance to bipolar devices, but also because of its importance as an interesting problem in semiclassical transport theory. The first measurements of minority-electron velocity–field characteristics[1], performed on lightly-doped $In_{0.53}Ga_{0.47}As$, suggested that the high-field drift velocity of minority electrons is significantly larger than that of majority electrons. Later, Monte-Carlo simulations of Osman and Grubin[2] indicated that statically screened electron–heavy-hole interactions do not account for sufficient minority-electron energy loss to explain large high-field velocities at low hole concentrations. They did, however, show larger high-field velocities for minority electrons than majority electrons. Sadra *et al.*[3, 4] took statically-screened multiband processes into account and found that although minority-electron energy loss was enhanced, it was still too small to explain the measurements. Subsequent calculations by Saito *et al.*[5] also showed larger minority-electron drift velocities than majority electron values. In 1990, new experimental results[6] showed that, at least in GaAs, the high-field minority-electron drift velocity is in fact lower than the majority-electron value. Taniyama *et al.*[7] performed Monte-Carlo calculations of minority-electron velocity–field characteristics and found reasonable agreement with measurements. However, the agreement was obtained by replacing the heavy-hole effective mass with the light-hole value in the calculation of the screening length. More recently, Lowney and Bennett[8] calculated the minority-electron mobility and found that a plasmon cut-off wave vector $q_c = r_s^{-1}$ was required to obtain agreement with experimental values. Interestingly, this relation differs from that used by Taniyama *et al.*, $q_c = 2\pi r_s^{-1}$. It appears then that there is no consensus on what the proper theory for minority-electron transport is. In fact, there may be disagreement over what constitutes a theoretical explanation of experimental data.

As noted in our earlier work[3], a number of theoretical issues deserve further investigation, among them the effect of dynamic screening and plasmon–phonon coupling in the two-band hole system. Despite the contributions of a number of investigators[5–10], a comprehensive treatment of dynamic screening and plasmon–phonon coupling is still lacking. To us, the use of a cut-off wave vector, an artificial, vaguely defined, yet sharp dividing line between excitations of collective and single-particle character, is fundamentally dissatisfying. Given the inherent vagueness in the position of the cut-off, its value can be adjusted to fit any given measurement. The treatment of these excitations is even less appealing when, as is usually the case, the plasmon is taken to extend to the cut-off wave vector undamped and undispersed. We believe that a theory of minority electron transport should not rely on a cut-off wave vector, particularly if calculated results are sensitive to its value. Fortunately, the frequency- and wave-vector-dependent dielectric function provides a natural framework for removing the cut-off wave vector. The dielectric function gives the complete spectrum of plasma excitations, from collective to single-particle, with dispersion and damping, for all wave vectors. As the wave vector increases and the plasma can no longer sustain collective oscillations, modes gradually disappear in the calculated spectra. The dielectric

DTIC QUALITY INSPECTED 4

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

function is also a natural vehicle for including the often-neglected effects of light holes, among them the rise in the plasma frequency and the appearance of the acoustic plasmon.

In this paper, we report bipolar Monte-Carlo simulations of steady-state minority-electron transport, including dynamic screening and plasmon–phonon coupling given by a phenomenologically damped quasi-nonequilibrium two-band dielectric function. A similar treatment has been recently used in connection with electron transport through an equilibrium hole plasma in heterojunction bipolar transistors[11]. However, it has not been extended to the problem of interest here, that of steady-state minority-electron velocity–field characteristics. We begin with a presentation of the theoretical formulation and the relevant computational issues, and then proceed to describe the results and discuss their implications.

2. THEORETICAL FORMULATION

The Monte-Carlo model employed in this study is identical to that described in our earlier work[4] except for the addition of degeneracy, and the inclusion of dynamic screening and plasmon–phonon coupling in lieu of statically screened multiband carrier–carrier and polar carrier–optical-phonon interactions.† Degeneracy is included in the simulation following the standard technique[12, 13]. As for dynamic screening and plasmon–phonon coupling, our approach is to use the simplest possible formalism, while maintaining frequency and wavevector dependence, as well as some measure of external damping. The treatment is described below.

For a carrier with wave vector k and energy E, the rate of scattering by the coupled phonon–carrier system is given by [14]:

$$P(\mathbf{k}) = \frac{2\pi}{\hbar}\int\frac{d^3q}{(2\pi)^3}\int_{-\infty}^{+\infty}\frac{d(\hbar\omega)}{2\pi}[1-f(\mathbf{k}+\mathbf{q})] \times [n(\omega)+1]\frac{2e^2}{q^2}\,\mathrm{Im}\left[\frac{-1}{\epsilon_T(\mathbf{q},\omega)}\right] \times \mathcal{G}(\mathbf{k},\mathbf{k}+\mathbf{q})\delta[E(\mathbf{k}+\mathbf{q})-E(\mathbf{k})+\hbar\omega], \quad (1)$$

where $\epsilon_T(\mathbf{q},\omega)$ is the total dielectric function of the crystal and in the standard approximation, the overlap factor $\mathcal{G}$ has the form:

$$\mathcal{G}(\mathbf{k},\mathbf{k}') = X(k,k') + Y(k,k')\cos\xi + Z(k,k')\cos^2\xi, \quad (2)$$

and ξ is the angle between $\mathbf{k}$ and $\mathbf{k}'$. We use the usual values for X, Y and Z for electrons and holes[3, 4]. Defining:

$$\Xi(\omega) \equiv \max\left\{\mathrm{Im}\left[\frac{-1}{\epsilon_T(\mathbf{q},\omega)}\right]\right\}, \quad (3)$$

where the maximum is with respect to $\mathbf{q}$, and then transforming to $\mathbf{k}' = \mathbf{k} + \mathbf{q}$, we find an upper bound on the scattering rate given by:

$$P(\mathbf{k}) \leqslant \frac{2\pi}{\hbar}\int\frac{d^3k'}{(2\pi)^3}\int_{-\infty}^{+\infty}\frac{d(\hbar\omega)}{2\pi}[1-f(\mathbf{k}')] \times [n(\omega)+1]\frac{2e^2}{(\mathbf{k}'-\mathbf{k})^2}\Xi(\omega)\mathcal{G}(\mathbf{k},\mathbf{k}')\delta[E_{k'}-E_k+\hbar\omega]. \quad (4)$$

Following the standard technique[12, 13] we omit the degeneracy term $1-f(\mathbf{k}')$ from the integrand and include it directly in the Monte-Carlo simulation. Integrating over $\hbar\omega$, transforming the integral over $\mathbf{k}'$ to spherical coordinates with the polar axis along $\mathbf{k}$, and integrating over the azimuthal angle we have:

$$P(\mathbf{k}) \leqslant \frac{e^2}{2\pi^2\hbar k}\int_0^{+\infty}k'\,dk'\int_{-1}^{+1}d\cos\xi \times\left[\frac{X(k,k')+Y(k,k')\cos\xi+Z(k,k')\cos^2\xi}{k^2+k'^2-2kk'\cos\xi}\right] \times\{n[(E_{k'}-E_k)/\hbar]+1\}\Xi[(E_{k'}-E_k)/\hbar]. \quad (5)$$

Integrating over $\cos\xi$ and transforming back to $\omega = E(k') - E(k)$, we find:

$$P(\mathbf{k}) \leqslant \frac{e^2m^{*\prime}}{2\pi^2\hbar^2k}\int_{-\infty}^{E/\hbar}d\omega[n(\omega)+1]\Xi(\omega) \times[1+2\alpha'(E-\hbar\omega)]\times\Big\{\{X[k,k'(\omega)] + Y[k,k'(\omega)]\Lambda[k,k'(\omega)] + Z[k,k'(\omega)]\Lambda^2[k,k'(\omega)]\}\ln\left|\frac{k+k'(\omega)}{k-k'(\omega)}\right| - \{Y[k,k'(\omega)]+Z[k,k'(\omega)]\Lambda[k,k'(\omega)]\}\Big\}, \quad (6)$$

where

$$k'(\omega) = \sqrt{2m^{*\prime}(E-\hbar\omega)[1+\alpha'(E-\hbar\omega)]}/\hbar. \quad (7)$$

Pre- and post-scattering effective masses and nonparabolicities m^*, $m^{*\prime}$, α and α' allow the application of eqn (6) to electrons in nonparabolic bands and to inter- and intraband scattering of holes.

In the random phase approximation, the total dielectric function of a semiconductor can be written as the sum[15]:

$$\epsilon_T(\mathbf{q},\omega) = \epsilon_L(\mathbf{q},\omega) + \epsilon_e(\mathbf{q},\omega) + \epsilon_h(\mathbf{q},\omega), \quad (8)$$

where ϵ_e and ϵ_h are the contributions of free electrons and holes, respectively. The contribution ϵ_L of the crystal in the absence of free carriers is[16]:

$$\epsilon_L(\mathbf{q},\omega) = \epsilon_\infty\frac{\omega_{LO}^2-\omega^2-i\omega\gamma}{\omega_{TO}^2-\omega^2-i\omega\gamma}, \quad (9)$$

where γ is the inverse of the nonelectronic LO-phonon lifetime, which we take to be 4.0 ps at room temperature[17]. Treating the limit of low minority

†The material parameters used in this study are identical to those used in our earlier work, except for the new parameters reported in this text.

electron concentrations, we ignore the free electron contribution. For the hole contribution, we use[18]

$$\epsilon_h(\mathbf{q},\omega)=\frac{e^2}{q^2}\sum_i\sum_j G_{ij}(\mathbf{k},\mathbf{k}+q) \times\frac{f_i(\mathbf{k})-f_j(\mathbf{k}+\mathbf{q})}{E_i(\mathbf{k}-\mathbf{q})-E_j(\mathbf{k})-\hbar\omega-i\hbar\alpha}, \quad (10)$$

where indices i and j span the heavy- and light-hole bands. Note that the overlap factor G is twice as large as the standard Monte-Carlo overlap factor $\mathcal{G}$ since G and $\mathcal{G}$ are summed and averaged over initial spin states, respectively. For a collisionless plasma, $\alpha\to 0$ and eqn (10) already includes intrinsic or Landau damping. For a plasma of particles undergoing scattering, extrinsic or collisional damping should be included as well. A simple and widely-used approach is to include collisional damping by interpreting α as a phenomenological damping term of the order of the inverse relaxation time. In our calculations, we take α to be 5×10^{12} s^{-1}. Significantly larger values are unphysical; values up to an order of magnitude smaller do not produce appreciably different minority-electron results; and much smaller values are unphysically small and cause severe difficulties in numerical integration of eqn (10).

The theory outlined above is applicable to equilibrium hole plasmas. Here, to obtain some measure of the effect of the nonequilibrium nature of the hole plasma, we take a quasi-nonequilibrium approach: we use the simulated nonequilibrium hole distribution functions in eqn (10). This seemingly benign complication creates inordinate difficulties. Evaluation of eqn (10) for a numerically-known simulated distribution function is extremely difficult, particularly for small α. Furthermore, in spherical bands, the spherical symmetry of equilibrium distribution functions allows us to calculate ϵ_h only as a function of ω and $|\mathbf{q}|$. Under an applied field, however, only cylindrical symmetry remains, necessitating evaluation of ϵ_h as a function of ω and two components of $\mathbf{q}$, say $|\mathbf{q}|$ and Υ, where Υ is the polar angle of $\mathbf{q}$ around the direction of the applied field.† The problem is compounded by the fact that numerical integration is not entirely suited to efficient vector processing. Fortunately, if all additive terms in eqn (10) are separated and regrouped according to the band i of $f_i(\mathbf{k})$ and the integral over $\mathbf{k}$ is transformed to spherical coordinates with $\mathbf{q}$ as the polar axis, the azimuthal integral becomes independent of $|\mathbf{q}|$ and ω and the remaining multiplicative portions of the integrand become independent of the distribution function. A vector double-integration algorithm can then be used to evaluate ϵ_h at a few thousand points $(|\mathbf{q}|,\omega)$ at the same time. Each time the integrands are required at a given point, the azimuthal integral is evaluated only once and the remaining portions of all the integrands are calculated in a vectorized fashion. It is then possible to perform the complete calculation of the dielectric function at an average rate of 150 MFLOPS on a single CRAY YMP processor. Higher FLOP rates are attainable for very smooth distribution functions or at the cost of larger memory. We calculate $\epsilon_h(\mathbf{q},\omega)$ for $q\le 3\times 10^7$ cm^{-1}, $\Upsilon=i\pi/10$ with $i=1,\ldots,10$ and $|\omega|\le\Omega_{max}$ where $\Omega_{max}=6.0\omega_{LO}$. A single-hole simulation is first performed to obtain the steady-state light- and heavy-hole distributions. Periodically, the simulation is interrupted and ϵ_h and the associated upper bound on the hole scattering rate are reevaluated using the converging distribution functions. Only intraband terms are included in the calculation of ϵ_h in the hole simulations to avoid problems associated with, crudely speaking, identifying the scattering holes interchanging bands with the scattered hole[19]. In the statically-screened pair-scattering picture, it is easy to find the single scattering hole and change its state. In the present dynamic treatment, however, the identities of the holes involved in the scattering of the scattered hole are somewhat obscured and further mixed by the damping term. If, in the single-pair picture, most interband processes are viewed as essentially a light hole and a heavy hole exchanging states, the net effect can be regarded as small. It should be clear that ignoring the interband contribution to the dielectric function is not equivalent to ignoring interband hole scattering events. Interband hole scattering is included, as in our previous papers, for ionized-impurity scattering, acoustic deformation-potential scattering, and nonpolar optical-phonon scattering. Also included are interband hole–coupled-mode scattering events in which the scattering of the Monte-Carlo hole does not involve interband transitions of other holes, but involves phonons and/or intraband transitions of other holes. It will be shown below that within our treatment, the details of the hole distribution do not appear to affect minority-electron results appreciably. As such, ignoring the interband contribution to the hole dielectric function in hole simulations is essentially inconsequential.

The final hole distributions obtained from the hole simulations are used to calculate ϵ_h, including both intra- and interband terms, and the upper bound on the minority-electron scattering rate given by eqn (6), for subsequent minority-electron simulations. Interband terms are no longer problematic since a low concentration of minority electrons will not affect the majority hole distribution significantly, making the identification of scattering holes unnecessary. During both electron and hole simulations, the coupled-mode energy is first chosen according to the distribution defined by the integrand of eqn (6). This is done using a search table containing values of

$$f(\Omega)=\int_{-\Omega_{max}}^{\Omega} d\omega[n(\omega)+1]\Xi(\omega) \quad (11)$$

†Note also that the property $\epsilon_h(-\mathbf{q},-\omega)=\epsilon_h^*(\mathbf{q},\omega)$ provides a factor of two reduction in the required computational effort.

and the rejection technique for the rest of the integrand, for which an upper bound is tabulated for all E. The mode energy uniquely determines the value of k. Next, the polar angle ξ is chosen according to the distribution given by the ξ-dependent part of the integrand in eqn (5). The azimuthal angle is uniformly distributed between 0 and 2π. Finally, the exact contribution of this event to the rate integral is evaluated using the dielectric function, and the rejection technique is used to decide whether or not the scattering event occurs.†

3. RESULTS AND DISCUSSION

We begin by considering the major features of the spectrum of the mode strengths[20] $\mathrm{Im}[-1/\kappa(\mathbf{q}, \omega)]$ of the light- and heavy-hole plasma, where $\kappa \equiv \epsilon \epsilon_0$. Figure 1A depicts the mode strength spectrum for the fictitious case of equal light and heavy hole concentrations, $p_h = p_l = 10^{19}\,\mathrm{cm}^{-3}$. For simplicity and clarity, we have used room-temperature Fermi–Dirac distributions and ignored overlap factors and interband excitations. The modes of the system are those of a two-component plasma[21]: a heavily-damped acoustic mode with linear dispersion for small q and an optical mode at the optical plasma frequency $\omega_{OP} = \sqrt{\omega_{ph}^2 + \omega_{pl}^2}$, where $\omega_{ph}^2 = e^2 p_h / \epsilon_\infty m_h^*$ and $\omega_{pl}^2 = e^2 p_l / \epsilon_\infty m_l^*$. Due to the large light-hole concentration, ω_{OP} is sufficiently high to allow the two modes to be widely separated and distinct. As the light-hole concentration decreases (Fig. 1B), ω_{OP} decreases and the two modes approach each other. For a realistic mix of light and heavy holes, $p_l = (m_h^* / m_l^*)^{3/2} p_h$ and $p_h + p_l = 10^{19}\,\mathrm{cm}^{-3}$, further decrease in ω_{OP} together with phenomenological damping results in a joining of the two modes into the characteristic two-lobe spectrum of Fig. 1C. The presence of the acoustic plasmon then serves to extend plasma modes to somewhat higher wave vectors. It should be noted that because of their small effective mass, and despite their relatively small concentration, light holes raise the plasma frequency by some 20% and should therefore not be ignored. Addition of overlap factors does not alter the spectrum of Fig. 1C significantly. Figure 1D depicts the effect of the interband terms, which result in a slight shift of the peak to lower energies, an overall expression of its strength, significant broadening, and the appearance of a high-energy tail due to interband transitions.

The coupled-mode spectra shown in Fig. 2 are obtained from mode simulations for $\mathscr{E} = 12\,\mathrm{kV\,cm}^{-1}$ and $p = 10^{19}\,\mathrm{cm}^{-3}$ as described above. Without interband terms, the mode spectra exhibit a small degree of anisotropy. With the addition of interband terms, however, anisotropic effects are all but overwhelmed. In fact minority-electron simulations performed with $\epsilon(\mathbf{q}, \omega)$ and $\epsilon(-\mathbf{q}, \omega)$ give very similar results, indicating that within this treatment, minority electrons are not seriously affected by the direction in which the holes travel. This rather interesting results is in agreement with the observations of Lovejoy *et al.*[22] concerning the apparent absence of a significant drag effect on the minority-electron mobility, although the absolute value of the mobilities calculated here are higher than the measured values (see below). Low-temperature experiments on modulation-doped quantum wells have shown that if minority electrons are effectively decoupled from phonons and ionized impurities and only interact with majority holes, and if the majority hole distribution is sufficiently anisotropic, minority electrons are drastically affected by the motion of the holes, even to the point of exhibiting negative mobility at low fields[23]. As we will see, at $p = 10^{19}\,\mathrm{cm}^{-3}$, phonon scattering is quite weak. Within the present treatment, strong static screening reduces the importance of ionized-impurity scattering as well. Minority electrons are then coupled mostly to the hole gas, although not as exclusively as the case above. Also, the hole distribution is only weakly anisotropic, even at $12\,\mathrm{kV\,cm}^{-1}$. It remains to be seen whether or not a more sophisticated treatment of the drifting nature of the hole plasma produces different results.

Typical mode strength spectra for $p = 1.5 \times 10^{17}$, 1.5×10^{18} and $10^{19}\,\mathrm{cm}^{-3}$ are shown in Figs 3–5. We begin by considering the long-wavelength region. At $p = 1.5 \times 10^{17}\,\mathrm{cm}^{-3}$, the low-energy plasmon-like mode appears at more or less the optical plasma frequency and the high-energy mode is slightly above ω_{LO}. As the concentration increases to $1.5 \times 10^{18}\,\mathrm{cm}^{-3}$, both the plasmon-like and phonon-like modes rise in energy. At $p = 10^{19}\,\mathrm{cm}^{-3}$, however, the uncoupled optical plasma frequency is already above ω_{LO}, resulting in a phonon-like mode below ω_{LO} and a plasmon-like mode above it. In the small-wavelength limit, screening effects are negligible and the small-wavelength phonon-like mode approaches an unscreened phonon. Figure 3 demonstrates this behavior clearly, indicating some screening of the phonon line at intermediate wavelengths.‡ As the concentration rises, the phonon line is increasingly screened, although it still approaches the unscreened case for sufficiently large q. It should be noted that plots of $\mathrm{Im}[-1/\kappa(\mathbf{q}, \omega)]$ do not give a direct indication of the effectiveness of various modes in electron scattering. As eqn (1) indicates, the long-range nature of the Coulomb term q^{-2} results in the disproportionate importance of the small-q region. (Equally important are the effects of the boson

†We have performed simulations in which the free-carrier contribution to the dielectric function is ignored. The resulting scattering rates, drift velocities, etc. are in excellent agreement with those obtained in simulations using the traditional phonon scattering treatment.

‡It is imperative to view the phonon line with sufficient energy resolution as to avoid erroneous conclusions concerning its behavior.

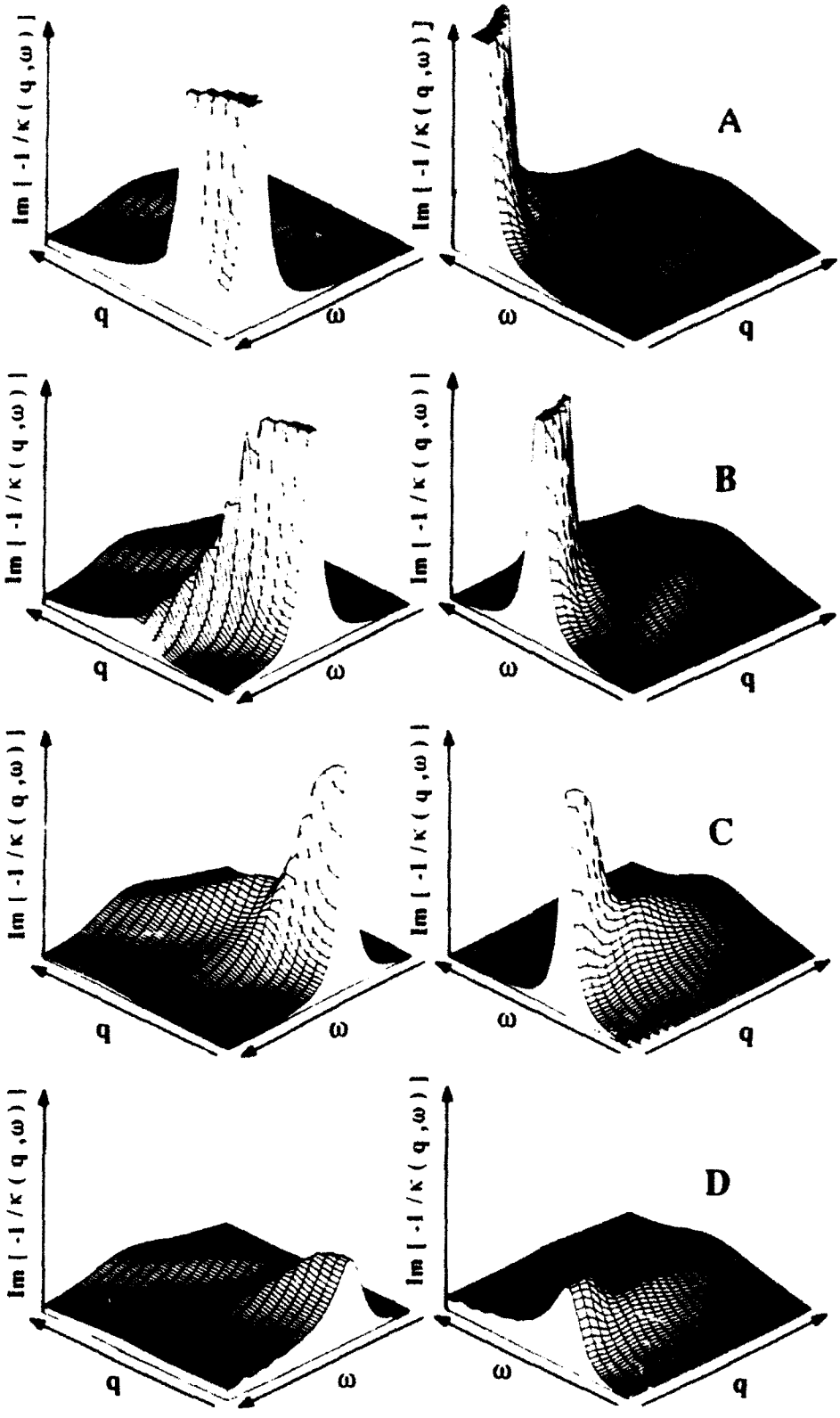


Fig. 1. Plasma mode strengths for $p_h = p_l = 10^{19}\,cm^{-3}$ (A); $p_h = 10^{19}\,cm^{-3}$ and $p_l = 3 \times 10^{18}\,cm^{-3}$ (B); and $p_l = (m_l^*/m_h^*)^{3/2} p_h$ and $p_h + p_l = 10^{19}\,cm^{-3}$ [(C and D)]. (A), (B) and (C) do not overlap factors or interband terms and $\alpha = 10^{13}\,s^{-1}$. The axes cover the ranges $\omega = 0 - 160$ meV, $q = 0 - 2 \times 10^6\,cm^{-1}$ and $Im[-1/\kappa] = 0 - 0.4$.

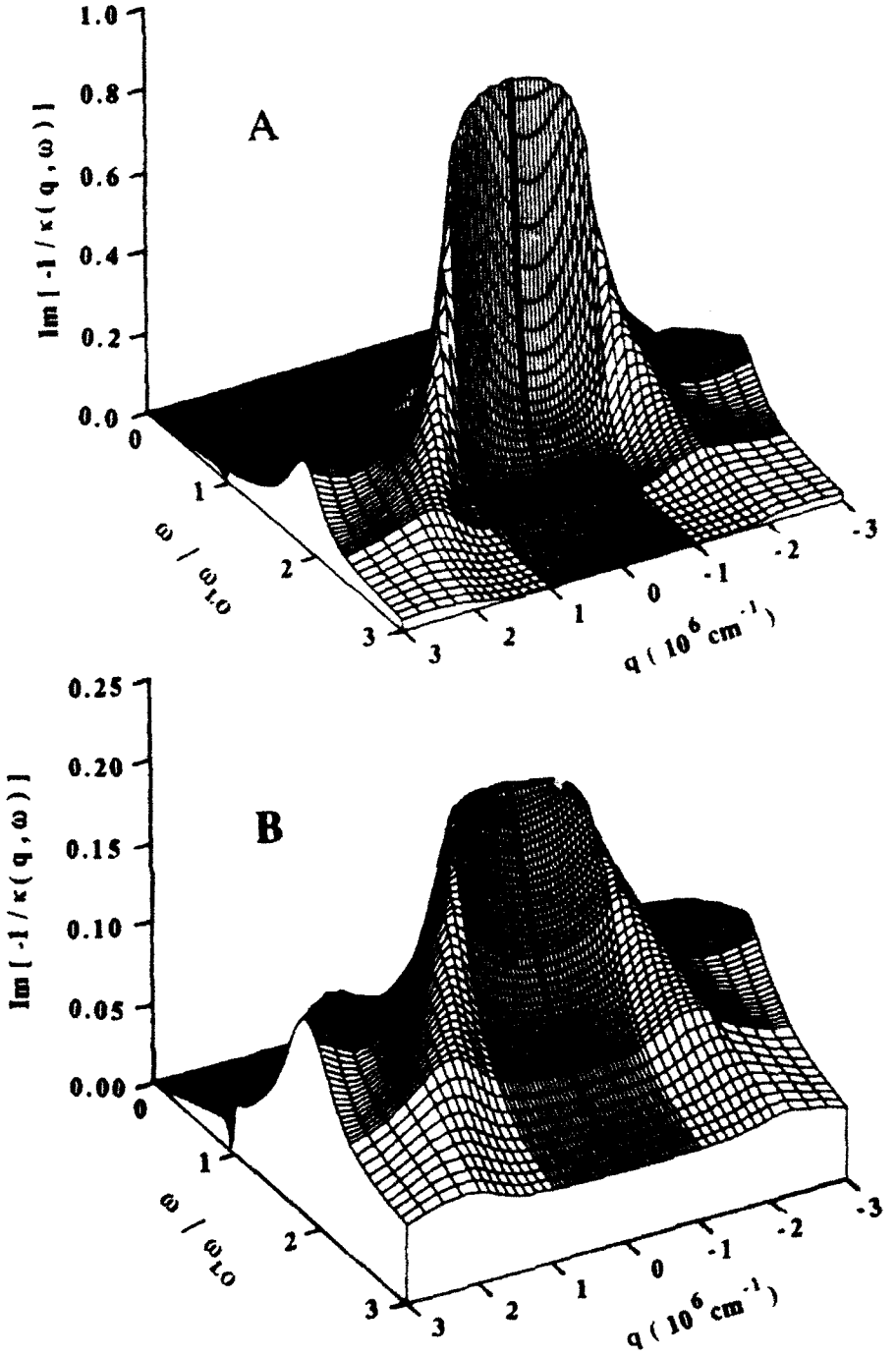


Fig. 2. The coupled-mode strength spectrum for $p = 10^{19}\,cm^{-3}$ and $\mathcal{E} = 12\,kV\,cm^{-1}$ without (A) and with (B) interband contributions. Positive and negative q-values are parallel and antiparallel to the applied field, respectively

population favoring small-ω events and the requirements of energy and momentum conservation.) At $p = 10^{19}\,cm^{-3}$, then, phonon scattering is expected to be essentially inoperative.

In Fig. 6, we have plotted the simulated minority-electron scattering rate density, the rate of minority-electron scattering by the coupled hole–phonon system per unit mode energy. Major features of the mode spectra of Figs 3–5 also appear in Fig. 6. The scattering rate density exhibits distinct features corresponding to the plasmon-like modes, which increase in energy with the concentration. The effect of the phonon-like modes diminishes as the concentration is increased. Due to the boson population, there are many scattering events at small ω. Figure 7 depicts the minority-electron energy loss rate density, the rate

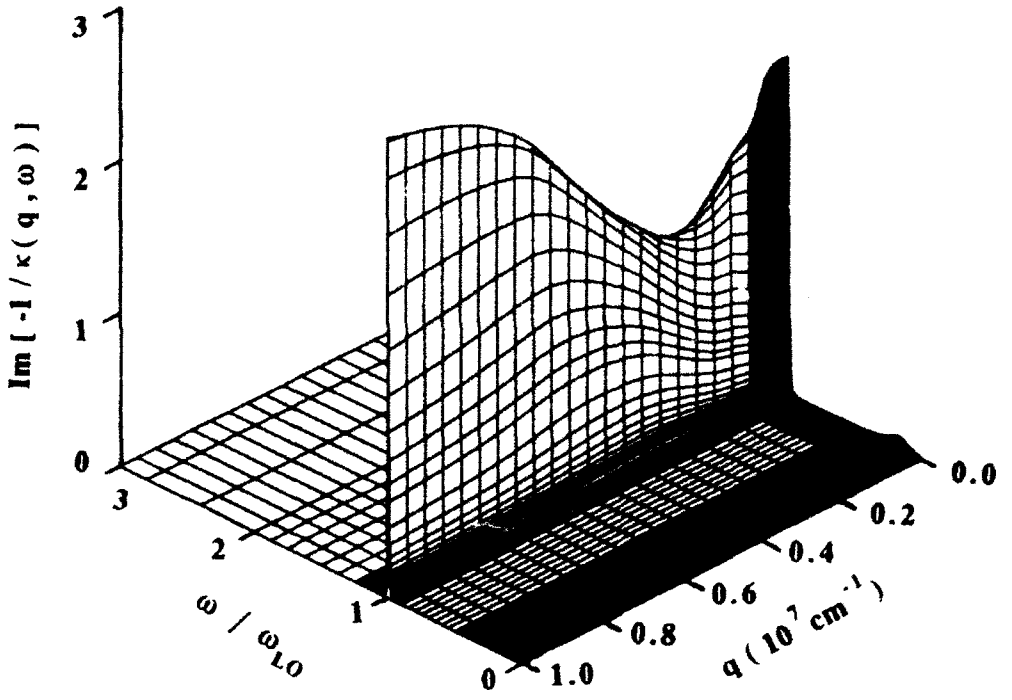


Fig. 3. The coupled-mode strength spectrum for $p = 1.5 \times 10^{17}\,\mathrm{cm}^{-3}$, $\mathcal{E} = 12\,\mathrm{kV\,cm}^{-1}$ and $\Upsilon = \pi$.

of minority-electron energy loss to the coupled hole–phonon system per unit mode energy. Due to the small energy involved in a small-ω event, as well as the relative unimportance of spontaneous emission, the net energy loss rate density for small ω is quite well behaved.

Theoretical and experimental velocity-field characteristics of minority electrons are shown in Figs 8–10. At $p = 10^{17}\,\mathrm{cm}^{-3}$, disagreement between theoretical and experimental velocities does not exceed 20%, which is comparable to the general range of experimental uncertainties. At $p = 10^{18}$ and $10^{19}\,\mathrm{cm}^{-3}$,

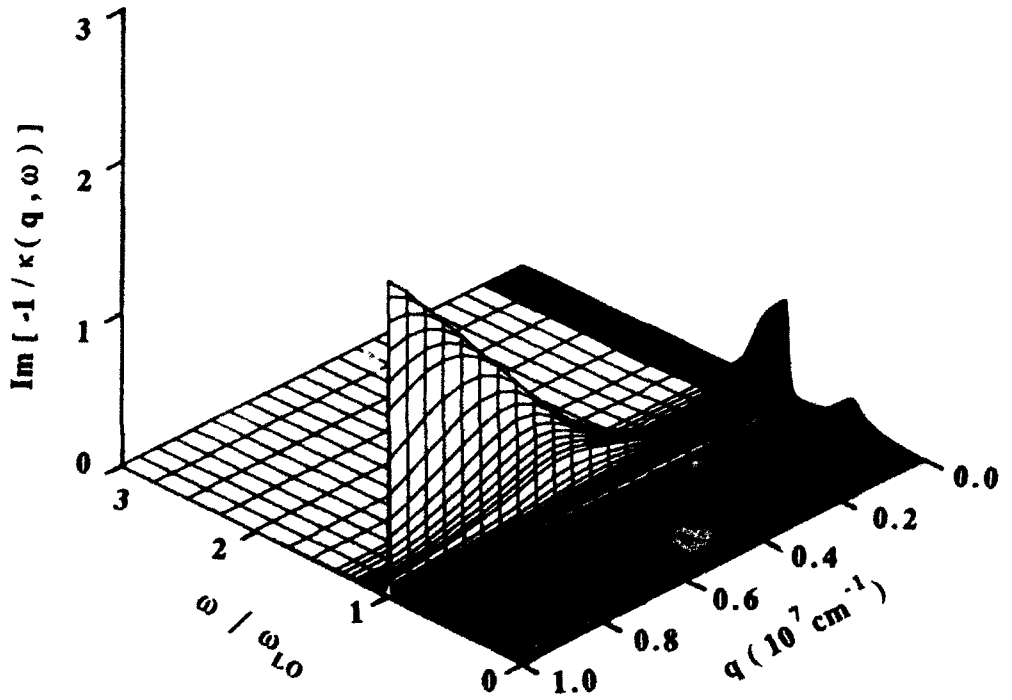


Fig. 4. The coupled-mode strength spectrum for $p = 1.5 \times 10^{18}\,\mathrm{cm}^{-3}$, $\mathcal{E} = 12\,\mathrm{kV\,cm}^{-1}$ and $\Upsilon = \pi$.

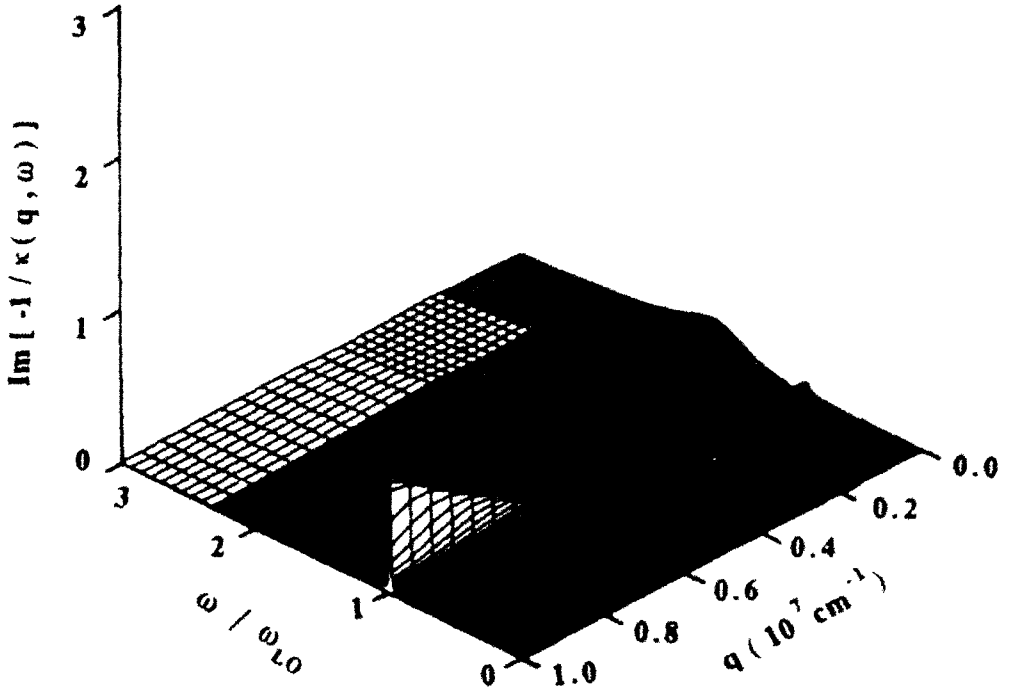


Fig. 5. The coupled-mode strength spectrum for $p = 10^{19}\,cm^{-3}$, $\mathcal{E} = 12\,kV\,cm^{-1}$ and $\Gamma = \pi$

however, theoretical characteristics are in striking disagreement with experimental data, exhibiting significantly larger high-field drift velocities. We focus on these higher hole concentrations because they provide more stringent tests of the ability of the theory to describe the interactions of minority electrons with holes. It is reasonable to expect that once the theory is able to explain the experimental data at the higher hole concentrations, an explanation of the data at $p = 10^{17}\,cm^{-3}$ would be naturally afforded. Although experimental problems such as the likely presence of large quantities of interstitial Be at $p = 10^{19}\,cm^{-3}$ may be invoked to explain away some of the discrepancy, it is clear that the present treatment is inadequate. The large calculated velocities are a consequence of strong energy relaxation, preventing the heating of electrons and their transfer to upper valleys, coupled with insufficient forward momentum relaxation, which allows the development of a highly anisotropic distribution. A comparison of the theoretical and experimental drift velocities of Figs 9 and 10 suggests that the present treatment results in significantly smaller forward momentum relaxation than what is required to explain the measurements. Furthermore, a comparison of the peak fields suggests that the calculated minority-electron energy loss is perhaps somewhat too large.

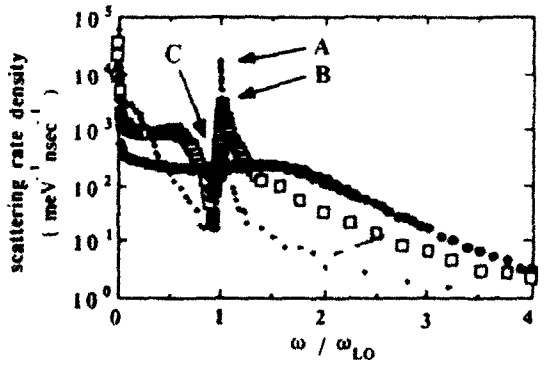


Fig. 6. The simulated rate of scattering of Γ-valley minority electrons by the coupled hole–phonon system per unit mode energy for $p = 1.5 \times 10^{17}\,cm^{-3}$ (+, phonon peak labeled A), $1.5 \times 10^{18}\,cm^{-3}$ (□, phonon peak labeled B) and $10^{19}\,cm^{-3}$ (●, phonon peak labeled C).

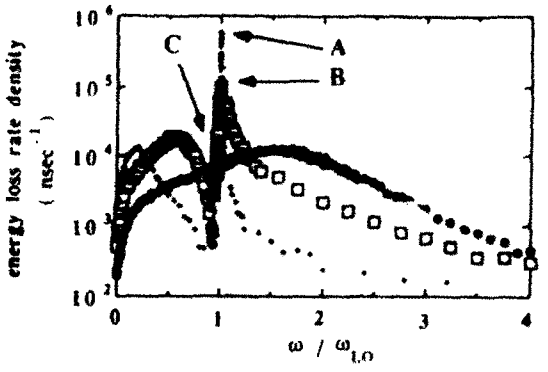


Fig. 7. The simulated rate of energy loss by Γ-valley minority electrons to the coupled hole–phonon system per unit mode energy $p = 1.5 \times 10^{17}\,cm^{-3}$ (+, phonon peak labeled A), $1.5 \times 10^{18}\,cm^{-3}$ (□, phonon peak labeled B) and $10^{19}\,cm^{-3}$ (●, phonon peak labeled C).

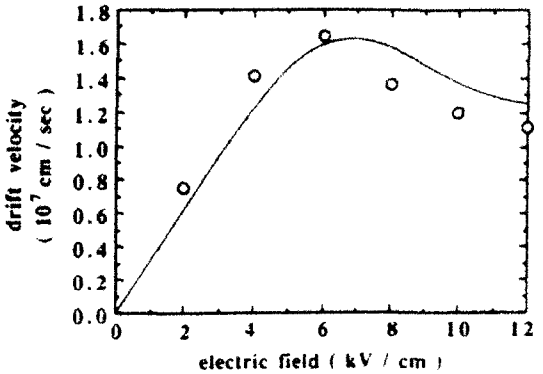


Fig. 8. Velocity–field characteristics of minority electrons for $p = N_A^- = 1.5 \times 10^{17}$ cm^{-3}. The line is experimental data from Ref. [6] and calculated points are marked.

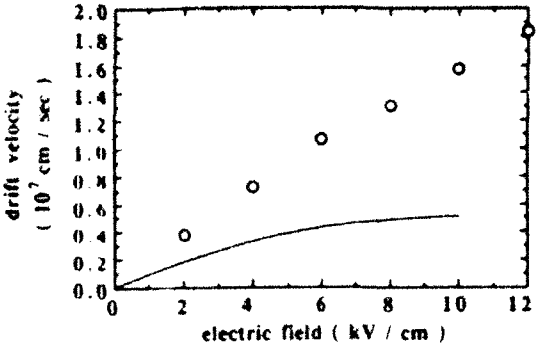


Fig. 10. Velocity field characteristics of minority electrons for $p = N_A^- = 10^{19}$ cm^{-3}. The line is experimental data from Ref. [9] and calculated points are marked.

Our heavy-doping results are essentially similar to those of Saito *et al.*[5], and differ from those reported by Taniyama *et al.*[7]. Saito *et al.* used a somewhat simpler treatment and calculated high-field drift velocities far in excess of experimental results. Taniyama *et al.*[7] argued that the static screening length of the two-component p-type system is not equal to the heavy-hole value alone. They replaced the heavy-hole effective mass with the light-hole value to calculate the proper screening length and obtain agreement with experiment. We believe that this procedure is unsatisfactory. The correct static screening length of the two-component system for arbitrary degeneracy is quite easy to calculate, and has been used by us in this work and in previous publications[3]. Both light- and heavy-hole contributions are included, and the interband contribution is zero due to the orthogonality of same-k states in the two bands. The heavy-hole contribution is dominant, indicating effective static screening by heavy holes. Therefore, although the light-hole contribution is not included in most papers and the treatment is not always extended to arbitrary degeneracy, all workers employ essentially the same theory of static screening. Furthermore, the screened potential must result in light- and heavy-hole phase shifts that, at least in some approximation, satisfy the Friedel sum rule. Not every screening length will satisfy this constraint.

In Fig. 11, we have plotted available experimental values for the minority-electron mobility, together with values calculated by us and by Lowney and Bennett[8]. Our calculated values are in agreement with the results of Lowney and Bennett for the plasma cut-off wave vector $q_c = 0.5\, r_s^{-1}$. To the extent that mobility values alone can be used as indicators, our treatment, which includes dispersion and intrinsic and extrinsic damping, suggests that plasma modes extend only so far in wave vector as to correspond to an effective cut-off at $0.5\, r_s^{-1}$. Accordingly, although extending some effective cut-off to r_s^{-1} may be eventually borne out by an improved treatment of the dielectric function, it is not supported by the present treatment. This is not necessarily surprising. The argument that a plasma cannot support oscillations with wavelengths much shorter than the screening length does not necessarily mean that it does support undamped and undispersed oscillations all the way down to the screening length. Extrinsic damping can destroy plasma oscillations at shorter

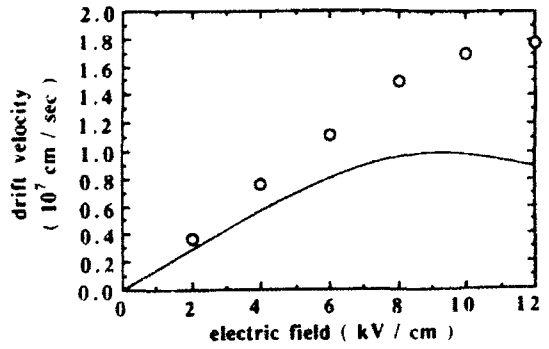


Fig. 9. Velocity–field characteristics of minority electrons for $p = N_A^- = 1.5 \times 10^{18}$ cm^{-3}. The line is experimental data from Ref. [6] and calculated points are marked.

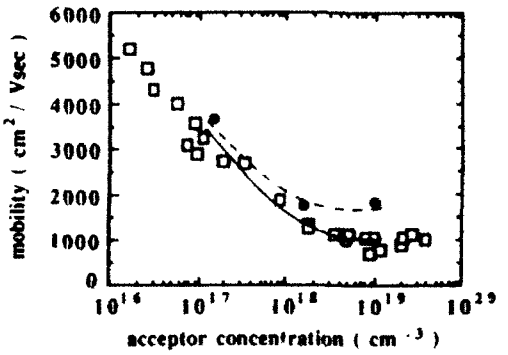


Fig. 11. Theoretical and experimental values of the minority-electron mobility. Empty squares are experimental points[24], filled circles are our calculated results and dashed and solid lines are Lowney and Bennett's value for $q_c = 0.5\, r_s^{-1}$ and r_s^{-1}, respectively[8].

wave vectors. Even in the collisionless case, the statement $q_r \approx r^{-1}$ may not rest on firm ground. In a p-type plasma, the screening length is essentially determined by heavy holes, whereas light holes and interband transitions provide additional channels for intrinsic damping at shorter wave vectors. We believe that these issues are strong incentives for pursuing calculations of the complete mode spectra using progressively improved dielectric functions.

We now turn to a critical evaluation of the treatment used in this work, and attempt to chart the road for future improvements. When all electrons are confined to the Γ-valley, as in the $N_A = 10^{19}\,\mathrm{cm}^{-3}$ case, the drift velocity is affected by two major processes: scattering from the coupled hole–phonon system and ionized-impurity scattering. As noted above, our strategy has been to use the simplest possible treatment for the former, while maintaining full frequency and wave vector dependence. An important limitation of this formulation is the phenomenological treatment of the effect of scattering on the dielectric function. Since an accurate and well-grounded analysis of the damping of plasma modes and the effect of scattering on screening may well be central to the problem at hand, we believe that one should proceed directly to a true relaxation-time treatment of the dielectric function, or perhaps go beyond the relaxation-time approximation altogether. Raman measurements can serve as a valuable guide for progress in this area. Equally important is the development of a nonperturbative theory for the interactions of carriers with the dynamically-screened hole–phonon system. More accurate hole overlap factors could also be beneficial. As the hole concentration increases to $10^{19}\,\mathrm{cm}^{-3}$, scattering events with larger energy and momentum exchange become prominent. Smaller hole overlap factors could reduce the relevant contributions to the dielectric function. An apparently unappreciated property of Kane's overlap factors[25] is their troubling behavior as $\mathbf{q} \to 0$ with nonzero ξ. This property could have an adverse effect on the important long wavelength regions of ϵ_h. Finally, our treatment attempts to gauge the importance of the nonequilibrium nature of the hole plasma by simply inserting nonequilibrium distributions into an equilibrium theory. Whereas this is a reasonable first step, it is clearly not the ultimate objective. Although experimental evidence for the apparent absence of a significant drag effect[22] suggests that the nonequilibrium nature of the hole plasma may not play a dominant role in determining the minority-electron mobility, a fully nonequilibrium treatment should be developed in due course. It is also known that as the impurity concentration increases, a number of complications arise in the theory of ionized-impurity scattering[26]. As increasingly sophisticated models for minority-carrier transport at high dopings are developed, such complications should be kept in sight.

In summary, we have calculated velocity–field characteristics of minority electrons in p^+-GaAs including dynamic screening and plasmon–phonon coupling using a phenomenologically damped quasi-nonequilibrium two-band dielectric function. Our results suggest that under conditions of interest here, the drifting nature of the hole gas does not affect the minority-electron drift velocity seriously. Moreover, calculated velocities are significantly higher than experimental values, indicating that larger forward momentum relaxation is required to account for the experimental data.

Acknowledgements—We would like to thank H. Taniyama, D. O. Yevick and K. Wan for providing us with preprints, and M. E. Kim, K. Hess, C. W. Horton, C. M. Maziar, F. A. Riddoch and D. S. Tang for encouraging conversations and useful discussions. This work was supported by the Joint Services Electronics Program through Contract No. AFOSR F 492620-89-C-0044. The computing resources for this work were provided by the University of Texas System Center for High Performance Computing.

REFERENCES

1. J. Degani, R. F. Leheny, R. E. Nahory and J. P. Heritage, *Appl. Phys. Lett.* **39**, 569 (1981).
2. M. A. Osman and H. L. Grubin, *Appl. Phys. Lett.* **51**, 1812 (1987).
3. K. Sadra, C. M. Maziar, B. G. Streetman and D. S. Tang, *J. appl. Phys.* **66**, 4791 (1989).
4. K. Sadra, C. M. Maziar and B. G. Streetman, *J. appl. Phys.* **66**, 2020 (1989).
5. K. Saito, T. Yamada, T. Akatsuka, T. Fukamachi, E. Tokimitsu, M. Konagai and K. Takahashi, *Jpn. J. appl. Phys.* **28**, L2081 (1989).
6. T. Furuta, H. Taniyama and M. Tomizawa, *J. appl. Phys.* **67**, 293 (1990).
7. H. Taniyama, M. Tomizawa, T. Furuta and A. Yoshii, *J. appl. Phys.* **68**, 621 (1990).
8. J. R. Lowney and H. S. Bennett, *J. appl. Phys.* **69**, 7102 (1991).
9. T. Furuta and M. Tomizawa, *Appl. Phys. Lett.* **56**, 824 (1990).
10. R. P. Joshi, A. M. Kriman and D. K. Ferry, *J. appl. Phys.* **68**, 4322 (1990).
11. P. H. Beton and A. F. J. Levi, *Appl. Phys. Lett.* **55**, 250 (1990).
12. S. Bosi and C. Jacoboni, *J. Phys. C* **9**, 315 (1976).
13. P. Lugli and D. K. Ferry, *IEEE Trans. Electron Devices* **ED-32**, 2431 (1985). Note that for the treatment of degeneracy to make sense, the relevant sampling condition must be met.
14. M. E. Kim, A. Das and S. D. Senturia, *Phys. Rev. B* **18**, 6890 (1978). Note the typo in eqn (12).
15. B. B. Varga, *Phys. Rev.* **137**, A1896 (1965).
16. See, for example, P. Bruech, *Phonons: Theory and Experiments II*. Springer-Verlag, Berlin (1982).
17. The actual value is probably somewhat smaller. See P. Lugli and S. M. Goodnick, *SPIE Proc.* **942**, 10 (1988); and J. S. Blakemore, *J. appl. Phys.* **53**, R123 (1982); as well as J. A. Kash, J. C. Tsang and J. M. Hvam, *Phys. Rev. Lett.* **54**, 2151 (1985); and J. C. Tsang, J. A. Kash and S. S. Jha, *Physica B*, **134**, 184 (1985) for the 77 K value; and P. G. Klemens, *Phys. Rev.* **148**, 845 (1966) for the temperature dependence.
18. H. Ehrenreich and M. H. Cohen, *Phys. Rev.* **115**, 786 (1959).

19. For a clarification of the terminology, see K. Sadra, C. M. Maziar and B. G. Streetman, *Proc. IEE Pt. I* **135**, 119 (1988).
20. B. K. Ridley, *Quantum Processes in Semiconductors*. Oxford University Press, New York (1982).
21. See, for example, P. M. Platzman and P. A. Wolff, *Waves and Interactions in Solid State Plasmas*. *Solid State Physics*, Suppl. 13. Academic Press, New York (1973).
22. M. L. Lovejoy, B. M. Keyes, M. E. Klausmeier-Brown, M. R. Melloch, R. K. Ahrenkiel and M. S. Lundstrom, *Jpn. J. Appl. Phys.* **30**, L135 (1991).
23. R. A. Hopfel, J. Shah, P. A. Wolff and A. C. Gossard, *Phys. Rev. Lett.* **56**, 2736 (1986).
24. See Ref. [8] and references therein.
25. J. D. Wiley, *Phys. Rev. B* **4**, 2485 (1971).
26. For a review, see D. Chattopadhyay and H. J. Queisser, *Rev. Mod. Phys.* **53**, 745 (1981).

Measurements of abrupt transitions in III–V compounds and heterostructures

B. G. Streetman and Y. C. Shih
Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712

(Received 20 February 1991; accepted 6 May 1991)

We review measurements of abrupt transitions in III–V compound semiconductors and heterostructures, including abrupt interfaces as found in quantum wells (QWs) and high electron mobility transistors (HEMTs), and abrupt doping profiles such as delta doping. The interface quality of QW and HEMT structures can be studied using photoluminescence (PL). Several methods are employed to characterize the abrupt doping profiles of delta-doped structures. Hall and Shubnikov–de Haas measurements are used to extract the 2D carrier concentrations in detail. C–V profiling is employed to study the carrier profiles normal to the dopant plane. The actual dopant distribution can be obtained by secondary ion mass spectrometry or by combining C–V profiling with theoretical calculations. The resolution of these measuring techniques will be discussed.

I. INTRODUCTION

Abrupt transitions in III–V compounds, including abrupt heterojunctions and doping profiles, have been of increasing importance in high speed electronic and optical device fabrication and design. The rapid development of advanced semiconductor growth technologies, such as molecular beam epitaxy (MBE) and metalorganic chemical vapor deposition (MOCVD), has greatly improved the crystal quality and reliability of materials. This leads to nearly ideal heterostructures with abrupt interfaces. The interface roughness plays a dominating role in determining the performance of heterostructure electronic and optoelectronic devices. To investigate the properties of the abrupt interfaces, photoluminescence (PL) at low temperatures, typically liquid helium or liquid nitrogen temperature, is a sensitive and powerful tool.[1,2] The PL spectra can provide information about the interface quality as well as two-dimensional electron (hole) gas (2DEG).

Extremely abrupt doping in III–V semiconductors has been demonstrated by delta-doping, also called atomic layer doping. Abrupt doping with donors such as Ge,[3] Si,[4,5] and Se (Ref. 6) in GaAs, and Si (Ref. 7) in InGaAs, has been studied, and p-type doping using C (Ref. 8) and Be (Ref. 9) in GaAs has been reported. The physics and applications of delta doping in III–V compounds and heterostructures are very interesting and have been reviewed in several recent articles.[10–13] In this doping technique the dopants are ideally deposited onto a single plane by interrupting epitaxial growth. The dopants are subject to diffusion and segregation during growth,[14,15] and therefore it is important to subsequently measure the doping distribution. Hall measurements of delta-doped semiconductors have shown high two-dimensional carrier concentrations[16] and enhanced mobilities. Shubnikov–de Haas (SdH) measurements assert the existence of a two-dimensional electron gas in the delta-doping layer,[17] and are used to calculate the subband carrier concentrations associated with the 2D system.[18] Along the growth direction, the carrier profiles are extracted from capacitance–voltage measurements, which is C–V profiling. The C–V profile widths of the delta doping are much less than the Debye length, which is the resolution of conventional C–V profiling. This suggests that electrolyte C–V profiling, with resolution typically larger than 50 Å, is not useful for such abrupt doping profiles. The actual dopant profile widths are of the same order as the carrier profiles, as verified by secondary ion mass spectroscopy (SIMS).[8] However, to quantitatively evaluate the actual doping distributions, theoretical calculations combined with measurements such as C–V profiling, Raman spectroscopy,[19] and SdH measurements, are still under study. It is believed that the dopants are confined in the regime where quantum size effects occur.

II. ABRUPT INTERFACES

III–V compound semiconductors provide a variety of lattice-matched heterostructures with abrupt interfaces, and a range of band discontinuities. The most widely studied ternary and quaternary alloys are in the InAlGaAs and InGaAsP systems, and most recently the InGaAsSb system. This variety of materials results in a wealth of applications in high speed electrical and optical devices, such as high electron mobility transistors (HEMTs) and quantum well lasers. The layer abruptness of such heterostructures play an important role in determining their performance. For example, in HEMT structures, imperfection of the interface along which the 2DEG conducts, reduces the mobility of the 2DEG due to the interface scattering of carriers. For optical devices employing quantum well structures, interface roughness as well as layer mixing results in optical spectral linewidth broadening.

Photoluminescence (PL) is a nondestructive technique to measure radiative transitions in semiconductors, and is especially useful in compound semiconductors. For QW structures, the PL spectra can extract important information such as recombination mechanisms,[20] interface roughness,[21] and the impurities in the QWs.[22,23] The lowest subband tran-

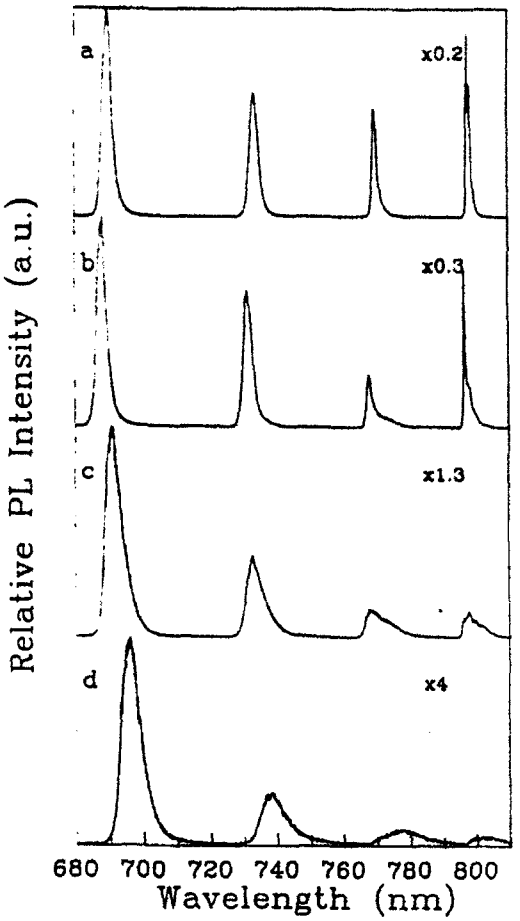


FIG. 1. PL spectra from four quantum wells with different thickness. Samples were grown by MBE at As cracking temperatures of increasing order of (a), (b), (c), and (d), corresponding to the increase of AsO. Significant degradation of the PL spectra is observed in the presence of AsO while the AsO cannot be detected by RGA (Ref. 25).

sition of QWs corresponds to a distinct peak distinguished from the bulk material signals.[19] This peak can therefore accurately determine the thickness of the quantum well.[24] The linewidth of the QW peak is a sensitive index of the interface roughness. For AlGaAs/GaAs QWs grown by molecular beam epitaxy, our group has demonstrated strong dependence of PL spectra on the QW degradation induced by AsO from the As_2 cracking source,[25] while the AsO cannot be detected by residual gas analysis (RGA). Figure 1 shows broadening of linewidth and reduction of intensity in PL spectra, as a function of AsO content. In electrical devices using HEMT or MDQW (modulation-doping quantum well) structures, PL spectra can provide information of the two-dimensional electron (hole) gas.[26] Our group has illustrated the relation between the PL spectral linewidth and the 2DEG densities for AlGaAs/GaAs (Ref. 27) and pseudomorphic AlGaAs/InGaAs/GaAs MDQWs.[28] The FWHM (full width at half maximum) of the 77-K PL spectra can accurately measure the 2DEG densities, as illustrated in Fig. 2(a). This is due to the linewidth broadening induced by the thermal motions of carriers. This technique provides a nondestructive method to extract the carrier concentrations, as verified by Hall measurement. The ratio of linewidths at 77 and 4.2 K is linearly dependent on the Hall mobilities, and indicates the crystal quality since the 4-K PL spectra is sensitive to the interface scattering while the 77 K linewidth is not. The 4.2-K PL spectra can also predict the Fermi energy position in the MDQWs.

III. DELTA DOPING

Delta doping in compound semiconductors produces the narrowest doping profile reported so far, and has realized novel electrical and optical devices, such as δ-FETs (delta-doped field effect transistors),[29,30] and saw-tooth superlattice lasers.[31] Delta doping has exhibited 2D carrier concentration much higher than that achieved by uniform doping. For Si-doped GaAs, delta doping has demonstrated[5] 2D electron densities $> 1\times10^{13}$ cm^{-2}, as compared to the saturation concentration about 1.7×10^{11} cm^{-2} in uniformly doped GaAs. The strong localization of carriers about the dopant plane, as shown in Fig. 3, gives a strong screening effect of the ionized impurity scattering to the free carriers, resulting in enhanced carrier mobilities. The use of delta doping in modulation doping devices[32,33] such as MODFETs has demonstrated improved transconductance and therefore current driving capability due to high density 2DEG, as verified by experiments and theories.[34–36] Delta doping can also be employed in submicron devices to improve short channel effects. It has been shown in conventional HEMT devices that the carrier mobilities are drastically reduced at high electric field,[37] setting a lower limit on the channel length. Devices using the delta-doping plane as a conduction channel such as δ-FETs,[38] however, suffer much less reduction of carrier mobilities at high field. The δ-FETs show better performance than the conventional HEMT, in submicron channel length regime. Recently, delta doping in quantum wells[39] has been employed in FET device fabrication[40] since the quantum wells provide additional carrier confinement, yielding improved breakdown voltages.

IV. CARRIER CONCENTRATIONS

The 2D carrier concentrations in the delta-doping layer are of great importance and provide the current flow used for devices. The understanding of the 2D carrier system can be of great help in optimizing device design and fabrication. Hall effect measurement is by far the most common technique to measure free-carrier densities and mobilities in semiconductors at weak magnetic fields,[41] typically of the order of kilogauss. Measured Hall concentrations of about 10^{13} cm^{-2} have been reported for Si delta doping in GaAs,[5] as compared to Hall concentrations of $< 1\times10^{12}$ cm^{-2} for AlGaAs/GaAs HEMTs. The Hall mobilities of Si delta doping in GaAs are enhanced compared to that of equivalent uniform doping, due to strong screening of Coulombic potential induced by the ionized Si dopants. Moreover, the low-temperature Hall concentrations of delta-doped GaAs show dependence on the applied magnetic field.[42] This results from the fact that in the delta-doped samples there are several conduction channels with different carrier mobilities,[43]

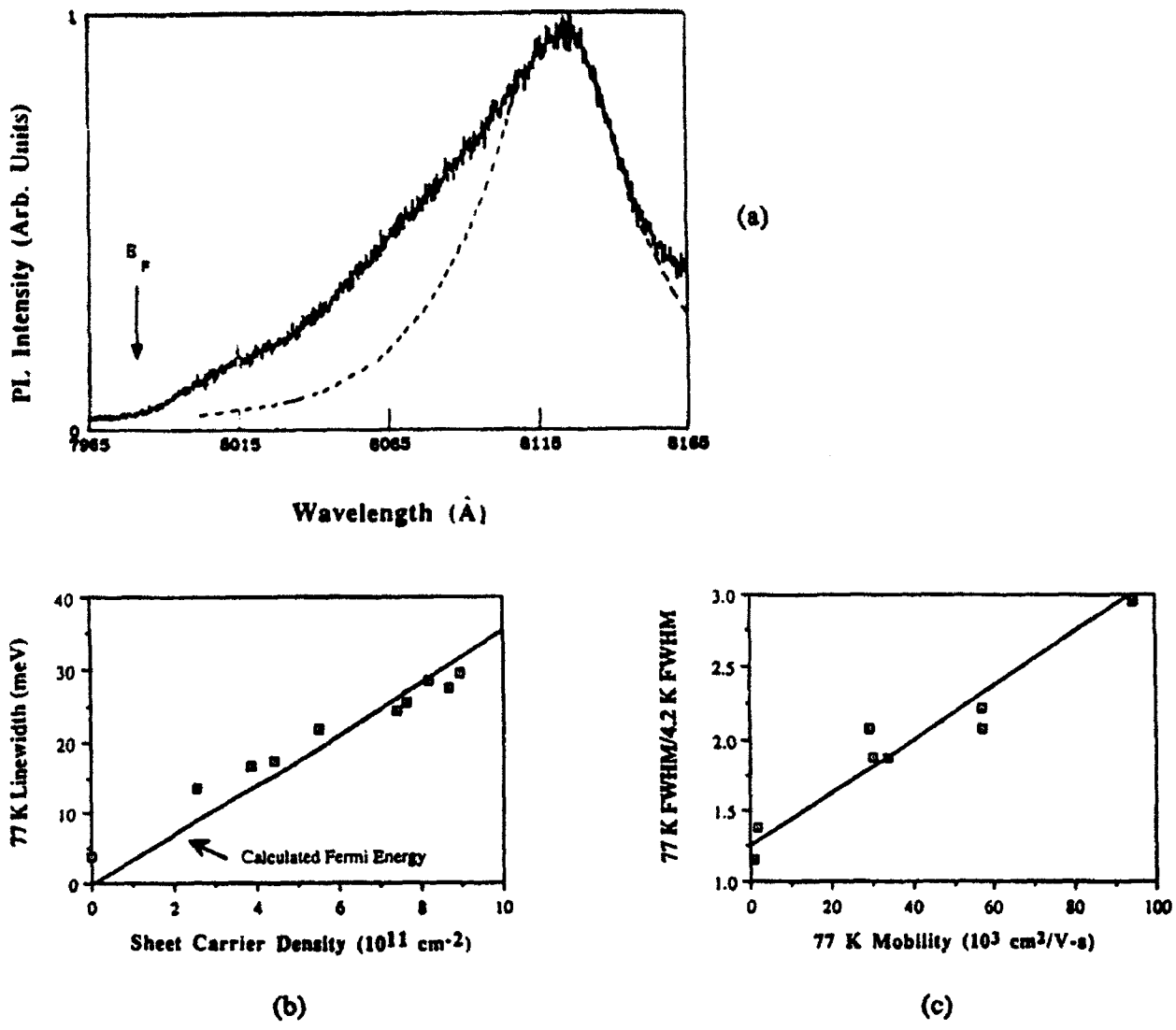


FIG. 2. (a) A PL spectrum of a MDQW from which the Fermi energy position can be extracted, (b) 77-K PL linewidths as a function of sheet carrier density measured by Hall measurements, showing a good linear dependence, (c) ratio of linewidths at 77 and 4.2 K as a function of 77 K Hall mobility (Ref. 27).

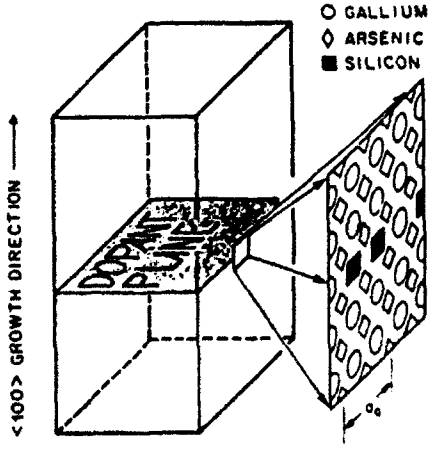


FIG. 3. Schematic diagram of Si delta doping in GaAs, where the Si dopants are ideally deposited on a single plane during the epitaxial growth interruption.

due to subbands induced by the dopant plane.[44] A similar field-dependent effect of Hall concentrations can also be found in AlGaAs/GaAs HEMTs,[45] where the carriers in the AlGaAs layer are not fully depleted and thus form another conduction channel. The Hall concentration is therefore a weighted sum of the carrier density of each subband, as verified by temperature-dependent Hall measurements.[46]

To investigate the subband system associated with the delta-doping layer, Shubnikov–de Haas measurements can be used to extract the carrier concentration of each subband from the oscillation period of magnetoresistance versus magnetic field.[47] The SdH measurement requires high magnetic field, of the order of tesla (10^5 G), and liquid helium temperature operation. Figure 4 shows typical SdH oscillations of longitudinal magnetoresistance versus inverse magnetic field for delta-doping structures. The data indicate the existence of the 2DEG and separate subbands.[42,48] Each

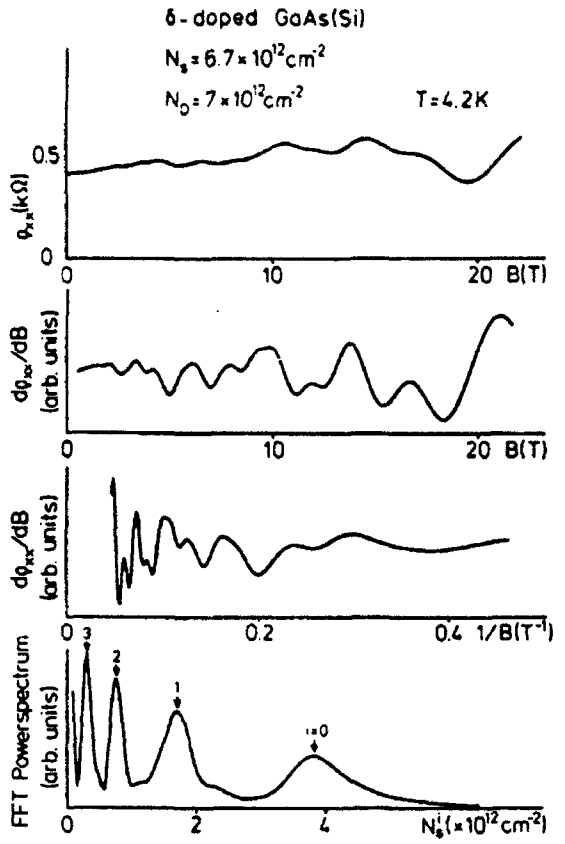


FIG. 4. SdH measurement of Si delta doping in GaAs, showing the processing of the transverse magnetoresistance ρ_{xx} to a FFT spectrum, indicating the location of each subband and its carrier concentration (Ref. 44).

group of oscillations corresponds to one occupied subband and subband carrier populations can be calculated from the oscillation periods. Fast Fourier transform (FFT) techniques have been employed to expedite the calculation. The subband mobilities of the 2DEG can also be estimated through the SdH measurement.[49]

V. CARRIER AND DOPANT PROFILES

The actual distributions of dopants and carriers along the growth direction in delta-doping structures should be measured to understand dopant incorporation during epitaxial growth and carrier localization mechanisms. Capacitance–voltage (C–V) measurement is a well-accepted electrical profiling technique for doping in semiconductors.[40] The C–V profiling method measures the free-carrier profile in a Schottky, p^+n, or n^+p diode structure, based on the depletion approximation. The carrier profile is a good approximation to the actual doping profile for homogeneous or slowly varying doping. The resolution of C–V profiling is of the order of Debye length,[50,51] typically > 50 Å. This has in the past led to a conclusion that carrier profiles any more abrupt than a Debye length cannot be detected by the C–V measurement. However, this conclusion does not hold in cases for which extremely sharp variation such as delta doping occurs, because of quantum size effects. In delta-doping layers,

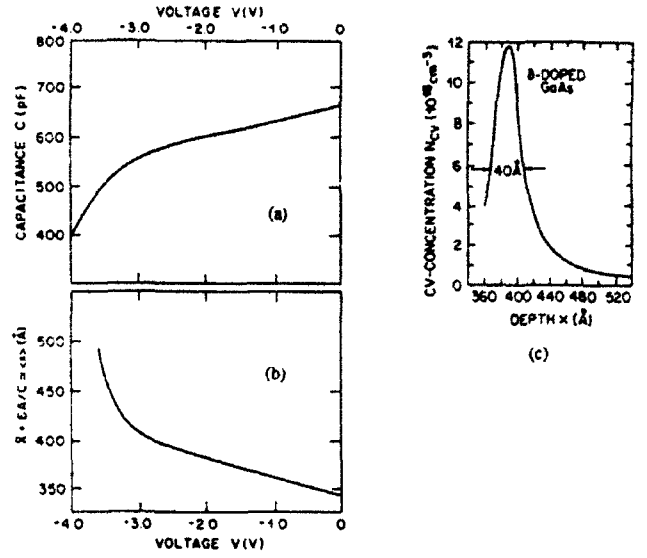


FIG. 5. (a) The measured C–V curve, (b) the C–V profile depth, and (c) the C–V profile with a FWHM of 40 Å, for Si delta-doped GaAs (Ref. 53).

the electrons are described by wave functions corresponding to subbands, rather than by dopant distribution functions.[52] The C–V profiling thus shows the expectation values through the conventional formulation.[53] Theoretical calculations demonstrate that the resolution of the C–V profiling in the delta-doping structures is the spatial extent of the ground state wave function,[54] typically of the order of lattice constants. For Si delta doping in GaAs, a typical C–V curve and the C–V depth are illustrated in Fig. 5. The resulting C–V carrier profile, also shown in Fig. 5, exhibits a FWHM of 40 Å, much less than the Debye length. The profile width strongly depends on the dopant diffusion and therefore the substrate temperature during growth.[55] The C–V profiles of Be delta doping in GaAs at low growth temperature < 550 °C are even narrower than that of Si delta doping, as illustrated in Fig. 6. The narrowest C–V profile FWHM of 12 Å yet reported was obtained in delta-doped quantum

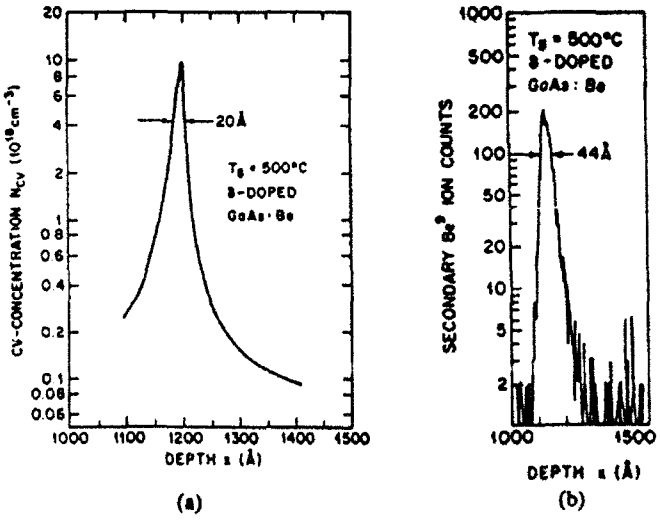


FIG. 6. (a) The C–V profile with FWHM of 20 Å and (b) the SIMS profile of the Be delta doping in GaAs grown at 500 °C by MBE (Ref. 8).

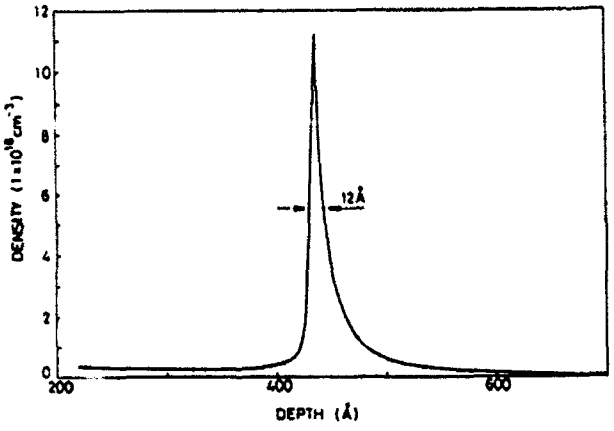


FIG. 7. The narrowest C–V profile with FWHM of 12 Å, as reported in Ref. 56 in a delta-doped GaAs/InGaAs quantum well.

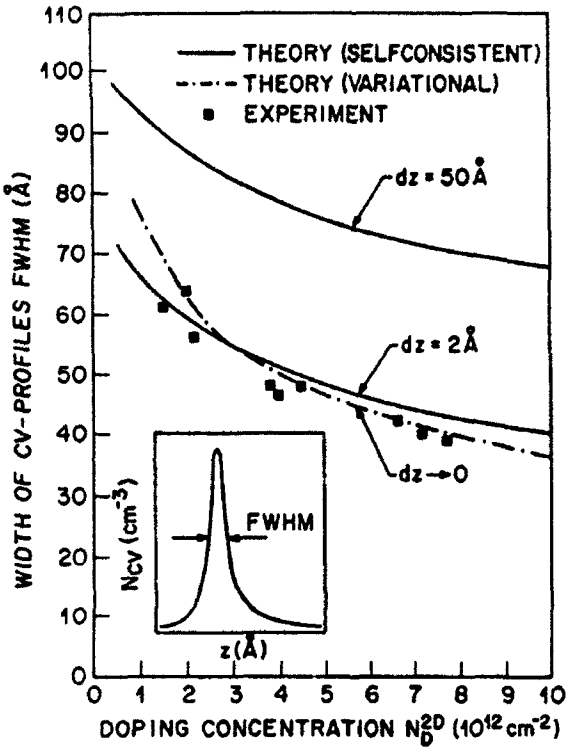


FIG. 9. Comparison of FWHM of C–V profiles based on theories and experiments, indicating that the dopants are in fact confined in a monolayer plane (Ref. 53). The dz represents the FWHM of the dopant distribution.

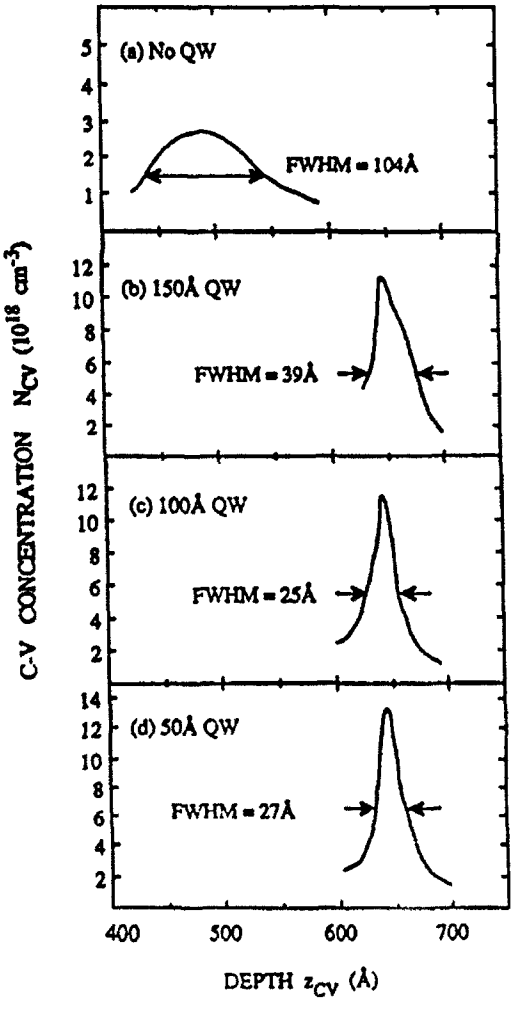


FIG. 8. The C–V profiles of delta doping in (a) bulk GaAs, (b) 150 Å QW, (c) 100 Å QW, and (d) 50 Å QW samples. All samples were grown at 600 °C by MBE with identical thermal treatment after the delta-doping layers were deposited, yielding identical dopant distributions.

wells,[56] as shown in Fig. 7. The additional confinement of the quantum wells gives rise to much narrower carrier profiles than that of delta doping in bulk materials. Our studies of delta-doped quantum wells have demonstrated a significant narrowing effect of the C–V profiles.[57] Figure 8 shows that the FWHM of the C–V profiles decreases as the quantum well width decreases.[57]

The actual dopant distributions of the delta doping are important information to determine the dopant diffusion during growth or annealing.[58] Secondary ion mass spectroscopy (SIMS) has been used to study the dopant diffusion[59] and verify that the dopant distributions of the delta doping are close to a delta function, as shown in Fig. 6(b). The resolution of the SIMS technique is limited by roughening of the surface and the "knock-on-effect," which tend to increase the uncertainty and width of the dopant profile.[8] This can be seen in Fig. 6, in which the C–V profile shows narrower width than the SIMS profile. Theoretical calculations obtained by fitting dopant distributions into the C–V carrier profiles, as illustrated in Fig. 9, reveal that the dopants are spatially confined within a few lattice constants. However, the presence of DX centers for heavily Si doped GaAs may in fact cause a narrowing effect on the C–V carrier profile, indicating that the actual dopant distributions may be wider than previously assumed.[60]

In conclusion, the measurements of the abrupt transitions in III–V materials require high resolution measuring tech-

nique as well as theoretical calculations involving quantum mechanics and statistical mechanics. The PL has been able to measure the abruptness of intrinsic or lightly doped heterostructures such as quantum wells, and the electrical properties of doped structures such as MDQWs. To measure abrupt doping profiles such as delta doping, it is essential to employ measuring tools with resolutions of the order of lattice constants. Full investigation of the delta-doping layers, however, needs the combining of experiments and calculations and is a very attractive topic in physics and device applications.

ACKNOWLEDGMENTS

The authors wish to thank T. R. Block for providing the PL spectra and useful information. The authors are also grateful to K. Sadra for theoretical discussion. This work was supported by the Joint Services Electronics Program under Contract No. AFOSR F49260-89-C-0044 and the Army Research Office under Contract No. DAAL 03-88-K-0060.

[1]D. C. Reynolds, K. K. Bajaj, C. W. Litton, P. W. Yu, J. Singh, W. T. Masselink, R. Fischer, and H. Morkoç, Appl. Phys. Lett. **46**, 51 (1985).
[2]R. C. Miller, D. A. Kleinman, W. A. Nordland, Jr., and A. C. Gossard, Phys. Rev. B **22**, 863 (1980).
[3]C. E. C. Wood, G. Metze, J. Berry, and L. F. Eastman, J. Appl. Phys. **51**, 383 (1980).
[4]H. Ohno, E. Ikeda, and H. Hasegawa, Jpn. J. Appl. Phys. **23**, L369 (1984).
[5]S. Sasa, S. Muto, K. Kondo, H. Ishikawa, and S. Hiyamizu, Jpn. J. Appl. Phys. **24**, L602 (1985).
[6]M. Hashemi, J. Ramdani, B. T. McDermott, K. Reid, and S. M. Bedair, Appl. Phys. Lett. **56**, 964 (1990).
[7]W-P. Hong, F. DeRosa, R. Baht, S. J. Allen, and J. R. Hayes, Appl. Phys. Lett. **54**, 457 (1989).
[8]N. Kogayashi, T. Makinoto, and Y. Horkoshi, Appl. Phys. Lett. **50**, 1435 (1987).
[9]E. F. Schubert, J. M. Kuo, R. F. Kopf, H. S. Luftman, L. C. Hopkins, and N. J. Sauer, J. Appl. Phys. **67**, 1969 (1990).
[10]K. Ploog, M. Hauser, and A. Fischer, Inst. Phys. Conf. Ser. **91**, 27 (1987).
[11]K. Ploog, J. Cryst. Growth **81**, 304 (1987).
[12]K. Ploog, M. Hauser, and A. Fischer, Appl. Phys. A **45**, 233 (1988).
[13]E. F. Schubert, J. Vac. Sci. Technol. A **8**, 2980 (1990).
[14]J. E. Cunningham, T. H. Chiu, B. Tell, and W. Jan, J. Vac. Sci. Technol. B **8**, 157 (1990).
[15]C. Webb, Appl. Phys. Lett. **54**, 2091 (1989).
[16]G. Gillman, B. Vinter, E. Barbier, and T. Tardella, Appl. Phys. Lett. **52**, 972 (1988).
[17]A. Zrenner, H. Reisinger, F. Koch, and K. Ploog, *Proceedings of the International Conference Physics Semiconductors*, edited by J. D. Chadi and W. A. Harrison (Springer, New York, 1985), p. 325.
[18]A. Zrenner, F. Koch, and K. Ploog, Inst. Phys. Conf. Ser. **91**, 171 (1987).
[19]J. Wagner, M. Ramsteiner, D. Richards, G. Fasol, and K. Ploog, Appl. Phys. Lett. **58**, 143 (1991).
[20]C. Weisbuch, R. C. Miller, R. Dingle, A. C. Gossard, and W. Wiegmann, Solid State Commun. **37**, 219 (1981).
[21]R. C. Miller and D. A. Kleinman, J. Lumin. **30**, 520 (1985).
[22]R. C. Miller, A. C. Gossard, and W. T. Tsang, Physica **117B and 118B**, 714 (1983).
[23]R. C. Miller, A. C. Gossard, W. T. Tsang, and O. Munteanu, Phys. Rev. B **25**, 3871 (1982).
[24]R. Dingle, Festkorperprobleme **15**, 21 (1975).
[25]T. R. Block and B. G. Streetman, J. Cryst. Growth **111**, 98 (1991).
[26]C. H. Yang, S. A. Lyon, and C. W. Tu, Appl. Phys. Lett. **53**, 285 (1988).
[27]A. Dodabalapur, A. Sadra, and B. G. Streetman, J. Appl. Phys. **68**, 4119 (1990).
[28]A. Dodabalapur, V. P. Kesan, D. P. Neikirk, and B. G. Streetman, J. Electron. Mater. **19**, 265 (1990).
[29]E. F. Schubert, A. Fischer, and K. Ploog, IEEE Trans. Electron Devices **ED-33**, 625 (1986).
[30]E. F. Schubert, J. E. Cunningham, and W. T. Tsang, Appl. Phys. Lett. **49**, 1729 (1986).
[31]E. F. Schubert, A. Fischer, Y. Horikoshi, and K. Ploog, Appl. Phys. Lett **47**, 219 (1985).
[32]W. E. Hoke, P. S. Lyman, W. H. Labossier, J. C. Huang, M. Zaitlin, and H. Hendriks, J. Vac. Sci. Technol. B **8**, 397 (1990).
[33]H. Ishikawa, H. Shibata, and M. Kamada, Appl. Phys. Lett. **57**, 461 (1990).
[34]T. Ishikawa, K. Ogasawara, T. Nakamura, S. Kuroda, and K. Kondo, J. Appl. Phys. **61**, 1937 (1987).
[35]E. F. Schubert, J. E. Cunningham, W. T. Tsang, and G. L. Timp, Appl. Phys. Lett. **51**, 1170 (1987).
[36]E. F. Schubert, L. Pfeiffer, K. W. West, and A. Izabelle, Appl. Phys. Lett **54**, 1350 (1989).
[37]T. J. Drummond, M. Keever, W. Kopp, H. Morkoç, K. Hess, and B. G. Streetman, Electron. Lett. **17**, 545 (1981).
[38]E. F. Schubert, J. B. Stark, T. H. Chiu, and B. Tell, Appl. Phys. Lett. **53**, 293 (1988).
[39]W-P. Hong, A. Zrenner, O. H. Kim, F. DeRosa, J. Harbison, and L. T. Florez, Appl. Phys. Lett. **57**, 1117 (1990).
[40]W-P. Hong, J. Harbison, L. Florez, and J. H. Abeles, Electron Device Lett. **10**, 310 (1989).
[41]D. K. Schroder, *Semiconductor Material and Device Characterization* (Wiley, New York, 1990).
[42]Y. C. Shih and B. G. Streetman (unpublished).
[43]T. Makomoto, N. Kobauashi, and Y. Horikoshi, J. Appl. Phys. **63**, 5023 (1988).
[44]A. Zrenner, F. Koch, and K. Ploog, Surf. Sci. **196**, 671 (1988).
[45]E. F. Schubert, K. Ploog, H. Dambkes, and K. Heime, Appl. Phys. A **33**, 63 (1984).
[46]T. Makimoto, N. Kobayashi, and Y. Horikoshi, J. Appl. Phys. **63**, 5023 (1988).
[47]D. G. Seiler, in *High Magnetic Fields in Semiconductor Physics II*, edited by G. Landwehr (Springer, Berlin, 1989), p. 578.
[48]A. Zrenner, H. Reisinger, F. Koch, K. Ploog, and J. C. Maan, Phys. Rev B **33**, 5607 (1986).
[49]S. Yamada and T. Makimoto, Appl. Phys. Lett. **57**, 1022 (1990).
[50]H. Kroemer and W. Y. Chien, Solid-State Electron. **24**, 655 (1981).
[51]W. C. Johnson and P. T. Panousis, IEEE Trans. Electron Devices **ED-18**, 965 (1971).
[52]E. F. Schubert and K. Ploog, Jpn. J. Appl. Phys. **25**, 966 (1986).
[53]B. Ullrich, E. F. Schubert, J. B. Stark, and J. E. Cunningham, Appl. Phys. A **47**, 123 (1988).
[54]E. F. Schubert, R. F. Kopf, J. M. Kuo, H. S. Luffman, and P. A. Garbinski, Appl. Phys. Lett. **57**, 497 (1990).
[55]M. Santos, T. Sajoto, A. Zrenner, and M. Shayegan, Appl. Phys. Lett. **53**, 2504 (1988).
[56]D. G. Liu, C. P. Lee, K. H. Chang, J. S. Wu, and D. C. Liou, Appl. Phys Lett. **57**, 1887 (1990).
[57]Y. C. Shih and B. G. Streetman (unpublished).
[58]T. H. Chiu, J. E. Cunningham, B. Tell, and E. F. Schubert, J. Appl. Phys. **64**, 1578 (1988).
[59]A-M. Lanzillotto, M. Santos, and M. Shayegan, Appl. Phys. Lett **55**, 1445 (1989).
[60]A. Zrenner, Appl. Phys. Lett. **55**, 156 (1989).

Degradation of photoluminescence from quantum wells grown on top of low-temperature buffers

A. Srinivasan, Y. C. Shih, and B. G. Streetman
Microelectronics Research Center, The University of Texas at Austin, Austin, Texas 78712

(Received 16 September 1991; accepted 17 September 1991)

The high-resistivity properties of low-temperature (200–300 °C) molecular-beam-epitaxy-grown GaAs and AlGaAs have great potential for device applications. The purpose of this study is to examine the effect of low-temperature buffers (LTB) on the crystal quality of epilayers grown on top. Photoluminescence (PL) from quantum wells grown at regular growth temperatures (620 °C) is found to be sensitive to changes in crystal quality introduced by the LTB. Limiting LT AlGaAs buffer thicknesses to a critical thickness is found necessary for growing layers with good quantum-well PL.

I. INTRODUCTION

Molecular-beam-epitaxy (MBE) grown low-temperature (LT) GaAs has been found to exhibit very interesting properties. Annealed LT GaAs exhibits high resistivity and has been used as buffer layers to reduce sidegating and backgating of metal-semiconductor field-effect transistors (MESFETs) and high-electron-mobility transistors.[1–3] Other applications include use of the LT layer as an isolation buffer layer for solar cells.[4] The LT GaAs has also been used as the photoconductor material for THz pulses.[5] Low-temperature buffers appear to reduce threading dislocations in lattice-mismatched devices.[6]

These layers have been found to contain about 1–2 at. % excess As which forms precipitates on annealing the layers to about 600 °C.[7–12] The high-resistivity properties have been attributed to the presence of these precipitates in both AlGaAs and GaAs. The addition of aluminum appears to increase the resistivity by an order of magnitude and makes LT AlGaAs preferable to LT GaAs as an isolation buffer.

These LT layers can be used as "active" layers or as isolation layers only if the epitaxial films grown on top are smooth and free from defects induced by the underlying LT layer.[13] Photoluminescence is sensitive to changes in quantum-well interfaces and material quality.[14,15] The rough interface or defects migrating to the quantum-well interface from the LTB would influence the PL.

II. EXPERIMENTAL STUDIES

Semi-insulating (100) GaAs wafers were used in this study. The wafers were first cleaned in de-ionized (DI) water for 2 min and blown dry in dry nitrogen. They were then mounted to molybdenum blocks using indium solder. After the cleaning and mounting process, the blocks were loaded into a Varian Gen II MBE system and outgassed at 450 °C for 1 h. The oxide was desorbed by heating the sample under an As overpressure at 670 °C. The layers were grown at an As/Ga incorporation ratio of 1.8 and a growth temperature of 620 °C. The As/Ga BEP ratio was about 15. The arsenic was obtained from a cracker source.[16]

All samples incorporated four quantum wells grown on a LT buffer. The thickness of the LT buffer layer was varied from 0 to 1.5 μm in different samples. LT GaAs buffers were grown at 230 and 300 °C, and LT $Al_{0.3}Ga_{0.7}As$ buffers were grown at 230 and 270 °C. These buffer layers were grown on top of a 0.5 μm smoothing layer of GaAs and were annealed for 10 min at 620 °C under an As_2 overpressure. Then four GaAs quantum wells (35, 19, 11, and 5 ML) were grown with 200 Å $Al_{0.3}Ga_{0.7}As$ barriers at 620 °C on top of the LTB. A four-quantum-well reference sample without any LT buffer was also grown for comparison studies.

Photoluminescence (PL) measurements were done on the quantum wells at 4.2 K. The 514.5 nm line from an argon-ion laser was used as the excitation source at power densities of 100 mW/cm^2. A typical spectrum for the reference (no LT buffer) sample is shown in Fig. 1.

III. RESULTS AND DISCUSSION

We observe a degradation in the PL of quantum wells on top of LT AlGaAs grown at either 230 or 270 °C. In fact, PL from the quantum well (35 ML) nearest to the LTB was not detectable. The spectral linewidth of the 5 and 11 ML wells increased as the thickness of the LT AlGaAs grown at 230 °C increased above 0.75 μm, as shown in Fig. 2. The peak luminescence of all the quantum wells with LTB thicknesses greater than 0.75 μm was red shifted (Fig. 3). The onset of this shift in the peak energy, together with a change in linewidths and a decrease in the integrated PL intensities, occurred at the same LTB thickness of 0.75 μm. LT AlGaAs grown at a higher temperature of 270 °C showed a smaller red shift of peak PL at about 1.5 μm (Fig. 4). In this case, the linewidth and the integrated PL intensity did not change significantly as the LTB thickness increased. The greater degradation of PL for lower growth temperatures may be due to greater incorporation of As in the lattice.

The migration of excitons to thicker wells would explain the red shifting of the peaks. This thermalization would only occur if the growth island size is smaller than the exciton diffusion length (about 1.3–1.8 μm). Quantum wells with a ragged interface would experience different

 0734-211X/92/020835-03$01.00 ©1992 American Vacuum Society

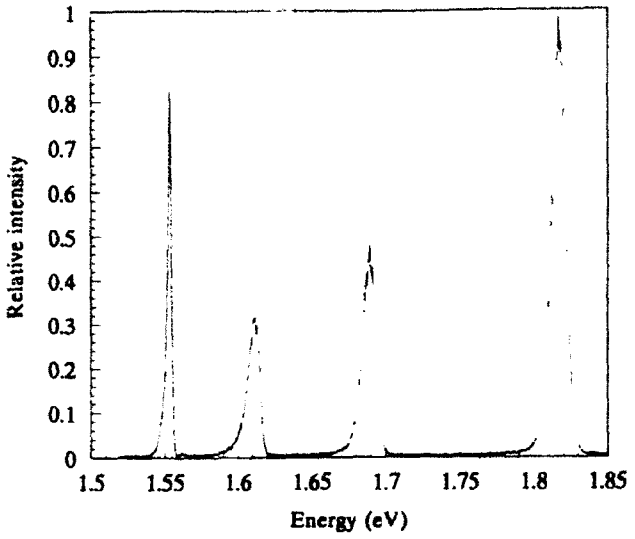


FIG. 1. Typical spectrum for the four-quantum-well structure without a LT buffer layer.

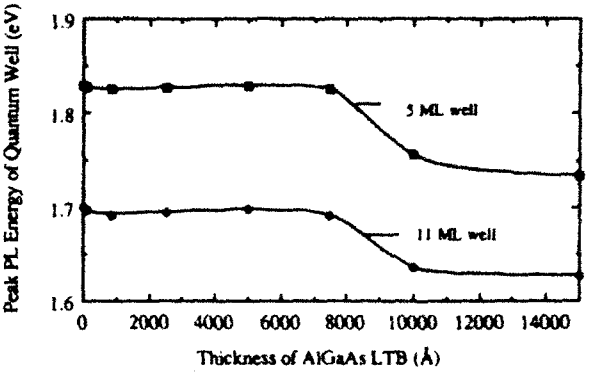


FIG. 3. Change in peak PL energy of quantum wells with the thickness of LT AlGaAs grown at 230 °C. The peak energy shifts to lower energies as the exciton themalizes to thicker wells.

confinements depending on the position of the exciton in the well. The PL from such structures would result in a relatively broad peak due to transitions at different energies.

LT GaAs grown at 230 °C under the same growth conditions did not result in any degradation in quantum-well PL linewidth or the red shift in peaks. All four quantum-well peaks appeared in the spectrum. The integrated PL intensity remained approximately constant for all the LT GaAs buffers. LT GaAs grown at 300 °C did not show any degradation for the buffer thicknesses (1.5 μm) studied.

After the 10 min anneal, the reflection high-energy diffraction (RHEED) (2×4) surface reconstruction was found to recover for both LT GaAs and LT AlGaAs. From qualitative studies of the RHEED patterns, the LT AlGaAs layers did not smooth out during the annealing period to the same extent as the LT GaAs layers. This is consistent with the complete disappearance of the first quantum-well PL in the LT AlGaAs layers. The bond energy for Al–As is greater than Ga–As, which contributes to increased surface roughness and slower recovery of RHEED intensities.

IV. CONCLUSION

The experiments found a degradation of PL for quantum wells grown on top of LT buffers which had been grown at different temperatures in the range of 230–300 °C. The study of LTB material included both GaAs and AlGaAs. We conclude that surface roughness from the LT GaAs is not a problem if the buffers are annealed for times long enough for the RHEED intensities to recover. There is a thickness limit on the LT AlGaAs layers that are grown under similar growth conditions. This limit appears to be related to the intrinsic roughness introduced by Al and the amount of excess arsenic that is incorporated in the LT AlGaAs buffer. A higher temperature for growth of these LT layers appears to reduce the associated surface roughness. The reduced surface mobility of Al and Ga at these low temperatures could have a role in the contribution to surface roughness.

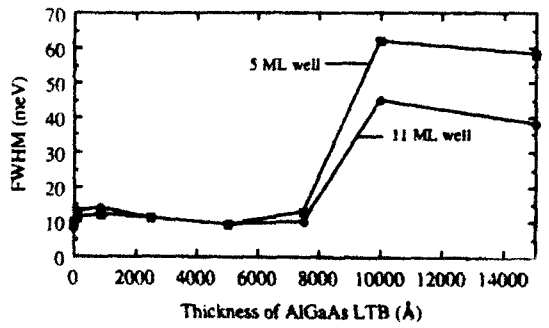


FIG. 2. Influence of LT AlGaAs layer thickness on FWHM of PL for two quantum-well widths.

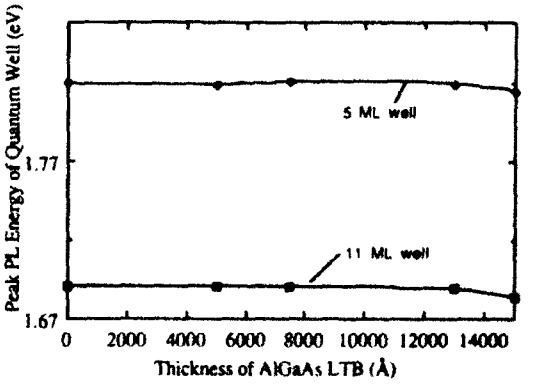


FIG. 4. Change in peak PL energy of quantum wells with the thickness of AlGaAs LT buffer grown at 270 °C.

ACKNOWLEDGMENTS

We would like to thank T. R. Block, K. Sadra, and T. Rogers for their helpful discussions. The technical support from T. J. Mattord is also appreciated. The work was supported by the Joint Services Electronics Program under Contract No. AFOSR F49620-89-C-0044, the Army Research Office under Contract No. DAAL-03-88-K-0060, and by the Science and Technology Center program of the National Science Foundation, Grant No. CHE 8920120.

[1]F. W. Smith, A. R. Calawa, C. Chen, M. J. Manfra, and L. J. Mahoney, IEEE Electron Device Lett. **9**, 77 (1988).
[2]F. W. Smith, C. L. Chen, G. W. Turner, M. C. Finn, L. J. Mahoney, M. J. Manfra, and A. R. Calawa, Proc. IEDM, 838 (1988).
[3]M. J. Delaneey, C. S. Chou, L. E. Larson, J. F. Jensen, D. S. Deakin, A. S. Brown, W. W. Hooper, M. A. Thompson, L. G. McCray, and S. E. Rosenbaum, IEEE Electron Device Lett. **10**, 355 (1989).
[4]G. Subramanian, A. Dodabalapur, J. C. Campbell, and B. G. Streetman, Appl. Phys. Lett. **58**, 2514 (1991).
[5]A. C. Warren, N. Katzenellenbogen, D. Grischkowsky, J. M. Woodall, M. R. Melloch, and N. Otsuka, Appl. Phys. Lett. **58**, 1512 (1991).
[6]R. A. Metzger, M. J. Delaney, L. McCray, H. Kanber, D. C. Wang, and T. Y. Chi, J. Vac. Sci. Technol. B **8**, 297 (1990).
[7]M. Kaminska, Z. Liliental-Weber, E. R. Weber, T. George, J. B. Kortright, F. W. Smith, B-Y. Tsaur, and A. R. Calawa, Appl. Phys. Lett. **54**, 1881 (1989).
[8]M. Kaminska, E. R. Weber, Z. Liliental-Weber, and R. Leon, and Z. U. Rek, J. Vac. Sci. Technol. B **7**, 710 (1989).
[9]A. C. Warren, J. M. Woodall, J. L. Freeouf, D. Grischkowsky, D. T. McInturff, M. R. Melloch, and N. Otsuka, Appl. Phys. Lett. **57**, 1331 (1990).
[10]M. R. Melloch, K. Mahalingam, N. Otsuka, J. M. Woodall, and A. C. Warren, J. Cryst. Growth **111**, 39 (1991).
[11]T. Y. Chu, A. Dodabalapur, A. Srinivasan, D. P. Neikirk, and B. G. Streetman, J. Cryst. Growth **111**, 26 (1991).
[12]A. C. Campbell, G. E. Crook, T. J. Rogers, and B. G. Streetman, J. Vac. Sci. Technol. B **8**, 305 (1990).
[13]B. J. F. Lin, C. P. Kocot, D. E. Mars, and R. Jaeger, IEEE Trans. Electron Devices **37**, 46 (1990).
[14]T. R. Block, Ph.D. dissertation, University of Texas, Austin, 1991.
[15]M. A. Herman, D. Bimberg, and J. Christen, J. Appl. Phys. **70**, R1 (1991).
[16]T. J. Mattord, V. P. Kesan, G. E. Crook, T. R. Block, A. C. Campbell, D. P. Neikirk, and B. G. Streetman, J. Vac. Sci. Technol. B **6**, 1667 (1988).

Enhancement of carrier concentration and spatial confinement in molecular-beam epitaxial Si and Be δ-doped GaAs by increasing As_4/As_2 flux ratio

Y. C. Shih, T. R. Block, and B. G. Streetman
Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712

(Received 16 September 1991; accepted 17 September 1991)

We report the effects of changing As_4/As_2 flux ratio from an As cracking source on Si and Be δ-doped GaAs grown by molecular-beam epitaxy. It is observed that the carrier concentration increases as the As_4/As_2 flux ratio increases. The spatial confinement of carriers in the δ-doped induced potential well is also enhanced using high As_4/As_2 flux ratio. These effects are attributed to the enhancement of dopant incorporation by As_4 during the δ-doping growth period.

I. INTRODUCTION

The technique of δ doping in semiconductors has motivated a large amount of research on device applications and material studies. By introducing a single plane of dopants into semiconductors, a variety of novel devices have been suggested.[1] In practice, the δ doping is accomplished by depositing dopant atoms onto a growth-interrupted semiconductor surface using epitaxial growth techniques, such as molecular-beam epitaxy (MBE) or metalorganic chemical vapor deposition (MOCVD). Abrupt doping profiles with widths on the order of few lattice constants have been confirmed by various characterizing techniques and theoretical calculations.[2] For example, an extremely narrow capacitance–voltage (C–V) profile with a full width at half-maxima (FWHM) of 5 Å has been reported in δ-doped quantum wells.[3] The abrupt dopant distributions in δ-doped structures give rise to many interesting phenomena, such as quantum-size effect from the dopant-induced potential well, enhanced carrier mobilities, and enhanced carrier concentrations.[4]

MBE is widely employed to synthesize δ doping in semiconductors, such as GaAs and Si. In δ-doped GaAs, the Ga flux is normally turned off to interrupt the GaAs growth and the dopant shutter is opened, while the As shutter is kept open to maintain surface stoichiometry. In other words, the growth-interrupted surface is exposed to both As and dopant fluxes during the δ doping growth period. Si and Be are used respectively as donor and acceptor species in δ-doped GaAs. Currently, MBE systems use tetrameric (As_4) and/or dimeric (As_2) arsenic sources. In practice, the As_2 cell consists of an As_4 cracking section and a sublimation section.[5] The As flux from the As_2 cracking source in fact contains a fraction of uncracked As_4, and by adjusting the cracking temperature the fraction of As_2 in the As flux can be varied. The growth mechanisms of GaAs with As_2 and As_4 are different, being first-order[6] and second-order processes,[7] respectively. The As_2 source has some advantages over the As_4 source in crystal growth;[5] however, for homogeneous Si doping in GaAs, it has been shown recently that the use of a high As_4/As_2 flux ratio from an As_2 cracking source leads to higher electron concentrations.[8] No similar study has been done on δ-doped GaAs prior to the present work.

We report the effects of the As_4/As_2 flux ratio on Si and Be δ-doped GaAs. Hall measurements are employed to measure the free-carrier concentrations, and the spatial extents of the carrier distributions are characterized by C–V measurements.

II. EXPERIMENTS

Samples were grown in a Varian Gen II MBE system with a structure consisting of a 5000 Å undoped GaAs smoothing layer grown at 600 °C and a δ-doping layer sandwiched between two 1000-Å-thick GaAs layers grown at 600 or 550 °C. During the δ-doping growth period, the Si (Be) shutter was open for 30 (10) s. The GaAs growth rate was 1.0 ML/s. The As_4/As_2 flux ratio was adjusted by changing the cracker furnace temperature, as measured by a pyrometer. The As-to-Ga incorporation ratio is kept constant at 1.7 by adjusting the sublimator power. Figure 1 shows the cracker furnace temperature and the As beam equivalent pressure (BEP) as functions of the voltage of the cracker power supply. As shown in Fig. 1, the cracker furnace temperature is roughly proportional to the voltage applied to the cracker heating element; therefore, subsequent discussions and figures refer to the cracker voltage as an indicator of temperature. Since As_4 has a lower sticking coefficient than As_2, a high As BEP would imply a high As_4/As_2 flux ratio for constant As-to-Ga incorporation ratio, which corresponds to low cracker voltages. Residual gas analysis (RGA) also confirmed that the relative amount of As_4 in the As flux increases as the cracker voltage decreases. Hall measurements were performed at room temperature and no significant carrier freezeout was observed at low temperature (77 K). An HP4194 Gain-phase/Impedance analyzer was employed to measure C–V profiles at frequencies ranging from 100 kHz to 1 MHz.

III. RESULTS AND DISCUSSION

Figure 2 shows the two-dimensional free-carrier concentration as a function of the cracker voltage at substrate temperatures of 550 and 600 °C for Si and Be δ doping,

 0734-211X/92/020863-03$01.00 ©1992 American Vacuum Society

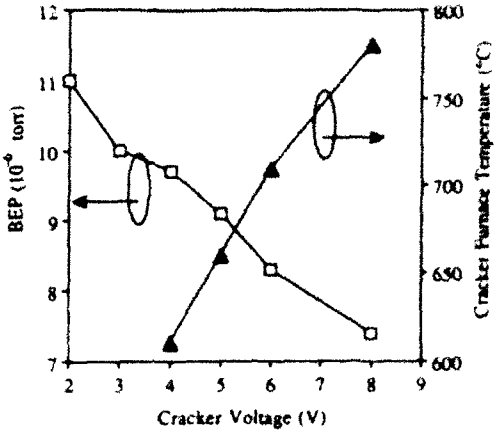


FIG. 1. The As BEP and the cracker furnace temperature as functions of the cracker voltage.

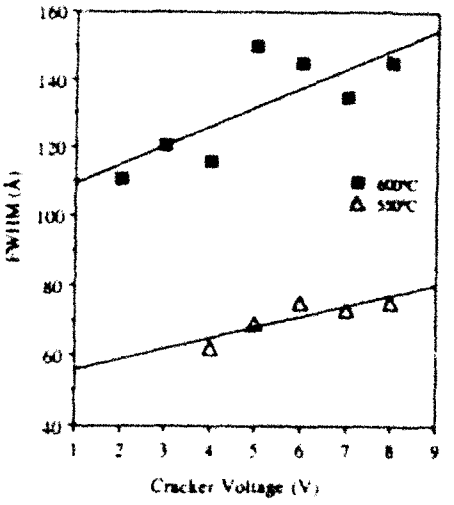


FIG. 3. The FWHMs of the C–V profiles of Si δ-doped GaAs grown at 600 and 550 °C.

respectively. No significant dependence of the carrier concentration on the substrate temperature is observed. It can be seen that as the cracker voltage increases (As_4/As_2 flux ratio decreases) the carrier concentration decreases by as much as a factor of 2 for both types of doping. The Hall mobilities are nearly constant at about 2300 and 100 $cm^2/V\,s$ in Si and Be δ-doped samples, respectively, for all cracker voltages. These observations suggest that during the δ-doping growth period As_4 molecules may have the following effects versus As_2 molecules: (1) higher incorporation of the dopant atoms onto the surface or (2) a higher electrical activation or less autocompensation of the dopants after the deposition of the δ-doping layer. In Be δ-doped samples, it is unlikely that the electrical activation of Be atoms depends on the species of As after the δ-doping growth period. Previous work[8] also indicates that autocompensation of Si in GaAs is a negligible effect. There-

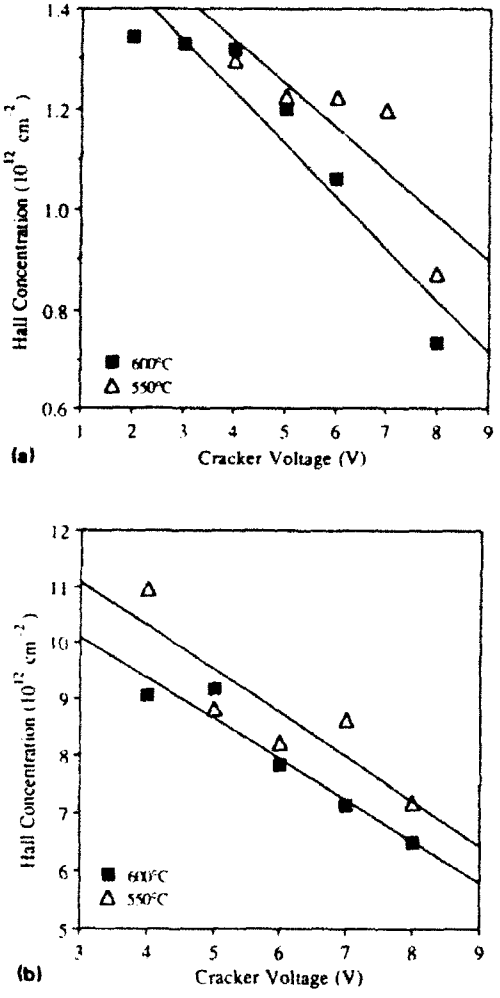


FIG. 2. The Hall concentrations of (a) Si and (b) Be δ-doped GaAs at substrate temperatures of 550 and 600 °C.

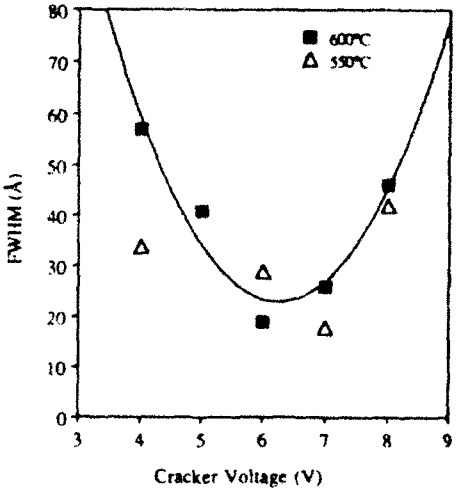


FIG. 4. The FWHMs of the C–V profiles of Be δ-doped GaAs with growth temperatures of 600 and 550 °C.

fore, the first possibility is the most probable explanation of the experimental results. Since As_2 is more reactive at the surface than As_4, it may be that the formation of volatile molecules involving As_2 and the dopants can cause the reduction of dopant incorporation in the crystal.

C–V measurements confirm that the dopant incorporation is higher for high As_4/As_2 flux ratio. As shown in Fig. 3, the FWHM of the *C–V* profiles of Si δ-doped GaAs increases as the cracker voltage decreases. This might be expected from the higher incorporation of dopants as the As_4 in the As flux increases. As a result, the confinement of carriers in the dopant-induced potential well is enhanced due to a larger number of ionized dopants and therefore stronger Coulombic attractive interactions. At low growth temperatures, the *C–V* profiles show less dependence on the cracker voltage because there is less As reevaporation. Similarly, in Be δ-doped GaAs, as shown in Fig. 4, the FWHM of the *C–V* profile decreases as the cracker voltage decreases. At low cracker voltages where the resulting dopant concentration is higher, however, the width of the *C–V* profile increases. This results from the fact that at very high Be concentration, the repulsive Coulombic interactions between the ionized Be acceptors become significant.[9] As a result, the dopant distributions and thus the carrier distributions are broadened when the spacing between Be acceptors reaches a minimum.

ACKNOWLEDGMENTS

This work was supported by the Joint Services Electronics Program under Contract No. AFOSR F49620-89-C-0044, the Army Research Office under Contract No. DAAL-03-88-K-0060, and the Science and Technology Center program of the National Science Foundation, Grant No. CHE 8920120.

[1] E. F. Schubert, J. Vac. Sci. Technol. A **8**, 2980 (1990).
[2] Recently summarized by B. G. Streetman and Y. C. Shih, J. Vac. Sci. Technol. B **10**, 296 (1992).
[3] Y. C. Shih and B. G. Streetman, Appl. Phys. Lett. **59**, 1344 (1991).
[4] E. F. Schubert, J. E. Cunningham, and W. T. Tsang, Solid. State Commun. **63**, 591 (1987).
[5] T. J. Matford, V. P. Kesan, G. E. Crook, T. R. Block, A. C. Campbell, D. P. Neikirk, and B. G. Streetman, J. Vac. Sci. Technol. B **6**, 1667 (1988).
[6] C. T. Foxon and B. A. Joyce, Surf. Sci. **64**, 293 (1977).
[7] C. T. Foxon and B. A. Joyce, Surf. Sci. **50**, 434 (1975).
[8] B. J. Wu, Y. J. Mii, M. Chen, and K. L. Wang, Appl. Phys. Lett. **58**, 391 (1991).
[9] E. F. Schubert, J. M. Kuo, R. F. Kopf, H. S. Luftman, L. C. Hopkins, and N. J. Sauer, J. Appl. Phys. **67**, 1969 (1990).

Scanning tunneling microscopy of GaAs multiple *pn* junctions

S. Gwo, A. R. Smith, and C. K. Shih
Department of Physics, University of Texas, Austin, Texas 78712

K. Sadra and B. G. Streetman
Department of Electrical Engineering, University of Texas, Austin, Texas 78712

(Received 13 April 1992; accepted for publication 22 June 1992)

Scanning tunneling microscopy and spectroscopy is used to study GaAs multiple *pn* junction samples cleaved in ultrahigh vacuum. Direct topographic contrast over the *pn* junctions can be observed in the constant current imaging mode. The topographic height in the *p*-type regions appears much lower (by about 5 Å) than that in the *n*-type regions. Tunneling spectroscopy measurements show consistency with the assignment of the *p*- and *n*-type regions. We discuss a possible mechanism for the observed contrast.

Recently, there have been numerous applications of scanning tunneling microscopy (STM) to study semiconductor junctions including homo- and heterojunctions.[1-7] The most appealing feature of this approach is the ability to simultaneously investigate both the structural and electronic properties at atomic or nanometer resolution. The understanding of these properties is very important for the development of ultrasmall devices for modern electronics.

Because STM is a surface sensitive technique, one needs to create a fresh surface which exposes the junction regions. It is generally agreed that the best way to do that is to cleave samples under ultrahigh vacuum (UHV). The tunneling tip is then brought into the junction region for STM studies. Since the device junction areas are usually less than one micron across, the major technical difficulty is the precise positioning of the tip in the junction region. Many groups have therefore combined STM with a scanning electron microscope (SEM)[1,2] to help position the tip in the junction region, with most of the studies concentrating on the heterojunction systems. In fact, it was only after an UHV-SEM/STM combination had been developed that the first atomically resolved images of semiconductor heterojunctions (AlGaAs/GaAs system) was achieved.[1]

As for the semiconductor homojunctions, most of the studies have concentrated on Si *pn* junctions,[5-7] although very few were studied with an UHV environment. Furthermore, most of those studies relied on biasing the *p*- and *n*-type regions differently in order to observe any contrast between the junction regions. On the other hand, STM studies of the GaAs *pn* junctions have never been explored except for a study using scanning tunneling potentiometry.[8]

STM studies of GaAs *pn* junctions have a major advantage over those of Si *pn* junctions. Since the intrinsic surface states of the cleaved GaAs (110) surface are outside the band gap, tunneling spectroscopy measures a surface band gap which is the same as the bulk band gap.[9] Furthermore, if there are no extrinsic surface states to pin the surface Fermi level, then the position of the Fermi level relative to the band edge measured by the tunneling spectroscopy is the same as that in the bulk.[9] Thus, the idea of mapping out the band profile across the *pn* junction using scanning tunneling spectroscopy is possible in the cleaved (110) surface of GaAs. This is certainly not the case for Si.

In this letter, we report scanning tunneling microscopy and spectroscopy (STM/S) studies of GaAs multiple *pn* junctions. To our knowledge, this is the first STM/S study of GaAs *pn* junctions. Our experiments were performed in a UHV-STM system equipped with a unique dual-axis sample translation stage (the design and the characteristic of which will be described elsewhere). GaAs multiple *pn* junctions were grown using molecular beam epitaxy (MBE) with varying doping densities and layer thicknesses. Si and Be were used as dopants and the layers were grown at 600 °C under As-stable conditions. Samples were cleaved in UHV with a base pressure of 5×10^{-11} Torr or less to create (110) surfaces, thus exposing the junction area. The tunneling tip was first brought to the sample within 20 μm from the edge by using an optical microscope. Instead of using a UHV-SEM/STM combination, we employed a different edge finding method which enables us to locate the tip over the junction regions very reliably. The details of this method will be described elsewhere.

Figure 1 shows a small area of a constant current image taken at the n^+ substrate right next to the *pn* junction regions. The image was obtained at a tip bias of +3 V relative to the sample, thus corresponding to that of the As sublattice.[9] The atomically resolved image is used to calibrate the piezoelectric scanner and to characterize the tip before the large area scan over the junction is performed. Figure 2(a) shows the grown device structure where the area enclosed by the dashed line corresponds to the scanned area shown in Fig. 2(b). This three-dimensional (3D) perspective-view image was obtained at tip bias of +2.9 V, thus corresponding to the image of the filled states of the sample. As one can clearly see in this perspective-view image, there is a large topographic height difference between the *n*-type and *p*-type regions in this constant current image. The topographic height in the *p*-type region appears lower than that in the *n*-type region. It also shows a dependence on either the device width or the doping level, namely, the shorter device region (or higher doping level) shows a higher contrast. The largest topographic height difference is about 5 Å. STM images obtained at the opposite bias (−2.9 V) also show similar contrast, namely, the topographic height in *n*-type regions being higher than that in the *p*-type regions.

The assignment of the *p*- and *n*-type regions was first

 0003-6951/92/341104-03$03.00 © 1992 American Institute of Physics

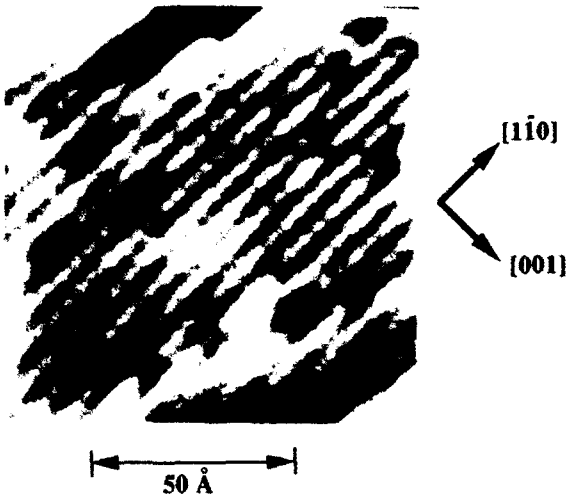


FIG. 1. Top-view STM image acquired at the n^+ substrate right next to the *pn*-junction regions. The image is obtained at a positive tip bias of 3 V, thus corresponding to the As sublattice. The scanning area is 100 by 100 Å.

based on a comparison of the length scale shown in the image with that of the grown device structure. In addition, tunneling spectroscopy was used to verify our assignment. Two typical tunneling spectra are shown in Fig. 2(c) which were acquired at the positions shown on the image 2(b) labeled as spots 1 and 2. The voltage scale on the *I-V* spectra represents the voltage of the tip relative to the sample; thus, positive voltages correspond to filled states of the sample. The positions of the band edges (also marked on the *I-V* spectra) are determined as the positions where there is no detectable tunneling current. The error bar is estimated to be ± 0.1 eV. As one can see, the spectra show that the Fermi level is closer to the valence band edge in the *p*-type region, whereas it is closer to the conduction band edge in the *n*-type region. However, our spectra show a slightly larger band gap than the bulk band gap of 1.4 eV. The measured gap values are 1.6 ± 0.2 and 1.8 ± 0.2 eV in the *n*- and *p*-type regions, respectively. This may be due to the effect of the tip-induced band bending in the measurement. Furthermore, the shift in the spectra acquired at the *n*-type region relative to that acquired at the *p*-type region is only about 0.7 ± 0.2 eV when one uses the assignment of the conduction band edges. This shift is smaller than the built-in potential. Nevertheless, the spectra show qualitative agreement with our assignment of the *p*- and *n*-type regions.

We now discuss the contrast mechanism. We interpret our image by considering both the work function difference in the *p*- and *n*-type regions and the effect of the tip induced band bending. Figure 3(a) shows energy diagrams of the tunnel junctions at the *p*- and *n*-type regions, respectively. It should be noted that the length scale in the tunneling region (order of 20 Å) and the length scale in the band bending region (order of a few hundred angstroms) is not the same. The work function at the *p*-type region is higher than that at the *n*-type region by about 1.4 eV. With the tip biased positively, the Fermi level of the tip is lower

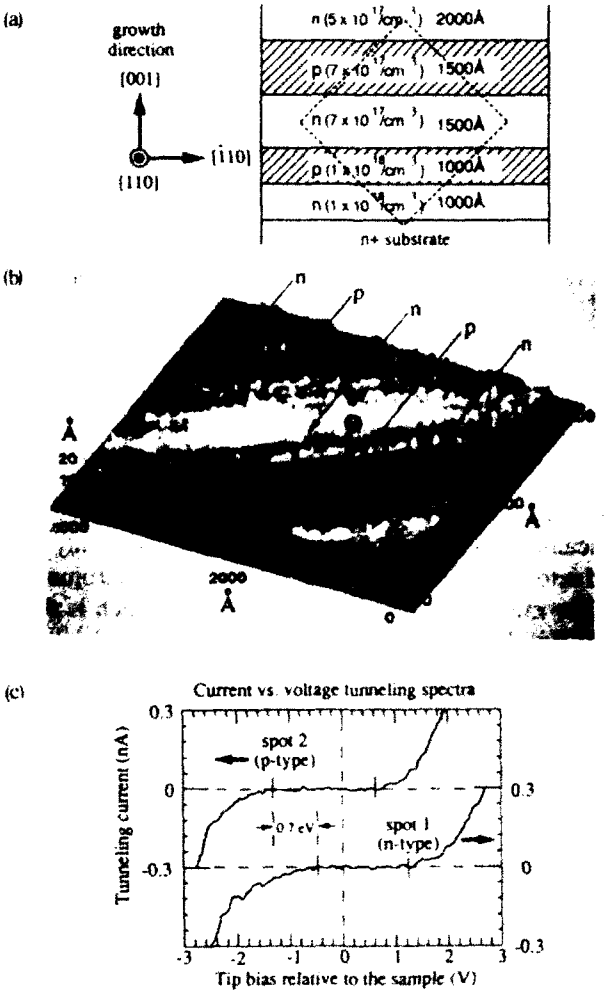


FIG. 2. (a) The device structure of GaAs multiple *pn* junctions grown by molecular beam epitaxy. The area enclosed by dashed line corresponds to the STM image shown in (b). (b) A 3D perspective-view STM image of the GaAs multiple *pn* junctions acquired at a tip bias of 2.9 V. The scanning area is 4000 by 4000 Å. (c) Tunneling spectra obtained on the *p*- and *n*-type regions. The spectra were taken at the locations labeled as spots 1 and 2 in the image.

than that of the sample. This will cause the band to bend downward resulting in a highly accumulated layer in the *n*-type and depleted layer in the *p*-type region (until the inversion occurs). A highly accumulated layer in the *n*-type regions means that there are appreciable occupied states in the conduction band. Since the tunneling occurs between the occupied states of the sample and the unoccupied states of the tip, the highest occupied states of the sample see the lowest tunneling barrier (shown as a thick dashed arrow in the figure). As a result, the effective tunneling barrier height for the *n*-type region is substantially smaller than that for the *p*-type region. This interpretation is consistent with the higher topographic heights observed in the *n*-type region.

The same interpretation applies for the case when the tip is negatively biased. Figure 3(b) shows the energy diagrams for this case where the depletion occurs in the *n*-type region and the accumulation occurs in the *p*-type region. Since the lowest tunneling barrier height occurs at the highest occupied states of the tip (shown by the thick

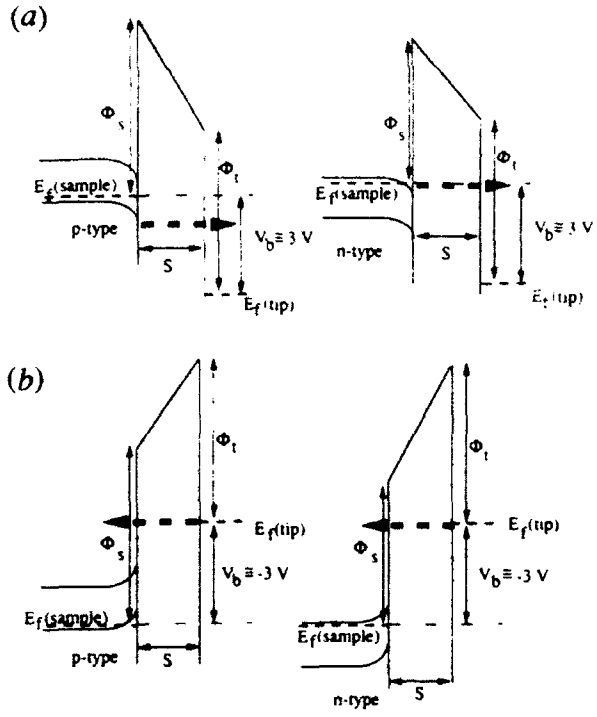


FIG. 3. (a) Energy diagrams of the tunnel junctions with the tip positively biased at the p- and n-type regions. The tip-induced band bendings are also shown. It should be noted that the length scale in the tunneling region (order of 20 Å) and the length scale in the band bending region (order of a few hundred angstroms) is not the same. (b) Same energy diagrams of the tunnel junctions as those shown in (a) except that the tip negatively biased.

dashed arrow in the figure), the effective tunneling barrier height is still smaller in the n-type region. Thus, this interpretation is also consistent with the same contrast observed in our study when the tip was biased negatively.

Consider that $I \sim \exp(-2\kappa S)$, where κ is related to the effective barrier height Φ, as

$$\kappa = (2m\Phi/\hbar^2)^{1/2}.$$

The observed topographic height difference can then be expressed, to first order, as

$$\Delta S \approx (\Delta\kappa/\kappa)\cdot S_0,$$

where S_0 is the average tip-to-sample distance. If we estimate that $\Delta\Phi$ is about 1–1.5 eV then $\Delta\kappa$ is about 0.15–0.2 Å^{-1}. A typical κ is 1 Å^{-1}. If we also assume that S_0 is about 15–20 Å, then ΔS is about 3–4 Å which is consistent with our experimental value of 5 Å. The assumption of S_0 being about 15–20 Å is based on previous measurements of the vacuum barrier resonance such as those used by others[10–12] to determine the absolute tip-to-sample distance. However, we do not rule out the possibility that there could exist other explanations for the pn contrast observed in our experiments.

Our interpretation of the contrast mechanism can be extended to explain the lack of contrast in the constant current image of the Si pn junctions cleaved in UHV.[7] The cleaved Si(111) surface has a small surface gap of about 0.2 eV,[11] thus pinning the surface Fermi level for both p- and n-type regions to within 0.2 eV. The lack of the variation in the surface Fermi level relative to the band edge results in little difference in the effective tunneling barrier heights for the p- and n-type regions. As a result, there is little difference in the observed topographic heights. Interestingly enough, there is a very recent STM study of Si pn junctions in air[13] with the oxide layer formed on the cleaved Si surface. The basic idea for this work was to use sample/oxide interface to replace the sample/vacuum interface, thus to eliminate the surface Fermi level pinning. Indeed, it was found that there exist large topographic height differences of about 5 Å (similar to ours) between the p-type and n-type regions for both the positive and negative tip biases. However, it should be noted that in that case, the tunnel junction is much less well-defined.

In summary, we have used scanning tunneling microscopy and spectroscopy to study GaAs multiple pn junctions. Direct to topographic contrast over the pn can be observed in the constant current image. Tunneling spectroscopy measurement shows consistency with the assignment of the p- and n-type regions. We believe that the primary mechanism for the observed contrast is due to the combination of the tip induced band bending and the difference in the work function between the p- and n-type regions. Our study has demonstrated the great potential of STM for providing an understanding of the structural and electronic properties of semiconductor junctions.

This work was supported by the Texas Advanced Research Program, and by the Joint Services Electronics Program under Contract No. AFOSR F49620-89-C-0044. We thank H.W.M. Salemink for sharing with us his method of cleaving samples. We also thank S. F. Alvarado for useful discussion in the development of the edge finding method. Fruitful discussion with R. M. Feenstra is appreciated.

[1] H. W. M. Salemink, H. P. Meier, R. Ellialtioglu, J. W. Gerritsen, and P. R. M. Muralt, Appl. Phys. Lett. **54**, 1112 (1989); O. Albreksen, D J. Arent, H. P. Meier, and H. W. M. Salemink, *ibid.* **57**, 31 (1990).
[2] J. M. Gomez-Rodriguez, A. M. Baro, J. P. Silveira, M. Vazquez, Y. Gonzalez, and F. Briones, Appl. Phys. Lett. **56**, 36 (1990).
[3] D. L. Abraham, V. Veider, Ch. Schonenberger, H. P. Meier, D. J. Arent and S. F. Alvarado, Appl. Phys. Lett. **56**, 1564 (1990); Ph. Renaud and S. F. Alvarado, Phys. Rev. B **44**, 6340 (1991).
[4] T. Kato and F. Osaka, Jpn. J. Appl. Phys. **30**, L1586 (1991).
[5] S. Hosaka, S. Hosoki, K. Takata, K. Horiuchi, and N. Natsuaki, Appl. Phys. Lett. **53**, 487 (1988).
[6] T. Takigami and M. Tanimoto, Appl. Phys. Lett. **58**, 2288 (1991).
[7] S. Kordic, E. J. van Loenen, and A. J. Walker, Appl. Phys. Lett. **59**, 3154 (1991); J. Vac. Sci. Technol. B **10**,496 (1992); IEEE Electron. Device Lett. **12**, 422 (1991).
[8] P. Muralt, Appl. Phys. Lett. **49**, 1441 (1986).
[9] R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, Phys. Rev. Lett. **58**, 1192 (1987).
[10] R. S. Becker, J. A. Golovchenko, and B. S. Swartzentruber, Phys. Rev. Lett. **55**, 987 (1985).
[11] J. A. Stroscio, R. M. Feenstra, and A. P. Fein, Phys. Rev. Lett. **57**, 2579 (1986).
[12] C. K. Shih, R. M. Feenstra, and G. V. Chandrashekhar, Phys. Rev. B **43**, 7913 (1991).
[13] R. Chapman, M. Kellam, S. Goodwin-Johnson, J. Russ, G. E. McGuire, and K. Kjoller, J. Vac. Sci. Technol. B **10**, 502 (1992).

First-order phase transition in a laser threshold

D. G. Deppe, D. L. Huffaker, T. J. Rogers, C. Lei, Z. Huang, and B. G. Streetman
Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712-1084

(Received 25 March 1992; accepted for publication 7 April 1992)

A mechanism for achieving a first-order phase transition in the threshold characteristics of a laser is discussed, and data demonstrating the effect are presented. It is suggested that a first-order phase transition can be realized through the design of a laser active region if the gain medium is localized to a spatial region with longitudinal dimensions much less than the wavelength of the resulting laser light. Evidence of the first-order phase transition is presented in the spontaneous emission characteristics of a microcavity semiconductor laser, which shows a decrease in the separation energy of the quasi-Fermi levels at threshold, along with hysteresis in the light versus current characteristics.

Early in the study of lasers it was recognized that the transition from below to above threshold exhibits characteristics that can be associated with a nonequilibrium phase transition.[1-3] In most lasers, threshold resembles a second-order transition, but a first-order transition can also be achieved in a Fabry–Perot cavity laser system through the use of, for example, a saturable absorber.[4] In this letter we suggest that a more fundamental means exists in which a first-order transition in laser threshold can be achieved, solely through the passive cavity design and the spatial distribution of the gain medium in the cavity. The design of such laser cavities has become possible with the advent of semiconductor microcavity devices, in which lasing can be achieved using a gain medium which exists in the form of a heterostructure quantum well (QW) with thickness much less than its emission wavelength. In this letter we present data on such a microcavity laser in which strong evidence of the first-order lasing transition has been found in the spontaneous emission characteristics of the device's threshold region. The spontaneous emission characteristics provide a direct means of analyzing the laser gain behavior at threshold, since both the spontaneous emission and optical gain depend on the quasi-Fermi level separation in the QW active region of the semiconductor laser. Saturation behavior of the gain at threshold is measured for the semiconductor microcavity laser and compared to a more standard longer cavity edge-emitting semiconductor laser. While the first-order lasing transition is demonstrated for the semiconductor laser, we believe the mechanisms to be quite general with potential application to other laser systems as well.

To illustrate the role of the cavity and gain medium design in controlling threshold characteristics, we consider a simple model in which a Fabry–Perot cavity contains a single optically thin gain region with a thickness $l_g \ll \lambda_0$ (gain path of l_g), where λ_0 is the center emission wavelength of the active medium. Coherence in spontaneous emission has similar descriptions in either the semiclassical or fully quantum mechanical approaches,[5] and it is illustrative in this discussion to consider the spontaneous emission from the semiclassical approach. The spontaneous emission then exists in the cavity in the form of wave packets, the coherence length of each wave packet being determined by some characteristic collision time which fully breaks coherence in the spontaneous emission.[6] Since the spontaneous wave packet in the cavity is a traveling wave, the optical gain felt by the wave packet when passing through the gain region, l_g, will be the same for a fixed number of excited dipoles whether the gain is confined to a region of $l_g \ll \lambda_0$, or distributed over a region of length $l_g \gg \lambda_0$. In either case the electric field amplitude is averaged over a cycle of oscillation to determine the photon-dipole interaction. However, above threshold the gain from the thin active region can be increased if the optically thin gain region is placed at a standing wave antinode of the laser field in the Fabry–Perot cavity. This is considerably different from the case for which the gain region is distributed over a region $l_g \gg \lambda_0$ in which the optical gain is identical for the traveling wave packet and the standing wave laser field. The change comes quite simply from strategically locating the active dipoles at a standing wave antinode. The result of the change is now an optical gain which depends on the light field in the cavity (that is a traveling wave due to spontaneous emission or a standing wave due to the laser field), and a hysteresis can be expected to develop in the threshold, characteristic of a first-order phase transition.

Gain in the semiconductor laser is characterized by the separation of the quasi-Fermi levels in the QWs, which will in turn be controlled by the drive current and the electron-hole recombination rate. Figure 1 shows a qualitative plot of the quasi-Fermi level separation in a laser active region versus drive current for three cases; a nonlasing device with no gain saturation, gain saturation due to a second-order lasing transition, and gain saturation due to a first-order lasing transition. When the gain region is distributed over a distance of $l_g > \lambda_0$, the saturation in the separation energy of the electron and hole quasi-Fermi levels occurs smoothly at threshold due to the increase of stimulated emission with a clamping of the population inversion, representative of the second-order phase transition. This behavior of the second-order phase transition is typical of the more standard edge-emitting semiconductor laser with a distributed gain region. In the first-order case the gain required for lasing decreases above threshold due to the es-

 0003-6951/92/253081-03$03.00 © 1992 American Institute of Physics

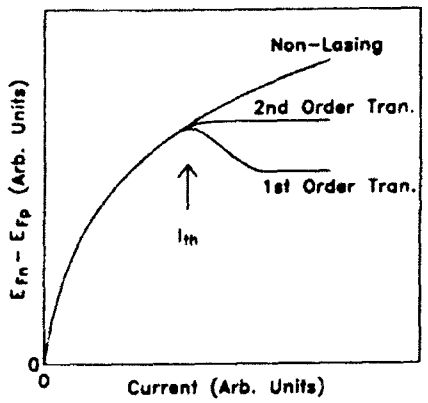


FIG. 1. Plot of the quasi-Fermi level separation in a p-n junction QW active region vs drive current showing the behavior for a nonlasing device, a laser with a second-order transition at threshold, and a laser with a first-order transition at threshold.

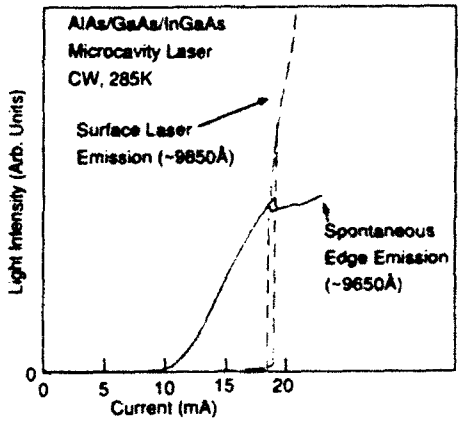


FIG. 2. Light vs current characteristics for the microcavity laser demonstrating hysteresis in the threshold characteristics (dashed curve) measured normal to the cavity, and also demonstrating a sharp decrease in the quasi-Fermi level separation (solid curve) for spontaneous emission measured out the edge of the cavity.

tablishment of the standing wave lasing field. This results in a decrease of the quasi-Fermi level separation with increasing drive current just above threshold (Fig. 1). In the analogy between phase transitions and laser threshold, the population inversion for laser threshold plays the role of temperature in an equilibrium phase transition, while the coherence length in the electric field is the ordering parameter. The population inversion in the semiconductor laser is given by the quasi-Fermi level separation, so voltage measured across the active region of the laser plays the role of temperature. Consequently, whether the phase transition is first or second order in the threshold of a laser diode can be determined by driving the diode with an ideal voltage source and measuring the threshold characteristics in the light output. A hysteresis will be observable for this ideal first-order case, due to the negative resistance provided by the laser diode junction impedance at threshold (Fig. 1). We note that a hysteresis in the operating characteristics would not be observed if the laser diode were driven by a current source even for the first-order transition (from the characteristics of Fig. 1), but instead an increased slope efficiency would be measured just above threshold.

Complete details of the device structure and operating characteristics can be found in Ref. 7, and we give only a brief description of the devices here. The structures consist of a Fabry–Perot microcavity composed of an optical wavelength thick $Al_{0.67}Ga_{0.33}As$ spacer region clad on the substrate side by a 19.5 pair GaAs/AlAs distributed Bragg reflector (DBR), and on the other side by a combination of a 3 pair GaAs/AlAs DBR and 4 pair CaF_2/ZnSe DBR. At the center of the $Al_{0.67}Ga_{0.33}As$ spacer region is an active region consisting of two $In_{0.2}Ga_{0.8}As$ 60 Å thick QWs with 60 Å GaAs barriers on either side. The combined thickness of the QW active region is 180 Å, compared to the emission wavelength from the QWs of $\lambda \sim 0.28$ μm (inside the semiconductor material). This system easily satisfies the requirement, therefore, of a gain region which is optically thin. Hysteresis develops simultaneously in both the light versus current and current versus voltage characteristics at lasing threshold.

To isolate the role of the quasi-Fermi levels and gain at lasing threshold, we have made a relative measurement of the spontaneous emission intensity from the QW active region versus current from below to above threshold. The spontaneous emission characteristics provide a sensitive probe as to the relative quasi-Fermi level behavior in the laser active region, and this technique has been used by others to characterize optical gain in semiconductor lasers.[8] To avoid modulation effects due to cavity filtering, the spontaneous emission is detected from the edge of the laser. Figure 2 shows the light versus current characteristics measured continuous-wave (cw) of the spontaneous emission measured from the device edge (solid curve), as well as from the laser aperture (dashed curve). The spontaneous emission peak is at ~ 0.98 μm which is also the lasing wavelength. As is characteristic of all these microcavity lasers, a hysteresis exists in the light versus current curve, with the same hysteresis always simultaneously observed in the current versus voltage curve. The spontaneous edge emission of Fig. 2 is measured through a spectrometer at a wavelength of 0.965 μm, well away from the lasing wavelength, to avoid effects of scattered laser emission. Laser threshold occurs at 18.5 mA with the characteristic hysteresis in the light versus current curve. In the cleaving of the device to obtain a measure of the spontaneous edge emission, some damage occurred giving a nonradiative current path some distance from the 10 μm laser active region. This current path increases the device threshold, and is clearly observed in the spontaneous edge emission which only "turns on" above 9 mA. An unmistakable dip is measured in the spontaneous emission at threshold (solid curve, Fig. 2), indicative of the abrupt decrease in the separation of the quasi-Fermi levels. A hysteresis can be seen in the switching of the spontaneous emission light intensity which occurs concomitantly with the switching in the laser light at threshold. This switching behavior in the spontaneous emission at threshold is observed up to the shortest measured wavelength of 9100 Å.

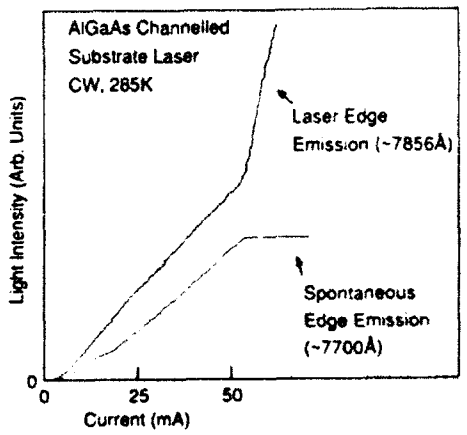


FIG. 3. Light vs current characteristics for an edge-emitting channeled-substrate laser for both the lasing wavelength at 7856 Å, and a much shorter wavelength of 7700 Å which shows ideal saturation behavior indicative of a second-order lasing transition.

Above threshold the quasi-Fermi level separation increases slightly, with a much smaller slope than that observed below threshold. We find the same threshold behavior in the spontaneous emission for other laser microcavity devices. The saturation of the quasi-Fermi level separation above threshold, which shows a slight increase with increasing drive current in Fig. 2, will be determined by intraband scattering rates relative to stimulated emission rate and pump rate. We note that a consistent explanation can be provided only by considering the role of the lasing mode on threshold characteristics for the following reasons. The *p*-side mirror of the laser is deposited using electron-beam evaporation after crystal growth. Therefore we are able to observe an individual device's current-voltage characteristics before deposition of the mirror, and before lasing action is observed. The lasing photon energy lies below the absorption gap of other layers in the semiconductor heterostructure, which consist of GaAs and AlAs. Switching behavior in the electrical and lasing characteristics is only observed at threshold after the *p*-side mirror reflectivity is increased sufficiently to achieve lasing.

A major drawback to microcavity semiconductor lasers is the high electrical series resistance measured in every device reported to date. This series resistance, which is several hundreds of ohms at threshold, will cancel the negative resistance due to junction impedance for the first-order lasing transition. Because of the biasing scheme used in the devices reported here, the negative resistance of the laser active region is coupled to an additional positive feedback current path which allows the direct observation of the bistability in our current-voltage and light-current characteristics at threshold. The additional feedback comes from a leaky *n*-type blocking layer which funnels most of the bias current into a 10 μm diam active region, but which also allows some leakage current and carrier buildup in the QW region surrounding the actual laser cavity. At threshold the collapsing quasi-Fermi levels pull excess charge into the 10 μm laser active region from the surrounding QW region. In this case the negative resistance intrinsic to the laser active region is "amplified" by this additional leakage current.

The switching behavior observed in the spontaneous emission characteristics of the microcavity laser can be compared to the more typical behavior expected for edge emitting lasers with distributed gain regions. For comparison, the spontaneous emission characteristics measured out the edge of an AlGaAs-GaAs channeled substrate laser are shown in Fig. 3. The continuous-wave edge emission is shown for measurements both at the lasing wavelength, in which threshold is observed at $\sim$50 mA, and at a somewhat shorter wavelength of 0.77 μm. The spontaneous emission measured at 0.77 μm shows the ideal behavior of gain saturation expected for a second-order phase transition, with no change in spontaneous emission intensity above threshold.

In summary, we have presented a mechanism for achieving a first-order phase transition in the threshold characteristics of a laser. The mechanism is based on the gain dependence on the lasing mode of light in the cavity when the gain region is placed at a standing wave antinode. The semiconductor microcavity laser provides a suitable system for building such lasers, since high gain can be achieved from an optically thin active region. That such characteristics might be tied to an intrinsic bistability in the cavity and gain region makes the devices particularly attractive for large optical systems due to the potential for high speed optical switching. However, the mechanisms as discussed in this letter are quite general, and may serve to impact the design of other laser systems in which a first-order lasing transition is desired.

This work has been supported by the Office of Naval Research Young Investigator Program, and the Joint Services Electronics Program Contract No. AFOSR F492620-89-C-0044.

[1] H. Haken, Rev. Mod. Phys. **47**, 67 (1975).
[2] V. DeGiorgio and M. O. Scully, Phys. Rev. A **2**, 1170 (1970).
[3] P. Lett, W. Christian, S. Singh, and L. Mandel, Phys. Rev. Lett. **47**, 1892 (1981).
[4] R. Salomaa and S. Stenholm, Appl. Phys. **14**, 355 (1977).
[5] H. Haken, *Light* (North-Holland, Amsterdam, 1985), Vol. 1.
[6] R. Loudon, *The Quantum Theory of Light* (Oxford University, New York, 1983), Chaps. 2 and 3.
[7] D. G. Deppe, C. Lei, T. J. Rogers, and B. G. Streetman, Appl. Phys. Lett. **58**, 2616 (1991).
[8] V. P. Gribkovski and V. A. Samoilyukovich, J. Appl. Spectrosc. **11**, 838 (1969); C. H. Henry, R. A. Logan, and F. R. Merritt, J. Appl. Phys. **51**, 3042 (1980).

Optical Memory Using a Vertical-Cavity Surface-Emitting Laser

D. L. Huffaker, W. D. Lee, D. G. Deppe, C. Lei, T. J. Rogers, J. C. Campbell, and B. G. Streetman

Abstract—**Data are presented demonstrating optical switching and memory in a bistable vertical-cavity surface-emitting laser. Optical switching from lasing to nonlasing or from nonlasing to lasing in the vertical-cavity laser is demonstrated using an AlGaAs probe laser at ~ 0.78 μm.**

WE have recently demonstrated a vertical-cavity surface-emitting laser (VCSEL) that exhibits concomitant bistable behavior in the light versus current and current versus voltage characteristics [1]. Such device operation is of considerable interest for applications of optical switching and optical memory. It is especially desirable to realize this type of operation in a VCSEL, given the natural extension to large area two-dimensional arrays, and the potential ease with which such a device geometry may be optically probed. There have been recent reports on optical switching in VCSEL's that incorporate pn-pn-layered structures [2] or optically addressable phototransistors [3]. The device structure presented in our work differs from those in terms of the coupling which exists between the lasing light of the VCSEL active region and the electrical switching in the surrounding pn-pn-layered structure [1]. Bistability in this device structure has been described in some detail [1], however, initial attempts to optically switch the device with an external probe were unsuccessful. Here we present the first demonstration of optically switching with an external probe, and demonstrate optical memory and an optical inverter.

The design and fabrication of the VCSEL studied in this work have been discussed previously [1], [4], therefore, only a brief description of the VCSEL structure is given here. The epitaxial layers of the device, illustrated in Fig. 1, are grown in two different growth sequences using molecular beam epitaxy (MBE). The layers deposited in the initial growth are composed of a 1 μm n^+ GaAs buffer, followed by a 21.5 pair n-type quarter-wave stack of GaAs/AlAs, a lower n-type $Al_{0.67}Ga_{0.33}As$ confining layer, an undoped active region consisting of two 60 Å $In_{0.2}Ga_{0.8}As$ pseudomorphic quantum-wells (QW's) with three 60 Å GaAs barriers, an upper p-type $Al_{0.67}Ga_{0.33}As$ confining layer, one p-type quarter-wave pair of AlAs/GaAs, a 200 Å p-type $Al_{0.67}Ga_{0.33}As$ etch stop layer, and finally a 600 Å n-type GaAs current blocking layer. The p- and n-type $Al_{0.67}Ga_{0.33}As$ confining layers and QW active region have a total combined thickness of an optical wavelength. The p-type AlAs/GaAs pair is grown with a graded superlattice between the layers in order to minimize series resistance. After selective etching steps to define the 10 μm-diameter VCSEL active region, a second MBE growth of two more p-type AlAs/GaAs quarter-wave pairs is performed, again with superlattice grading layers inserted at each AlAs/GaAs interface. To form the dielectric stack deposited on the p-type semiconductor layers of Fig. 1, we have used either $ZnSe/CaF_2$ or Si/SiO_2 quarter-wave layers. The $ZnSe/CaF_2$ combination has been found to yield significantly improved device performance [5].

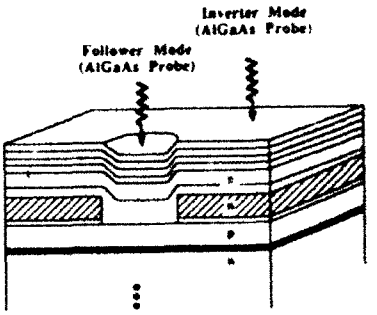


Fig. 1. Schematic cross-section (not to scale) of the VCSEL structure used in this work. Two modes of optical switching are possible by optically addressing either the 10 μm-diameter VCSEL active region (follower mode), or off of the VCSEL active region (inverter mode).

Light versus current characteristics of the VCSEL in the operating region around threshold are shown in Fig. 2. Lasing threshold is ~ 3.4 mA, and a hysteresis window of ~ 250 μA in current exists in the threshold region with a window of ~ 200 μW in light output power. The lasing wavelength is ~ 0.98 μm. A hysteresis also exists in the current versus voltage characteristics (not shown), providing for ~ 0.5 V of hysteresis from the "on" (lasing) switch point to the "off" (nonlasing) switch point [1]. As we show below, the hysteresis in the current versus voltage characteristics due to the lasing bistability makes it possible to bias the VCSEL in the voltage hysteresis window (or current hysteresis window shown in Fig. 2) and optically switch the device from either a lasing state to a nonlasing state (inverter mode), or from a nonlasing to a lasing state (follower mode). The selection of either the follower or inverter mode is made by positioning the AlGaAs laser probe beam either on the

Manuscript received July 5, 1991; revised September 9, 1991. This work was supported through an Office of Naval Research Young Investigator award, the Joint Services Electronics Program Contract AFOSR F492620-89-C-004, and the Army Research Office Contract DAAL 03-88-K-0060.

The authors are with the Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712.

IEEE Log Number 9104128.

1041-1135/91$01.00 © 1991 IEEE

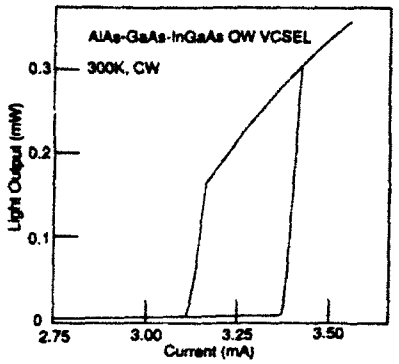


Fig. 2. Light versus current characteristics measured under continuous-wave room-temperature operation of the VCSEL at threshold showing the hysteresis. The hysteresis window is ~250 μA for current and ~200 μW for the light output.

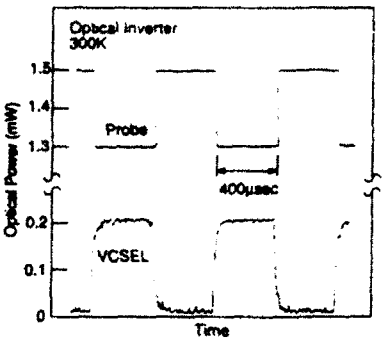


Fig. 3. Intensity versus time for the optical probe signal at a wavelength of 0.78 μm (upper trace) and VCSEL output at a wavelength of 0.98 μm (lower trace) for the VCSEL operated in the inverter mode (as illustrated in Fig. 1).

VCSEL active region for the follower mode, or somewhat off of the active region for the inverter mode, as shown in Fig. 1. The follower mode has been achieved while driving the VCSEL with either a current source or a voltage source, whereas the inverter mode has been achieved with only current source biasing.

Fig. 3 shows the optical switching characteristics of the VCSEL operating in the inverter mode, with the laser biased just above the hysteresis window. For this biasing scheme the VCSEL simply switches from above threshold (at ~0.2 mW output power) when the probe laser is low to below threshold (~0.01 mW) when the probe laser is high, without exhibiting memory.

Fig. 4 demonstrates the function of optical memory using the VCSEL in the optical follower mode. For the device measured for Fig. 4 as compared to the device measured for Fig. 3, a reduced contact dimension in which the contact consists of a metallized ring of 60 μm outer diameter and 30 μm inner diameter has been used, which greatly improves the device response time. The data of Fig. 4 demonstrate optical memory using 3 μs voltage pulses (lower curve, Fig. 4) to bias the VCSEL below threshold but in the hysteresis window of the voltage current curve. During the 3 μs voltage pulse to the VCSEL, a 200 ns voltage pulse is applied to the AlGaAs probe laser (upper curve, Fig. 4), the optical pulse of which (at a wavelength of 0.78 μm) triggers the VCSEL into the "on" (lasing) state. The internal optical latching of the VCSEL holds its output above threshold after the 200 ns pulse from the AlGaAs probe laser goes low. The middle curve of Fig. 4 shows the light output of the VCSEL measured using a Ge photodiode. The VCSEL turns on when triggered by the AlGaAs probe laser, and stays on until the end of the voltage pulse to the VCSEL. Also observable in the lower curve of Fig. 4 is the increase in current which occurs for the VCSEL when it is switched to the "on" state. The increase in current results from the optical switching of the pn-pn regions of Fig. 1 by the VCSEL laser light, which occurs only in the immediate vicinity of the VCSEL 10 μm active region. The impedance change is only observable in these devices at lasing threshold. Using the identical biasing scheme of Fig. 4, but with the optical path between the

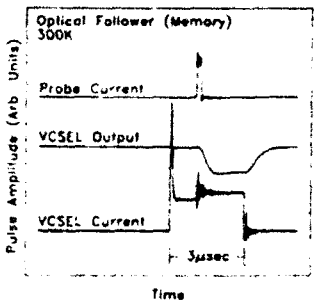


Fig. 4. Optical memory demonstrated using the VCSEL in the follower mode. The upper trace shows the current pulse applied to the AlGaAs probe laser which switches the VCSEL into the "on" (lasing) state. The lower trace shows the VCSEL bias current, and the middle trace shows the VCSEL light output. An impedance change in the VCSEL can be seen in the lower curve at the time of the probe pulse, due to the switch from nonlasing to lasing.

AlGaAs probe laser and VCSEL blocked, no switching of the VCSEL occurs, as indicated by the Ge-photodiode response and the lack of an impedance change in the VCSEL.

Optical memory has been achieved with AlGaAs probe pulses as short as 50 ns, which is the limit of the voltage pulser used to drive the AlGaAs probe, at the minimum probe power required for switching. Delay time between the leading edge of the optical pulse from the AlGaAs probe and optical switching of the VCSEL to the lasing state has been measured to be less than 10 ns, also at the minimum optical probe power required for switching. A delay time of ~100 ns exists for the leading edge of the bias voltage pulse to the VCSEL before which the VCSEL cannot be optically switched. It is assumed that this delay time arises due to parasitic capacitance of the pn-pn region peripheral to the VCSEL active region which does not contribute to the device bistability, but which must be charged at the beginning of the bias pulse. After the initial charging (on the leading edge of the VCSEL bias pulse) this parasitic capacitance does not limit VCSEL switching speed.

In summary, we have presented data which demonstrate optical switching and optical memory in a bistable VCSEL. These functions are thought to be important for such future applications as optical computing. The realization of these functions in a VCSEL is particularly attractive given the

natural extension of this device geometry to large area two-dimensional arrays.

REFERENCES

[1] D. G. Deppe, C. Lei, T. J. Rogers, and B. G. Streetman, "Bistability in a vertical-cavity surface-emitting laser," *Appl. Phys. Lett.*, vol. 58, pp. 2616–2618, 1991.

[2] T. Numai, M. Sugimoto, I. Ogura, H. Kosaka, and K. Kasahara, "Surface-emitting laser operation in vertical-to-surface transmission electrophotonic devices with a vertical cavity," *Appl. Phys. Lett.*, vol. 58, pp. 1250–1252, 1991.

[3] W. K. Chan, J. P. Harbison, A. C. von Lehmen, L. T. Florez, C. K. Nguyen, and S. A. Schwarz, "Optically controlled surface-emitting lasers," *Appl. Phys. Lett.*, vol. 58, pp. 2342–2344, 1991.

[4] C. Lei, T. J. Rogers, D. G. Deppe, and B. G. Streetman, "InGaAs-GaAs quantum well vertical-cavity surface-emitting laser using molecular beam epitaxial regrowth," *Appl. Phys. Lett.*, vol. 58, pp. 1122–1124, 1991.

[5] ——, "ZnSe/CaF_2 quarter-wave Bragg reflector for the vertical-cavity surface-emitting laser," *J. Appl. Phys.*, vol. 69, pp. 7430–7434, 1991.

Electronics Letters

IEE

AN INTERNATIONAL PUBLICATION

CONTENTS

9th April 1992 Vol. 28 No. 8

pages 701–808

ANTENNAS page

Direct synthesis of near-optimum difference patterns for planar arrays 753
E Botha and D A. McNamara *(South Africa)*

Performance of millimetre-wave frequency selective surfaces in large incident angle quasioptical systems 788
R Cahill and E A. Parker *(UK)*

CIRCUIT THEORY & DESIGN

High speed, current conveyor based voltage mode operational amplifier 742
E Bruun *(Denmark)*

Improved temperature compensation of OTAs 763
C. A. Karybakas C. Kosmatopoulos and Th. Laopoulos *(Greece)*

COMMUNICATIONS & SIGNAL PROCESSING

Baseband-processing feedforward carrier frequency drift compensation for DPSK mobile radio 771
M Ikura, K. Ohno and F. Adachi *(Japan)*

Compact image coding using two-dimensional DCT pyramid 791
K. H Tan and M. Ghanbari *(UK)*

Filter banks with rational decimation and interpolation rates 726
W H. Yim and F. P. Coakley *(UK)*

Neural network for singular value decomposition 784
A. Cichocki *(Poland)*

New class of polyphase pulse compression code with unique characteristics 769
T Felhauer *(Germany)*

Nonlinear continuous-phase frequency shift keying 758
R. Liyanapathirana, S. Le-Ngoc *(Canada)* and N. Ekanayake *(Sri Lanka)*

Performance evaluation of FDDI under asymmetric traffic conditions 710
S. Cusworth *(UK)*

Reducing multipath tracking errors in spread-spectrum ranging systems 729
D. J. R. van Nee *(The Netherlands)*

Resource auction multiple access (RAMA): efficient method for fast resource assignment in decentralised wireless PCS 799
N. Amitay *(USA)*

Waiting time jitter in FDDI to SONET payload mapping page 783
W. D. Grover and T Friesen *(Canada)*

COMPUTERS & LOGIC

Bit decision table for hypercube decomposition 736
C. H. Chang, H K Azzam and H. H. Dediu *(USA)*

Linear complexity of type of clock-controlled sequence 709
Liu Xian and Xiao Guozhen

ROM-based special purpose multiplication and its applications 718
H.-M Jong, L.-G Chen and T.-D Chiueh *(Taiwan)*

ELECTROMAGNETIC WAVES

New wire interface for graded 3-D TLM 728
A. J. Wlodarczyk and D. P. Johns *(UK)*

Surface wave propagation along system of parallel metallic strips and along unidirectionally conducting plane 762
M. L. Pereyaslavets *(Russia)*

Novel geodesic splitting on paraboloid of revolution with applications to ray-theoretic analysis 701
R. M. Jha, D. J. Edwards *(UK)* and V. Sudhakar *(India)*

FILTERS

New adaptive IIR notch filter and its application to howling control in speakerphone system 764
S. M. Kuo and J. Chen *(USA)*

Use of oversampled sigma-delta modulator for analogue adaptive FIR filters based on LMS algorithm 751
Huang Qiuting *(UK)* and G S. Moschytz *(Switzerland)*

INSTRUMENTATION & MEASUREMENT

Noncontact induced current impedance imaging 773
R. C. Tozer, J. C. Simpson, I. L. Freeston *(UK)* and J. M. Mathias *(USA)*

INTEGRATED CIRCUITS

Digit-pipelined on-chip clique-finding VLSI processor for real-time 3-D object recognition 722
T. Hanyu, T. Kodama and T. Higuchi *(Japan)*

(continued on back cover)

CONTENTS

(continued from front cover)

page

Multiple-valued pads for binary chips 794
A. J. Acosta, M. Valencia, A. Barriga, J. L. Huertas and M. J. Bellido *(Spain)*

INTEGRATED OPTICS

Codirectional grating-coupled filter for wavelength selective photodetection 749
H. Sakata, S. Takeuchi and T. Hiroki *(Japan)*

Composite rare-earth-doped glass waveguides 746
S. Honkanen, S. I. Najafi and W. J. Wang *(Canada)*

Integrated laser and add-drop optical multiplexer for narrowband wavelength division multiplexing 712
C. M. Ragdale, T. J. Reid, D. C. J. Reid, A. C. Carter and P. J. Williams *(UK)*

$LiNbO_3$ waveguide SHG device with ferroelectric-domain-inverted grating formed by electron-beam scanning 721
M. Fujimura, T. Suhara and H. Nishihara *(Japan)*

Monolithic integration of GaAs-waveguide optical intensity modulator with MESFET drive electronics 702
R. W. Ade, D. E. Bossi, R. P. Basilica and J. M. Berak *(USA)*

Monolithically integrated 2 × 2 InGaAsP/InP laser amplifier gate switch arrays 776
M. Janson, L. Lundgren, A.-C. Mörner, M. Rask, B. Stoltz, M. Gustavsson and L. Thylen *(Sweden)*

LASERS

Cascadability of optically latching vertical-cavity surface-emitting laser 734
D. L. Huffaker, D. G. Deppe, C. Lei, T. J. Rogers, B. G. Streetman, S. C. Smith and R. D. Burnham *(USA)*

Characteristic sideband instability of periodically amplified average soliton 806
S. M. J. Kelly *(UK)*

High power, singlemode InGaAs-GaAs-AlGaAs strained quantum well lasers with a new current blocking scheme using GaAs layers grown by molecular beam epitaxy at low substrate temperatures 803
Y. K. Sin, H. Horikawa, I. Matsuyama and T. K. Kamijoh *(Japan)*

Nd : YAG laser pumped picosecond Yb^{3+}/Er^{3+} fibre laser 766
A. B. Grudinin *(Russia)*, D. J. Richardson *(UK)*, A. K. Senatorov *(Russia)* and D. N. Payne *(UK)*

Novel principle of confinement in quantum-well structures 786
J. Barrau, B. Brousseau, M. Brousseau, R. J. Simes and L. Goldstein *(France)*

Proposal and realisation of new type distributed reflector (DR) laser with multiquantum-well structure 741
J. I. Shim, T. Nomaguchi, K. Kudo and S. Arai *(Japan)*

page

Spontaneous emission efficiency of compressively strained 1·5 μm MQW lasers 717
A. R. Adams, J. Braithwaite and V. A. Wilkinson *(UK)*

Strained InGaAs quantum well lasers grown on (111)B GaAs 705
I. W. Tao and W. I. Wang *(USA)*

MICROWAVE GUIDES & COMPONENTS

4–40 GHz MMIC distributed active combiner with 3 dB gain 739
R. Majidi-Ahy, C. Nishimoto, J. Russell, W. Ou, S. Bandy, G. Zdasiuk, C. Shih, Y. C. Pao and C. Yuen *(USA)*

MICROWAVE MEASUREMENTS & TECHNIQUES

Calculation of characteristic impedance of curved microstrip lines for MMIC and MHMIC 778
Y. Xu and R. G. Bosisio *(Canada)*

Collective-processing correlator system for thinned array imaging radiometer 779
K. Komiyama, Y. Kato, and T. Iwasaki *(Japan)*

Multimode oscillation and mode locking of magnetostatic wave delay line oscillator 789
S. N. Dunaev and Y. K. Fetisov *(Russia)*

NEURAL NETWORKS

Error and variance bounds on sigmoidal neurons with weight and input errors 760
D. Lovell, P. Bartlett and T. Downs *(Australia)*

Neural network synthesiser of pause duration for Mandarin text-to-speech 720
S.-H. Hwang and S.-H. Chen *(Taiwan)*

Neuron model with potential applications in pulse-stream neural nets 755
Q. Gan and Y. Wei *(China)*

OPTICAL COMMUNICATIONS

All-optical synchronisation with frequency division using selfpulsating laser diode 738
G. Farrell, P. Phelan and J. Hegarty *(Ireland)*

Demonstration of error-free soliton transmission over more than 15 000 km at Gbit/s single/channel and over more than 11 000 km at 10 Gbit/s in two-channel WDM 792
L. F. Mollenauer, E. Lichtman, G. T. Harvey, M. J. Neubelt and N. M. Nyman *(USA)*

Multifibre bus for rack-to-rack interconnects based on opto-hybrid transmitter/receiver array pair 801
J. W. Parker, P. J. Ayliffe, T. V. Clapp, M. C. Geear, P. M. Harrison and R. G. Peall *(UK)*

OPTICAL FIBRES

Passive, all-fibre source of 30 fs pulses 778
D. J. Richardson, A. B. Grudinin and D. N. Payne *(UK)*

Profiling of drawing induced defects in optical fibre preforms 768
G. R. Atkins, M. G. Sceats, S. B. Poole and H. W. Simmons *(Australia)*

(continued on inside back cover)

transistor: type A with a high emitter resistance and type B with a low emitter resistance. The common emitter current gain is 100 for both types of devices. The cause for the high emitter resistance for the type A device is a high contact resistance between the polysilicon and aluminium, which appeared from Kelvin structures on the same wafer. The type B devices were identically processed as the type A devices except for a tungsten-silicide layer between the aluminium and polysilicon. This greatly improves the contact resistance without affecting other transistor parameters.

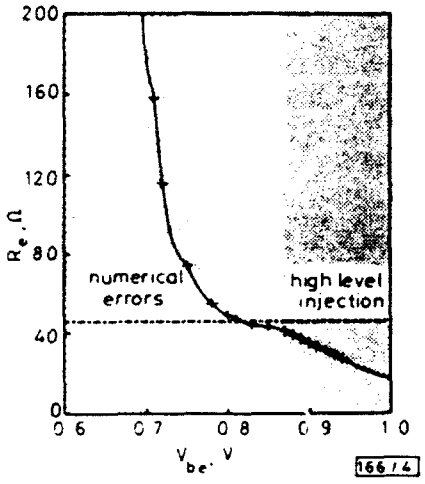


Fig. 4 *Emitter resistance as function of forward bias for type A transistor with emitter area of 80 × 5 μm²*

The measured emitter resistance for a type A device with an emitter area of $80 \times 5\,\mu m^2$ is plotted in Fig. 4. It illustrates that the calculated R_e is much too high for low V_{be} due to numerical errors, whereas the value at high V_{be} tends to be too low due to high injection effects. In the region in between, a rather constant value for the emitter resistance is found. Table 1 summarises the results obtained from the three different methods. However, the open collector method as well as the method of Ning and Tang led to ambiguous results. Therefore, the results had to be matched to the ones measured with the new method, which gave more reproducible results.

Table 1 MEASURED EMITTER RESISTANCES

Device	R_e New method	R_e Ning and Tang	R_e Open collector
	Ω	Ω	Ω
Type A $A_E = 5 \times 5\,\mu m^2$	200	?	160
Type A $A_E = 80 \times 5\,\mu m^2$	40	36	40
Type B $A_E = 5 \times 5\,\mu m^2$	22	15	15
Type B $A_E = 80 \times 5\,\mu m^2$	6	4	5

Conclusions: We have proposed a new method to determine the emitter resistance directly from the Gummel plot. This simple method is based on impact ionisation in the collector-base depletion region, which balances out the external base current.

20th February 1992

R. C. M. Wijburg and J. H. Klootwijk (*MESA Research Institute, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands*)

References

1 GETREU, I. E.: 'Modeling the bipolar transistor' (Tektronix Inc., Beaverton, Ore., 1976)

2 SANSEN, M. C., and MEYER, R. G.: 'Characterization and measurement of the base and emitter resistances of bipolar transistors', *IEEE J. Solid-State Circuits*, 1972, SC-7, pp. 492–498

3 NING, T. H., TANG, D. D.: 'Method for determining the emitter and base series resistances of bipolar transistors', *IEEE Trans. Electron Devices*, 1984, ED-31, pp. 409–412

4 WIJBURG, R. C., and RAGAY, F. W.: 'On the recombination in the quasi-neutral base of polysilicon emitter transistors with interfacial oxides', *Solid-State Electron.*, 1991, **34**, pp. 1469–1471

5 SAKUI, K., HASEGAWA, T., FUSE, T., SESHITA, T., ARITOME, S., WATANABE, S., OHUCHI, K., and MASUOKA, F.: 'A new reverse base current (RBC) of the bipolar transistor induced by impact ionization', *Jpn. J. Appl. Phys.*, 1989, **28**, pp. L2150–L2152

CASCADABILITY OF OPTICALLY LATCHING VERTICAL-CAVITY SURFACE-EMITTING LASER

D. L. Huffaker, D. G. Deppe, C. Lei, T. J. Rogers, B. G. Streetman, S. C. Smith and R. D. Burnham

Indexing terms: Lasers, Semiconductor lasers

The cascadability of an optically latching vertical-cavity surface-emitting laser is investigated through the use of a wavelength-tunable optical probe. The bistable vertical cavity laser emits at a wavelength of 0·98 μm, and shows a sensitivity peak to probe light generated from a grating-tunable edge-emitting laser in the 0·98 μm wavelength range. The sensitivity peak occurs in a window of ~50 Å around the vertical-cavity lasing wavelength, and corresponds to the Fabry-Perot transmission window of the surface-emitter.

Optical latching is an important function for the realisation of optical memory. Several reports of various schemes to achieve optical latching in vertical-cavity surface-emitting lasers (VCSELs) have been recently reported [1–4]. The realisation of optical memory in the VCSEL is particularly attractive because the device geometry is suitable for 2-dimensional arrays, and allows convenient optical addressing. Demonstrations of wavelength tuning across a 2-dimensional array may also be of use in the parallel transmission of optical data [5]. In the future it may be feasible to cascade optically latching VCSELs of similar wavelength. This operation may be contrasted with previously demonstrated optical switching and memory using a VCSEL probed at a wavelength outside of the Fabry-Perot reflectivity curve [2]. It has recently been emphasised that monolithic integration of the VCSEL with a separate *pnpn* switch provides cascadability due to the wide wavelength range over which the *pnpn* can be switched. This may prove to be an inherent advantage over structures in which the *pnpn* region is incorporated directly into the device active region [3, 4]. However, the integration of the *pnpn* switching layers directly into the VCSEL active region offers advantages due to the ease of device fabrication, potential for improved device speed, and denser device packing. In addition, wavelength sensitivity in the VCSEL switching may also be a desirable characteristic in some applications. These considerations have led us to investigate the switching characteristics of a bistable VCSEL using an optical probe beam with a wavelength near the lasing wavelength of the VCSEL. Optical switching is demonstrated at wavelengths inside the stopband of the Fabry-Perot cavity, with an increased sensitivity near the Fabry-Perot resonance.

Details of the device structure and fabrication have been described previously [8]. The VCSEL emits at a wavelength of 0·98 μm and demonstrates hysteresis in both light against current and current against voltage characteristics. The hysteresis in the current against voltage characteristics makes it possible to biss the VCSEL in the hysteresis window and optically switch the device from a nonlasing state to a lasing state [2]. A schematic diagram of the optical setup used to investigate the VCSEL switching sensitivity to the probe wavelength is shown in Fig. 1. The optical probe consists of an edge-emitting laser with a single 100 Å $In_{0.2}Ga_{0.8}As$ quantum well active region, with one of the facets coated with

an antireflecting film. The probe wavelength is selected through the use of an external grating [7]. The laser is tunable from 0·96 to 1·0 μm. Tuning of the probe laser with the external grating occurs in mode hops of ~2·0 Å, which corresponds to weak reflections from the edge-emitting Fabry-Perot cavity. Light emitted from the rear facet of the probe laser is focused onto the 10 μm active region of the VCSEL. The VCSEL is mounted on a thermoelectric cooler (TEC) which maintains an operating temperature of ~285 K. The VCSEL and probe output signals are each collected by separate Ge photodiodes.

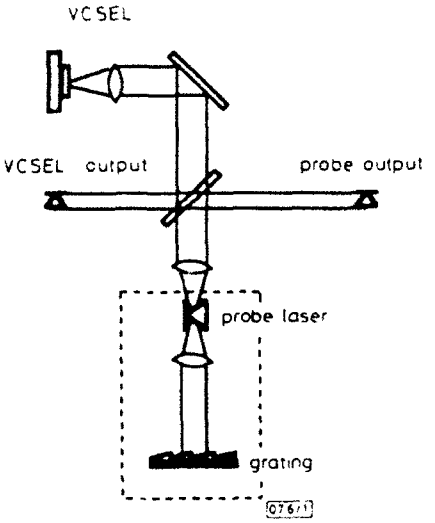


Fig. 1 *Schematic diagram of experimental setup used to demonstrate cascadability of VCSELs using wavelength-tuneable edge-emitting probe laser*

Wavelength is selected by rotating external diffraction grating to provide positive feedback to active region probe laser

Switching of the bistable VCSEL with ~0·98 μm wavelength light is shown in Fig. 2, where four traces represent electrical bias pulses to both the VCSEL and the probe, and the corresponding light output from each device. The upper trace shows the 5 μs pulse to the VCSEL, which biases the device just below lasing threshold in the hysteresis window [8]. During the 5 μs pulse to the VCSEL, a 0·25 μs bias pulse is applied to the probe laser which drives the probe laser above threshold, as indicated by the two traces labelled probe current and probe output. The light pulse produced by the probe laser at a wavelength 0·98 μm is seen to switch the VCSEL into its lasing state, where it remains for the duration of the electrical biasing pulse to the VCSEL as shown by the trace labelled VCSEL output.

There is significant sensitivity of the VCSEL switching response to the probe laser wavelength in this wavelength range, due to the transmission characteristics of the VCSEL Fabry-Perot cavity. Fig. 3 shows the relative probe intensity against wavelength needed to switch the VCSEL. The VCSEL emission spectrum above threshold is also shown, to specify its spectral location with respect to the sensitivity peak. The data indicate that the VCSEL is most sensitive to the probe beam near its lasing wavelength, where the probe intensity corresponds to ~3·5 mW. The probe power required for switching, therefore, is significant compared to the power emitted by the VCSEL just above threshold which is ~0·2 mW [2]. However, we have observed selfswitching of the VCSEL in which the light emitted from the device below threshold is coupled back into the VCSEL by means of an external reflector, and triggers the switching process. If this external reflector is considered as an external probe source, the light power required for switching is extremely small. It is not clear at this time what the exact mechanism of this self-switching is, and whether the switching is either extremely wavelength sensitive beyond our tuning resolution of wavelength with the external probe source, or possibly phase sensitive.

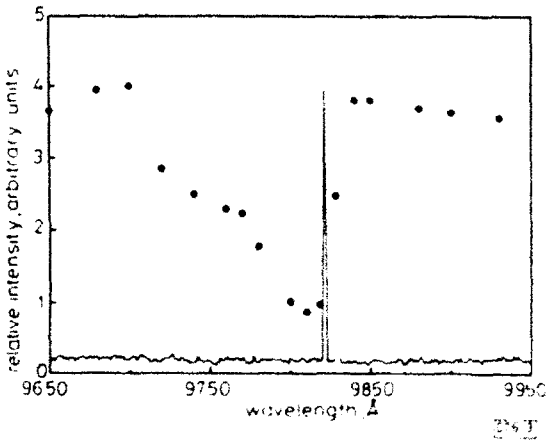


Fig. 3 *Relative probe intensity (data points) necessary to optically switch VCSEL, as function of wavelength*

VCSEL lasing spectrum is superimposed on same wavelength scale for comparison

To investigate the mechanism of optical switching when the external probe laser is used, we have studied the sensitivity of the current-voltage characteristics to wavelength of the isolated *pnpn* structure which is integrated with the VCSEL [8]. For this case a region of the *pnpn* structure is isolated from the VCSEL active region by etching into the epitaxial layers creating a deep current blocking trough. Although the current-voltage characteristics of the isolated *pnpn* layers are quite sensitive to probe wavelengths of ~0·79 μm, we have found that they are insensitive to 0·98 μm light under the forward biasing condition used in the VCSEL switching experiment. This suggests that the switching due to the 0·98 μm probe light shown in Fig. 2 is likely triggered by absorption of the probe light in the InGaAs quantum well active region sufficient to drive the VCSEL above threshold.

In summary, we have presented data which demonstrates the cascadability of a optically latching VCSEL at its lasing wavelength, and the sensitivity of the switching to probe light at other wavelengths. These functions should be useful in future applications such as optical computing and memory

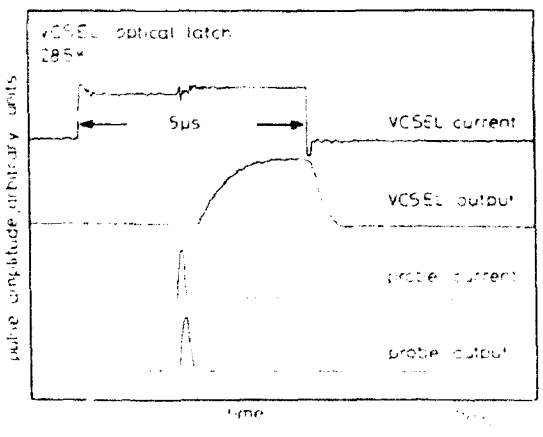


Fig. 2 *Oscilloscope traces showing biasing to VCSEL and probe laser, along with optical output and switching behaviour*

Acknowledgment: This work has been supported by the Office of Naval Research Young Investigator Award, the National Science Foundation Presidential Young Investigator program, the Joint Services Electronics Program under Contract No. AFOSR F49620-89-C-0044, and the Army Research Office under Contract No. DAAL-03-91-6-0334.

11th February 1992

D. L. Huffaker, D. G. Deppe, C. Lei, T. J. Rogers and B. G. Streetman (*Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712-1084, USA*)

S. C. Smith and R. D. Burnham (*Amoco Research Center, Amoco Technology Company, Naperville, Illinois 60566, USA*)

References

1 NUMAI, T., SUGIMOTO, M., OGURA, I., KOSAKA, H., and KASAHARA, K.: 'Surface-emitting laser operation in vertical-to-surface transmission electrophotonic devices with a vertical cavity', *Appl. Phys. Lett.*, 1991, **58**, pp. 1250–1252

2 HUFFAKER, D. L., LEE, W. D., DEPPE, D. G., LEI, C., ROGERS, T. J., CAMPBELL, J. C., and STREETMAN, B. G.: 'Optical memory using a vertical-cavity surface-emitting laser', *Photonics Technol. Lett.*, 1991, **3**, pp. 1064–1066

3 CHAN, W. K., HARBISON, J. P., VON LEHMEN, A. C., FLOREZ, L. T., NGUYEN, C. K., and SCHWARZ, S. A.: 'Optically controlled surface-emitting lasers', *Appl. Phys. Lett.*, 1991, **58**, pp. 2342–2344

4 ZHOU, P., CHENG, J., SCHAUS, C. F., SUN, S. Z., HAINS, C., ZHENG, K., ARMOUR, E., HSIN, W., MYERS, D. R., and VAWTER, G. A.: 'Cascadable, latching photonic switch with high optical gain by the monolithic integration of a vertical-cavity surface-emitting laser and a pn-pn photothyristor', *Photonics Technol. Lett.*, 1991, **3**, pp. 1009–1012

5 CHANG-HASNAIN, C. J., WULLERT, J. R., HARBISON, J. P., FLOREZ, L. T., STOFFEL, N. G., and MAEDA, M. W.: 'Rastered, uniformly separated wavelengths emitted from a two-dimensional vertical-cavity surface-emitting laser array', *Appl. Phys. Lett.*, 1991, **58**, pp. 31–33

6 LEI, C., ROGERS, T. J., DEPPE, D. G., and STREETMAN, B. G.: 'InGaAs/GaAs quantum well vertical-cavity surface-emitting laser using molecular beam epitaxial regrowth', *Appl. Phys. Lett.*, 1991, **58**, pp. 1122–1124

7 EPLER, J., HOLONYAK, JR. N., BURNHAM, R. D., PAOLI, T. L., and STREIFER, W.: 'Supermodes of multiple-stripe quantum-well heterostructure laser diodes operated (cw, 300 K) in an external-grating cavity', *J. Appl. Phys.*, 1985, **57**, pp. 1489–1493

8 DEPPE, D. G., LEI, C., ROGERS, T. J., and STREETMAN, B. G.: 'Bistability in a vertical-cavity surface-emitting laser', *Appl. Phys. Lett.*, 1991, **58**, pp. 2616–2618

BIT DECISION TABLE FOR HYPERCUBE DECOMPOSITION

C. H. Chang, H. K. Azzam and H. H. Dediu

Indexing terms: Parallel processing, Computers

A new algorithm for subcube assignment and decomposition is introduced to solve the fragmentation problem in hypercube (*n*-cube) processor allocation. The sharing density vector introduced in the prime cube graph strategy [1] is used as a vehicle to develop the new algorithm. A bit decision table is devised to provide the key parameters for determining the decomposition. It is proven that the result of the algorithm is at least 22% better than the sharing density vector approach.

Terms and definitions: Let α be a prime cube of dimension k in a given configuration C_d. We denote the positions of bits in α that correspond to *s by $s_1, \ldots, s_k$. Assume also that α is connected to each element of the set $C_\alpha = \{\beta_1, \ldots, \beta_l\}$, where C_α is a subset of C_f, and β_is are prime cubes. The zero index, one index and star index of α at bit s_i, denoted by $ZERO_\alpha(i)$, $ONE_\alpha(i)$ and $*_\alpha(i)$, are integers equal to the numbers of prime cubes in C_α that contain 0, 1 and *, respectively, in bit s_i.

The bit decision table provides a simplified view of zero and one patterns in a binary subcube representation. The first row consists of the bits of the target prime cube. For each prime cube adjacent to the target, a row consisting of the bits of the prime cube is added to the table. What follows is the penultimate row (I) consisting of two indication numbers and one indication symbol per column providing a summary of the zero, one, and star statistics in that column. The first indication number is min ($ZERO_\alpha(i)$, $ONE_\alpha(i)$). Following the indication number is a symbol denoted the split, which is equal to 0 if $ZERO_\alpha(i) < ONE_\alpha(i)$, 1 if $ZERO_\alpha(i) > ONE_\alpha(i)$, or * otherwise. The last indication number is a count of the stars in the corresponding column, $*_\alpha(i)$. Finally, the last row (S) contains the sum of the first and last indication numbers of each

Example 1: Let $C_f = \{\alpha, \beta, \gamma\}$, where $\alpha = ***00*$, $\beta = 0010**$, $\gamma = 00**01$, and a task requiring a subcube of dimension 3 arrives. The following bit decision table is generated:

α	*	*	*	0	0	* (Target)
β	0	0	1	0	*	*
γ	0	0	*	*	0	1
I	010	010	001			001
S	0	0	1			1

For example, in the first column, the first indication number is 0, followed by the indication symbol 1 and the indication number 0, which means that in the last row the select value is 0 + 0 = 0. The select value is a better indicator in selecting a subcube from a prime cube than the sharing density vector.

Star search algorithm: An algorithm (we choose to call star search) for determining the subcube that will be assigned to the ask is discussed in this Section. The algorithm proceeds as follows:

(1) determine the zero index ($ZERO_\alpha(i)$), the one index ($ONE_\alpha(i)$), and the star index ($*_\alpha(i)$) of each bit of the target prime cube α that contains a star

(2) find the select values

(3) choose the bits with the least select values as targets for splitting and find the split at that position

(4) substitute each bit selected in step 3 with its corresponding split (if *, choose randomly); assign the resulting subcube to the task (this is one of 2^r subcubes, where r = dimension of the target prime cube-dimension of the task).

Consider the configuration $\{\alpha, \beta, \gamma\}$, where $\alpha = 00**00$, $\beta = 0**00*$, $\gamma = *0*1*0$. For a task requiring a subcube of dimension 1, the following bit decision table is formed:

α	0	0	*	*	0	0 (Target)
β	0	*	*	0	0	*
γ	*	0	*	1	*	0
I			0*2	1*0		
S			2	1		

The smaller of the two select numbers is in column 4. Hence bit 4 should be selected as target for decomposition. Now we have to choose from one of the two following subcubes: (00*000) and (00*100). The second indication bit in the selected column is *, which means either of these two subcubes can be assigned to the incoming task.

The three indication numbers are the key to deciding which bit should be decomposed and which subcube is allocated to the task. The select value is the number of prime cubes lost in case the decomposition is carried out at the corresponding bit in the target and the task is assigned to the subcube with the second indication number (the split). Therefore, minimising this select value assures the minimum loss of prime cubes.

Performance evaluation: We prove in this Section that the result of our algorithm is 22% or more better than the sharing density vector approach [1]. Consider the ith and jth bits of the target prime cube α. For example, let us assume that the number of prime cubes adjacent to α is 10. Also assume that the total number of ones at bit i in all the adjacent prime cubes is 4, the number of zeros is 2, and the number of stars is 4. At the same time assume that the total number of ones at bit j in all the adjacent prime cubes is 5 and the number of zeros is 5. According to the SD vector method we should choose i although a better decision to choose j is detected by the bit decision table.

When using the bit decision table as opposed to the SD vector method an improvement occurs with the probability:

$P[V]$ = (probability of choosing i when using the SD vector method and choosing j when the bit decision table is

TWENTIETH ANNUAL CONFERENCE
PHYSICS AND CHEMISTRY of SEMICONDUCTOR INTERFACES

January 25-29, 1993

Colonial Williamsburg Hotel
Williamsburg, Virginia

ORGANIZING AND PROGRAM COMMITTEE

L. J. Brillson (Xerox), Chairman
H. F. Dylla (CEBAF), Local Arrangements Chairman
S. A. Chambers (Battelle Pacific Northwest Laboratory)
L. C. Feldman (AT&T Bell Labs)
R. W. Grant (Rockwell)
F. J. Grunthaner (JPL/CalTech, Past Chairman)
M. S. Hybertsen
A. Kahn (Princeton)
T. C. McGill (CalTech)
D. J. Wolford (IBM Research)
C. G. Van de Walle (Xerox)
J. M. Woodall (IBM Research, Past Chairman)

EX OFFICIO

L. R. Cooper (ONR)
H. R. Wittmann (AFOSR)

CONFERENCE SECRETARY

B. J. Tegas (Xerox)

PROCEEDINGS

T. C. McGill (CalTech), Special Editor
M. Hudson, Editorial Assistant

SPONSORS

American Vacuum Society
American Physical Society
Office of Naval Research
Air Force Office of Scientific Research
Xerox Corporate Research and Technology

Observation of room temperature quantum confined Stark shift in $Si_{1-x}Ge_x$/Si multi-quantum wells

Andalib A. Chowdhury, M. Mahbub Rashed, C. M. Maziar, S. S. Murtaza and J. C. Campbell

Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712 . (512) 471-8838

Electric fields applied perpendicular to quantum wells are observed to shift the associated electron transition energies and optical absorption edges of the structure. This effect is known as the quantum confined Stark effect or QCSE. Although the QCSE has been studied extensively for III-V compound semiconductor systems[1], observation of the QCSE in $Si_{1-x}Ge_x$/Si multi-quantum wells (MQW) has rarely been cited[2,3]. Recently, room temperature photocurrents for *p-i-n* photodiodes containing an intrinsic region comprised of $Si_{1-x}Ge_x$ MQW's grown by RPECVD (remote plasma enhanced chemical vapor deposition) have been reported[3]. The objective of this work is to report the observation of a room temperature Stark shift in these $Si_{1-x}Ge_x$/Si multi-quantum wells. The absorption edges estimated from the measured photocurrent of the reverse biased *p-i-n* diodes show large, linear shifts of the transition energies of the MQWs under applied electric field. The experimentally obtained shift of the transition energies matches very well with that calculated from the envelope function approximation method[5].

Figure 1 shows a schematic cross-section of the device studied in this work. MQWs grown on unstrained $Si_{0.8}Ge_{0.2}$ buffer layers are used to allow the strain to be distributed symmetrically in $Si_{0.6}Ge_{0.4}$ and in Si layers. This allows thick MQW structures to be grown without the critical thickness limitation over the entire structure thickness. The band alignment of these structures is staggered type II (i. e. the band edges of the narrower band gap $Si_{0.6}Ge_{0.4}$ lie above the respective band edges of the wider gap Si (Fig. 2)). Under applied electric field $(\mathcal{E})$, the change in transition energy shift may be expressed as $\Delta E_l + \Delta HH_1 \pm e\mathcal{E}L$ ($\Delta E_l + \Delta HH_1$ is the shift of minimum bound state energies). The last term represents the potential drop between the center of two adjacent layers and dominates the expression. The total energy shift is nearly linear, leading to large transition shifts[4]. Figure 3 shows the measured photocurrents as a function of photon energy under different electric fields. To eliminate the effect of increased carrier transport due to increased bias, the photocurrent spectra were normalized with respect to the wavelength 1190nm (1042meV), where any changes in the signal with bias had very small contributions from the QCSE. For the same incident optical power, the photocurrent (I) is proportional to $(1-\exp(-\alpha L))/h\nu$, where α is the absorption coefficient, L is the MQW thickness and hν is the photon energy. For $L \ll 1/\alpha$, α is proportional to Ihν. Plotting Ihν as a function of hν (Figure 4), it is observed that near 800 meV (the estimated transition energy of the MQW), α varies as $(h\nu - Eg_o)^{1/2}$ for different applied electric fields. From this fit, the absorption edge (Eg_o) may be estimated at different electric fields. A linear edge shift towards lower energies with increased electric field may be observed in Figure 5. Figure 5 illustrate the good agreement found between the experimentally obtained shift and that calculated within the framework of the envelope function approximation[5], neglecting intervalley interaction. The bandoffsets were estimated from the combination of self-consistent *ab initio* pseudopotential results[6] and the phenomenological deformation potential theory[7]. To understand the spectra at higher energies, we have subtracted the $(h\nu - Eg_o)^{1/2}$ dependency from the measured data and obtained the absorption shape of the $Si_{0.8}Ge_{0.2}$ buffer layer (Figure 6). The shape obtained in our work is similar to that of Braunstein, et al. [8]. It is interesting to note here that the square root dependence of absorption coefficient on photon energy (which is incidentally characteristic of the bulk direct band gap material) is different from that of ref. 2, where Park et al. observed a square dependence. We believe that more experimental information is needed to understand the recombination mechanism in type II $Si_{1-x}Ge_x$/Si heterostructures.

In summary, we have demonstrated large linear quantum confined Stark shift in $Si_{1-x}Ge_x$/Si MQWs at room temperature which may prove to be useful in optoelectronic device applications.

Acknowledgement : This work was supported in part by the joint Services Electronics Program, F49-620-92-C-0027 and the Science and Technology Center Program of the National Science Foundation, NSF grant No. CHE-8920120.

References

[1] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, Physical Rev. B 32, 1043 (1985).

[2] J. S. Park, R. P. G. Karunasiri, and K. L. Wang, J. Vac. Sci. Technol. B 8, 217 (1990).

[3] S. S. Murtaza, R. Qian, D. Kinosky, R. Mayer, A. Tasch, S. Banerjee and J. C. Campbell to be published.

[4] Andalib A. Chowdhury, K. H. Jung, D. L. Kwong, and C. M. Maziar, J. Appl. Phys. 70, 3946 (1991).

[5] G. Bastard, and J. A. Brum, J. Quantum Electronics 22, 1625 (1986).

[6] C. G. Van de Walle and R. M. Martin, J. Vac. Sci. Technol. B 4, 1055 (1986).

[7] R. People, Physical Rev. B 34, 2508 (1986).

[8] R. Braunstein, A. R. Moore, and F. Herman, Phys. Rev. 109, 695 (1958).

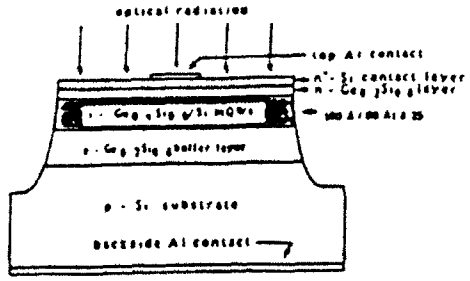


Fig. 1: A schematic cross-section of the device studied.

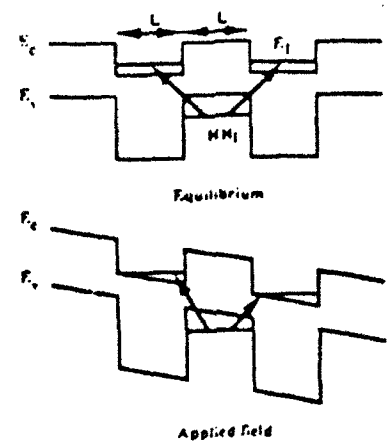


Fig. 2: QCSE in $Si_{1-x}Ge_x$ /Si MQW. E_1 and HH_1 are the minimum electron and hole energy levels in the quantum well.

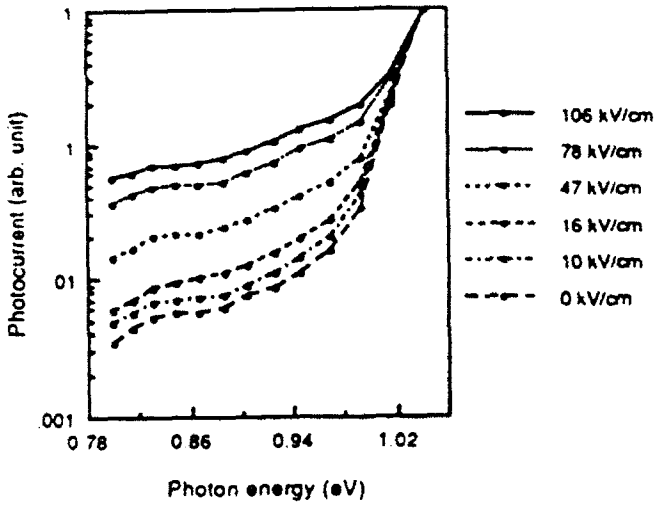


Fig. 3: The measured room temperature photocurrents as a function of photon energy under different estimated electric fields.

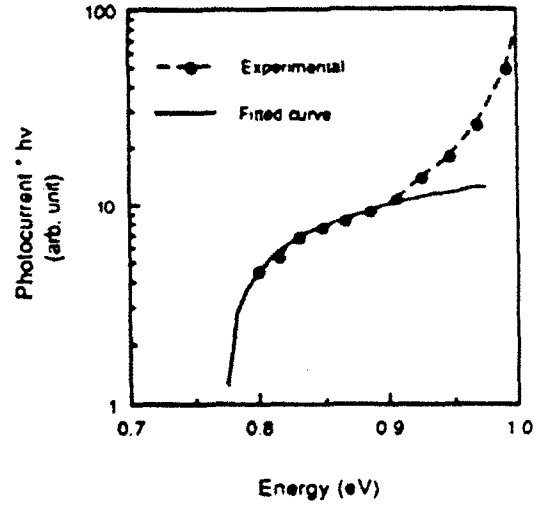


Fig. 4: Typical example of the curve fitting procedure at lower energies. The applied electric field in this case was 16 kV/cm.

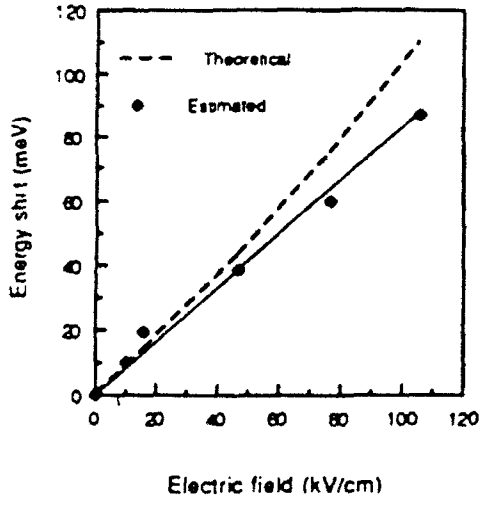


Fig. 5: Magnitude of the absorption edge shift as a function of electric field. The reference point was taken at zero electric field.

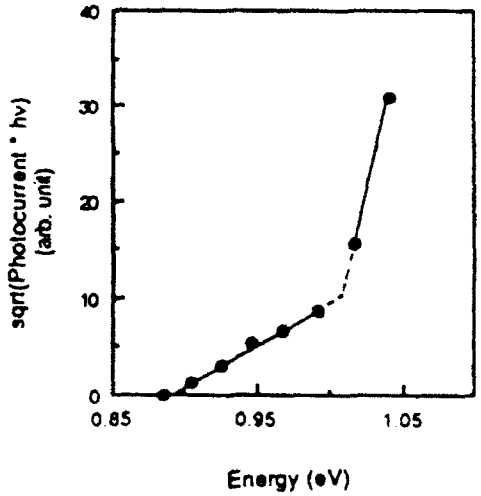


Fig. 6: Sqrt of (photocurrent * hv) as a function of photon energy. These points were obtained by subtracting the $(hv-Eg_0)^{1/2}$ dependency (Fig. 4) from the measured results.

PROCEEDINGS REPRINT

SPIE—The International Society for Optical Engineering

Reprinted from

Quantum Well and Superlattice Physics IV

23–24 March 1992
Somerset, New Jersey



Volume 1675

©1992 by the Society of Photo-Optical Instrumentation Engineers
Box 10, Bellingham, Washington 98227 USA. Telephone 206/676-3290.

Modeling of the surface states of a terminated superlattice using a muluband Hamiltonian

Andalib A. Chowdhury and C. M. Maziar

Microelectronics Research Center, Department of Electrical and Computer Engineering,
University of Texas at Austin, Austin, TX 78712

ABSTRACT

In most electronic bandstructure calculations of semiconductor superlattices (SLs), periodically extended boundary conditions are used. However the presence of a terminating layer, which breaks the periodicity of the SL, is expected to modify the electronic bandstructure. In this work, we report a model for calculating the localized surface-like states of a terminated SL which includes, for the first time, the nonparabolic nature of the host bandstructures by way of band mixing, unlike the simple one band Kronig-Penney type model. This model is within the framework of the multiband envelope function approximation, which has proven to be a useful and efficient method for the calculation of SL electronic bandstructure.

1. INTRODUCTION

Semiconductor superlattices (SLs) are obtained by growing alternately, layers of different host semiconductors. In the SL, the charge carriers see in addition to the normal crystal potential, a periodic potential which is due to a series of quantum wells (QWs) and barriers due to the presence of band offsets and energy gap differences at the heterojunction. When there is a considerable amount of overlap of wave functions between neighboring QWs, bands of allowed energy levels of delocalized states are formed with the simultaneous formation of forbidden minigaps (MGs). Superlattices are important model systems to study bulk crystal properties, due to the correspondence between a miniband of a SL and a band of energy of a bulk crystal. These artificial structures are currently of great technological interest and importance because of their promise for flexible tuning of the electronic structure of the resulting system by controlled modifications of parameters such as layer thicknesses, alloy compositions, strain, growth axis etc.

In order to understand the origin of the fundamental physical phenomena and other interesting optical and transport properties the SLs exhibit, it is essential to calculate their electronic band structure. In most band structure calculations of SLs, periodically extended boundary conditions are generally used and the SL is looked upon as an infinite, uniform periodic structure, so that it is readily amenable to theoretical treatment. In this picture there are allowed bands of extended states separated by forbidden gaps. The analogous picture for real bulk solid is also the same. But, in a real crystal, due to the presence of the surface, a perturbation of the band structure at the surface is expected. In the 1930s Tamm[1] and Shockley[2] modeled the presence of surface boundaries in solids and showed the presence of states inside the forbidden gaps at the surface. These were called surface states. Although surface science is one of the most fertile fields in physics, these surface states have not been observed in the pure form Tamm and Shockley predicted.[3] This was due to the simultaneous presence of many effects (such as the presence of foreign atoms) in real surfaces apart from the cases Tamm and Shockley studied. Also the deviation of the surface potential from that in the bulk was still unknown. Similarly, in the semi-infinite SL picture where the periodicity of a SL is broken by the addition of an interface, localized surface-like states occur at the interface. Recently Ohno[3] et al. have experimentally verified the existence of such surface states by simulating an abrupt higher barrier internal "surface" by terminating a GaAs/AlGaAs SL with an AlAs barrier. They observed, for the first time, the presence of localized "Tamm"-like "surface" states (called SL surface states) inside the minigaps.

 0-8194-0836-0/92/$4.00

We believe superlattices can be an excellent experimental tool to study different surface-like behavior (and to verify different theories of surface states) in a periodic structure such as a semiconductor crystal. This is due to the fact that SLs allow the construction of an arbitrary model surface for a particular study. As an example, the GaAs/AlGaAs system allows one to realize and manipulate a wide range of one dimensional potential profiles to simulate a model surface in a controlled manner.

2. MODEL

Recently Steslicka et al.[4] have proposed a simple model to calculate these SL surface states. This model is an extension of the one band effective mass type method in which the bandstructures of the host materials are taken to be parabolic. However this approach fails to describe even the band structure of infinitely long InAs-GaSb SL[5] in which the valence band maximum of GaSb is higher than the conduction band minimum of InAs. Even for wide band gap GaAs-AlGaAs material systems, the inclusion of band nonparabolicity has been found to be the primary reason for improved agreement with the experimental results.[6] Therefore for accurate determination of energy levels, band mixing i.e., band nonparabolicity should be taken into account. In this work, we report the calculation of the SL surface states of a terminated SL by including the nonparabolic nature of the host bandstructures by way of band mixing. We found that the results differ significantly from those using the simple parabolic band model. The approach used here is similar to that of the one band model but the Hamiltonian and the boundary conditions used as well as the final results obtained are substantially different as will be shown below.

Let us suppose that a SL is composed of host materials A and B of thicknesses L_A and L_B stacked alternately one on the other and the terminated well region is ended by a higher potential V_{sc} (region C), perhaps representing the vacuum region (Figure 1). Since the multiband Kane model[7] incorporates both the conduction and the valence bands taking into account the band mixing, it is most useful for describing the band structure of the host materials. This method has been found to give more accurate match with experimental results than any other similar methods which neglect the band nonparabolicity.[6] Using this Hamiltonian (i.e., following Bastard's envelope function approximation formalism), the SL dispersion relationship is given by

$$\cos(qd) = \cos(k_A L_A)\cos(k_B L_B) - \frac{1}{2}\left(\beta + \frac{1}{\beta}\right)\sin(k_A L_A)\sin(k_B L_B)$$

$$E - V_{S_i} = \frac{k_i^2 P^2}{3}\left[\frac{2}{E - V_{P_i}} + \frac{1}{E - V_{O_i}}\right] \qquad i = A, B, C \tag{1}$$

$$\beta = \frac{k_A}{k_B}\left[\frac{2}{E - V_{PA}} + \frac{1}{E - V_{OA}}\right]\left[\frac{2}{E - V_{PB}} + \frac{1}{E - V_{OB}}\right]^{-1}$$

$$d = L_A + L_B$$

$$BCM_i = \frac{2}{E - V_{P_i}} + \frac{1}{E - V_{O_i}} \qquad i = A, B, C$$

where q is the SL wave vector, m_i is the band edge effective mass, V_{si},V_{pi} and V_{oi} are the conduction band minimum, light hole band maximum and the split-off band maximum respectively of the i layer (i=A,B,C), P is the momentum matrix element and E is the electron energy. This method is a generalization of the one band parabolic Kronig-Penney type model applied to a multiband case. The boundary condition used are the continuity of the envelope function and the first derivative of the envelope function divided by BCM_i at a hetero-interface. The last boundary condition reduces for the case of one band parabolic model to the first derivative of the envelope function divided by the effective mass at the band extrema

point. When a SL with cyclic boundary conditions is considered, we can apply on the envelope functions $\Phi(z)$, the Bloch condition $\Phi(z+jd)=\exp(iqjd)\Phi(z)$. Since the envelope functions must describe an electron having equal probability of being found in any quantum well of the SL, all the components of q have to be real. But when the SL is terminated, the complex values of the SL wave vector q may describe states that are physically possible. The envelope function of this type decreases exponentially as one moves away from the terminating interface towards the SL, and therefore corresponds to a localized state. Suppose in the nth MG q is given by $i\,q_i + n\,\pi/d$ ($q_i > 0$, n=0,1 ,2,3,.....). For complex q, the L.H.S of (1) becomes $(-1)^n \cosh q_i d$.

From the transfer matrix approach[8] and the Bloch theorem, the wave function at the first well layer is known apart from a normalization coefficient. Assuming an exponentially decaying function in region C (higher barrier) and applying the above mentioned boundary conditions at the interface of the terminated well layer and region C, we obtain:

$$(-1)^n exp(-q_i d) = \frac{sin(k_A L_A) + C1cos(k_A L_A)}{C1cosh(k_B L_B) - C2sinh(k_B L_B)} \tag{2}$$

$$C1 = \frac{k_A * (BCM_A)}{k_C * (BCM_C)} \qquad , \qquad C2 = \frac{k_A * (BCM_A)}{k_B * (BCM_B)}$$

Now eliminating q_i from (1) and (2) gives us the surface states in the nth MG. This solution depends on the depth and the width of the QWs and on the effective masses at the extrema points of the bulk band structures of the host materials.

3. RESULTS

Figure 2 shows the calculated energy levels of InAs-GaSb SL as a function of the SL period. Hatched regions indicate the allowed energy bands as was calculated in the now classic paper by Bastard.[5] The important result is the presence of the SL surface states inside the MGs. Each MG has two energy levels due to the presence of a terminating layer (taken as vacuum). The shape of these SL surface states closely follow those of the allowed energy bands. The values of the input parameters used in this calculation were taken from[5] and the value of V_{sc} used was 4.55 eV. To make comparison with the earlier report,[4] we have also calculated the SL surface states of a terminated GaAs-$Al_xGa_{1-x}As$ SL as a function of the mole fraction x as shown in Figure 3. The barrier height of the internal quantum wells changes as the mole fraction is changed. As before the hatched regions indicate the allowed energy bands and the dotted lines, the surface states. If these results are compared with those calculated from the simple one band parabolic model, we notice a downward shift of the allowed energy bands as expected,[8] and consequently a shift of the SL surface states. In addition to this, there are two energy levels found from the nonparabolic model as opposed to one found from the parabolic model. In Figure 4, the effect of the variation of the barrier width (keeping the well width constant) and well width (barrier width constant) is shown. In both the cases, the presence of the SL surface states are shown inside the MGs. In Figure 5, a comparison is made between the SL surface states calculated from a one band parabolic model and the multiband nonparabolic model. The result clearly shows that the inclusion of nonparabolicity significantly affects the energy level positions of the SL surface states particularly at the higher lying minigaps. The terminating barrier height V_{sc} was taken as 4.07 eV for GaAs-$Al_xGa_{1-x}As$ SLs in all the above cases. In Figure 6, the effect of changing the terminating barrier height on the SL surface states are shown. For terminating barrier heights greater than the internal barrier height, the SL surface states are found above the minibands and for terminating barrier heights less than that of the internal barriers the surface states are found below the minibands, the details of which will be published elsewhere. It should be mentioned here that this result is expected from the tight-binding picture.[3] In this picture, the isolated quantum well energy levels broaden into a miniband as the coupling between the wells is turned on. But the end quantum well state which has a different energy level from that of the internal quantum wells (due

to the difference in barrier heights) does not participate in the formation of the minibands and it becomes a surface state. So, for terminating barrier higher (lower) than the internal quantum well barrier, the surface state is at higher (lower) energy than the minibands. Interestingly, we found that one of the SL surface states in each MG does not vary with the variation of the terminating potential. The other SL surface state disappears as the height of the terminating potential approaches that of the internal barriers. In all these calculations, a two band model using the conduction band and the light hole valence bands were used and only the electron energy levels are shown in the figures.

4. CONCLUSIONS

In summary, we have extended the envelope function approximation model to calculate the SL surface electronic states of a terminated SL to the case of multiband Hamiltonian to incorporate the nonparabolic nature of the band structure of the host materials. The surface electronic states derived from the multiband nonparabolic band structure model depart significantly from those reported for the one band parabolic model. Also, two SL surface states are obtained from the nonparabolic model rather than one found from the parabolic band model. Therefore band mixing is important for these SL surface states. Although, only InAs-GaSb and GaAs-AlGaAs SLs have been considered here, this model can be applied to any material system such as lattice matched $Ga_{0.47}In_{0.53}As$-InP SL terminated by $Al_{0.48}In_{0.52}As$ barrier. This method can also be extended for the case of SLs composed of complicated bases (triangular, parabolic, polytype[7] etc.). We believe that this model should be useful for accurate modeling of transport across a terminated SL and may aid in the understanding and studying of surface and interface physics.

5. ACKNOWLEDGMENTS

This work was supported in part by the Joint Services Electronics Program and the Science and Technology Center Program of the National Science Foundation, NSF grant No. CHE- 8920120.

6. REFERENCES

1. I. Tamm, " ÜBER EINE MÖGLICHE ART DER ELECKTRONENBINDUNG AN KRISTALLOBERFLÄCHEN," Phys. Z. Sowjetunion vol. 1, pp. 733-746, 1932.

2. W. Shockley, " On the Surface States Associated with a Periodic Potential," Phys. Rev., vol. 56, pp. 317-323, Aug. 1939.

3. H. Ohno, E. E. Mendez, J. A. Brum, J. M. Hong, F. Agullo-rueda, L. L. Chang and L. Esaki, "Observation of 'Tamm States' in Superlattices," Phys. Rev. Lett., vol. 64, pp. 2555-2558, May 1990.

4. Maria Steslicka, Robert Kucharczyk, and M. L. Glasser, "Surface states in superlattices," Phys. Rev. B, vol. 42, pp. 1458-1461, July (1990).

5. G. Bastard, "Superlattice band structure in the envelope-function approximation," Phys. Rev. B, vol. 24, pp. 5693-5697, Nov. 1981.

6. P. Dawson, G. Duggan, H. I. Ralph and K. Woodbridge, "Energies of sub-band minima in GaAs-(AlGa)As quantum well heterostructures," GaAs and Related Compounds 1984, ed. by B de Cremoux, pp. 391-396, Adam Hilger Ltd, Boston, Bristol, 1985.

7. G. Bastard, " Theoretical investigations of superlattice band structure in the envelope-function approximation," Phys. Rev. B, vol. 25, pp. 7584-7597, June 1982.

8. P.Yuh and K. L. Wang, " Formalism of the Kronig-Penney model for superlattices of variable basis," Phys. Rev. B, vol. 38, pp. 13307-13315, Dec. 1988.

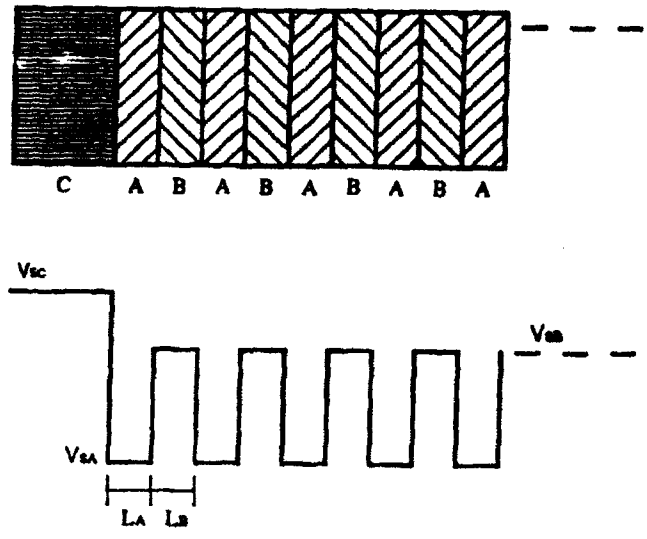


Fig. 1: Semi-infinite A-B superlattice terminated by C. V_{si} is the conduction band minimum in region i (i=A,B,C). L_a and L_b are the thicknesses of region A and B respectively.

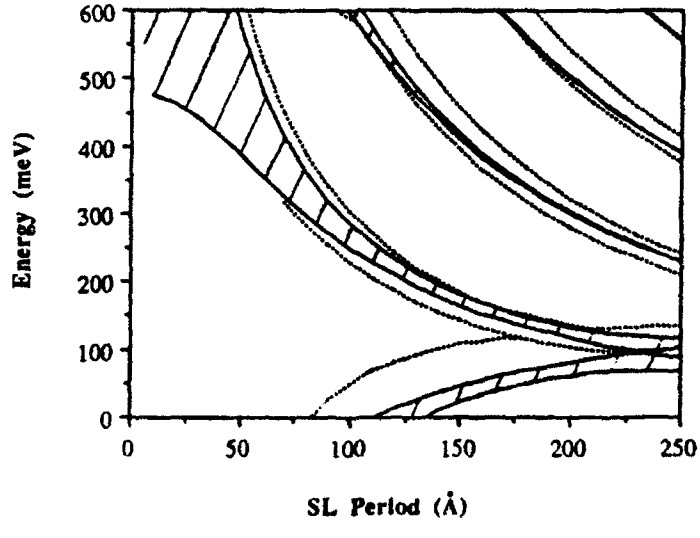


Fig. 2: The allowed energy levels of InAs-GaSb superlattice terminated in vacuum, as a function of the superlattice period with equal well and barrier width. The hatched regions are the minibands and the dotted lines correspond to the superlattice surface states. The origin of energy is taken at the conduction band minimum of InAs.

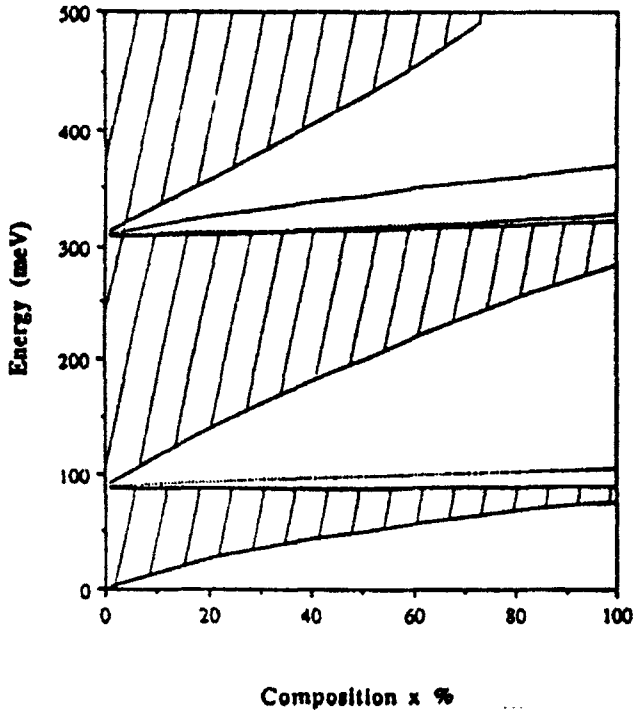


Fig. 3: The energy levels of a terminated GaAs-$Al_xGa_{1-x}As$ superlattice with well width=60 Å and barrier width=15 Å. The hatched regions are the minibands and the dotted lines correspond to the SL surface states. The origin of energy is taken at the conduction band minimum of GaAs.

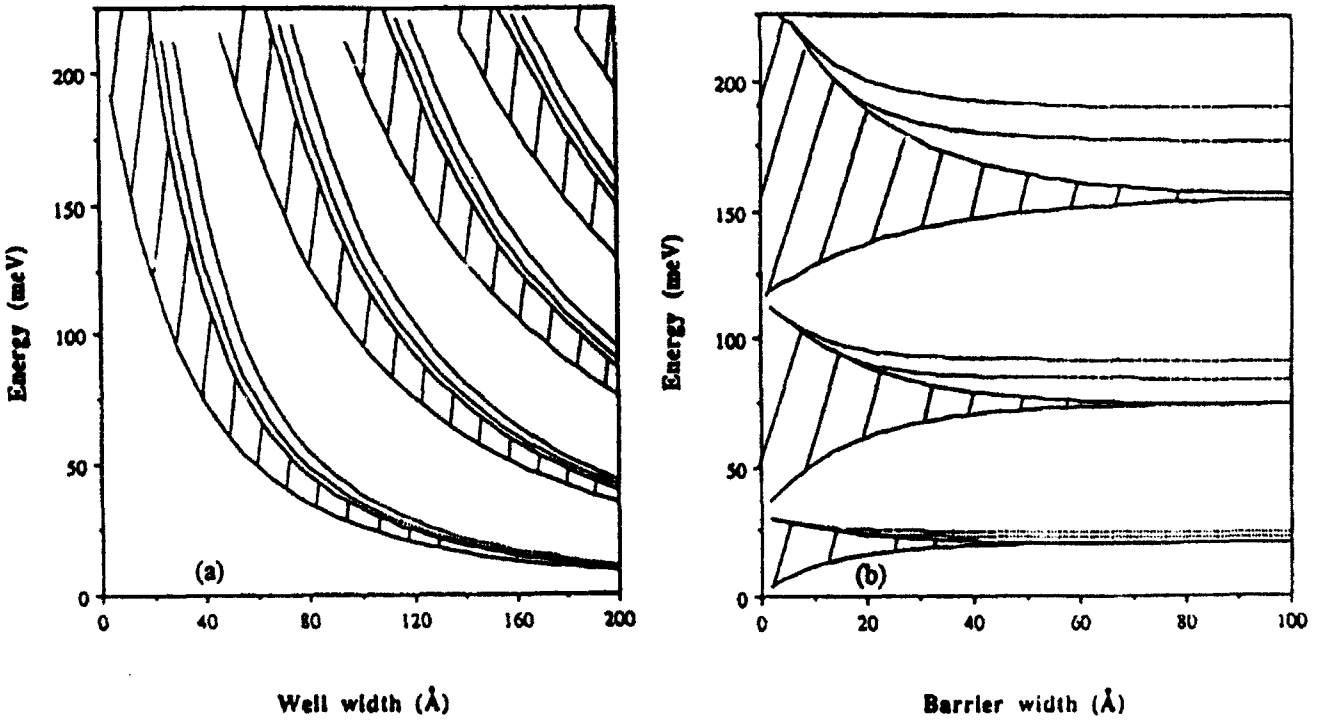


Fig.4: The allowed energy levels of a terminated GaAs-$Al_xGa_{1-x}As$ SL (a) as a function of well width with the barrier width=31 Å and (b) as a function of barrier width with well width=130 Å. The hatched regions are the minibands and the dotted lines correspond to the SL surface states. The origin of energy is taken at the conduction band minimum of GaAs.

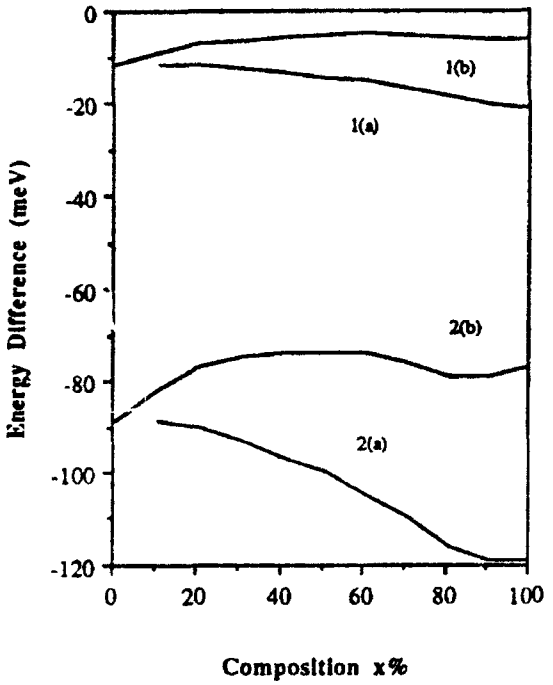
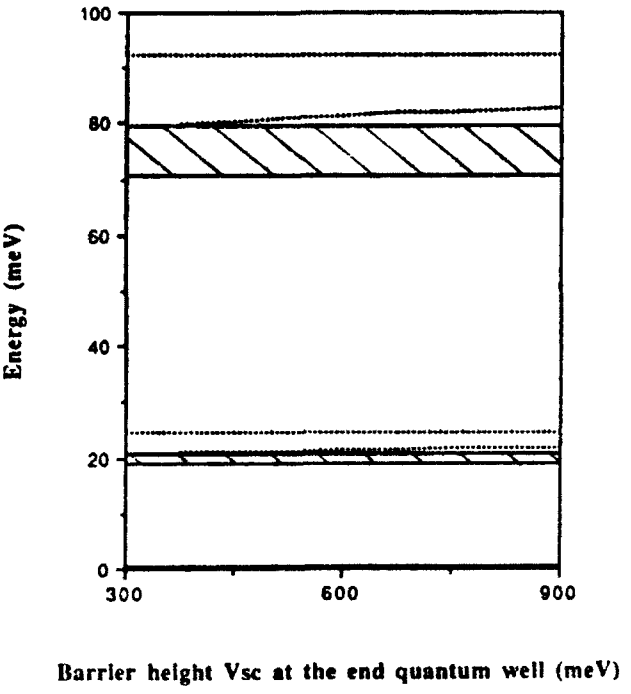


Fig. 5: The difference in energy levels of the SL surface states calculated using the nonparabolic and the parabolic band model for a GaAs-$Al_xGa_{1-x}As$ SL terminated in vacuum as a function of the composition of $Al_xGa_{1-x}As$. 1(a) is for the lower lying state and 1(b) is for the higher lying state in the first minigap. The corresponding states are also shown for the second minigap.

Fig. 6: The allowed energy levels of a terminated GaAs-$Al_xGa_{1-x}As$ SL as a function of the interface barrier height at the end quantum well. The well width was taken as 130 Å and the barrier width as 40 Å. The hatched regions are the minibands and the dotted lines correspond to the SL surface states. The origin of energy is taken at the conduction band minimum of GaAs.



REAL-TIME FEMTOSECOND ELLIPSOMETRY OF Si_xGe_{1-x} EPILAYERS

H. R. CHOO, M. C. DOWNER
Department of Physics, University of Texas at Austin, Austin, TX 78712

V. P. KESAN
IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598

ABSTRACT

We have developed a femtosecond ellipsometer by incorporating ellipsometric probe optics into a rapid scan femtosecond pump-and-probe experiment. The system allows near *real-time* display of the photo-induced reflectivity changes and provides complete characterization of the time-varying dielectric function. This ellipsometer is used *ex situ* to characterize the femtosecond response of relaxed, MBE-grown Si_xGe_{1-x} alloys over the complete composition range. The results show that the femtosecond response depends strongly on alloy composition in optically thick samples. Ge-like samples ($x<0.37$) show a characteristic two-component response which may be caused by intervalley $L \rightarrow \Gamma$ hole scattering and impact ionization. For a given alloy composition, the presence of interfacial strain or surface oxidation strongly alter the femtosecond response.

INTRODUCTION

Noncontact optical probing methods, such as cw ellipsometry, are being used increasingly for *in situ* characterization and control of epitaxial semiconductor growth in non-ultrahigh vacuum (UHV) environments, where electron diffraction cannot be used[1]. An *in situ* monitor of the femtosecond optical response of semiconductors could enrich the diagnostic power of optical probes, since the ultrafast relaxation of photogenerated carriers can depend strongly on process-induced defects, interface roughness, surface Fermi level pinning, and other surface conditions which arise during growth[2]. Nevertheless long data acquisition times have limited its development as an *in situ* diagnostic. Moreover, for some important electronic materials such as the Si_xGe_{1-x} alloys, very few quantitative experiments have measured the femtosecond response over the complete composition range, even for *ex situ* samples.

In this paper we present *ex situ* measurements of the femtosecond response of MBE-grown, optically thick, relaxed epilayers over the complete alloy composition range, using an unamplified colliding pulse mode-locked (CPM) dye laser. In addition we attempt to understand qualitatively how these responses arise from underlying bulk carrier dynamics. Finally we show how interfacial strain within the optical probe depth or surface oxidation radically alter the character of the femtosecond response. In order to achieve complete, sensitive and rapid optical characterization in these measurements, we have developed a prototype femtosecond ellipsometer that monitors probe pulses reflected obliquely from the

Mat. Res. Soc. Symp. Proc. Vol. 263. ©1992 Materials Research Society

samples through ellipsometric optics following excitation of electron-hole pairs by normally incident pump pulses. Our system achieves nearly real-time display of the femtosecond response by using a novel "rapid scan" technique[3] in which pump-probe delay is rapidly modulated.

EXPERIMENT

A cavity dumped output (~1MHz) from an unamplified CPM laser of 620nm, 80 fs pulses, is divided into strong pump (~1nJ) and weak probe beams. The probe beam is further divided into a reference beam and a sample probe for subsequent differential detection. The probe beam goes through a sequence of polarizer, compensator (quarter waveplate), sample, and analyzer. The time delay between pump and probe is introduced by a retroreflector attached to a rapid shaker. The pump-induced reflectivity change is continuously digitized and averaged by a fast AD converter interfaced to an IBM PC. No lock-in detection is used. Real time display of reflectivity change is possible for Ge rich alloys, where the reflectivity changes are largest. For Si rich alloys, about a minute is usually required to acquire the reflectivity changes, depending on the pump pulse energy. With more powerful Ti-sapphire femtosecond lasers now widely available, we anticipate even greater ease in obtaining real time data acquisition. From the results for different settings of the compensator angle, time-varying dielectric constants are calculated. Further experiment and calculation detail on the femtosecond ellipsometric procedure will be discussed elsewhere.

RESULTS AND DISCUSSION

Fig.1 shows an example of the measured femtosecond response of a series of optically thick Si_xGe_{1-x} alloy samples ranging from pure Si to pure Ge. The Si_xGe_{1-x} epilayers were grown by MBE on Si[100] substrates to thicknesses of approximately 1 μm. Consequently, the Si substrate and the substrate-epilayer play a negligible role at our probe wavelength of 620nm. The measurements shown were performed in air immediately after the samples were grown, using pump energy ~1nJ, and a relatively small probe incidence angle ($\Theta_{probe} = 30°$), with probe polarization tuned ellipsometrically to pure linear s-polarization. For pure Si (x=1.0) and the Si-rich ($x \geq 0.5$) alloys, the response for these conditions consists of a pulsewidth-limited reflectivity decrease upon photoexcitation ($\Delta t = 0$) of an electron-hole plasma, followed by a slow monotonic recovery governed by ambipolar diffusion, electron-phonon interaction, and bulk and surface recombination[2]. The magnitude of the initial reflectivity drop increases from $\Delta R/R \sim 10^{-5}$ at x=1 to $\Delta R/R \sim 10^{-4}$ at x=0.5, consistent with the increasing pump absorption coefficient. However, the temporal recovery pattern changes very little over this composition range. The pump injects carriers entirely via the indirect L $\rightarrow$ X valley transition in this composition range. Consequently, electrons relax to the X valley conduction band (CB) minimum, with negligible intervalley transfer of either electrons or holes.

As Ge content increases further (x<0.5), major changes in the femtosecond response become evident. First, the initial signal magnitude begins to increase with Ge content (1-x) at a much faster rate than at lower Ge concentrations. This effect correlates with the transition

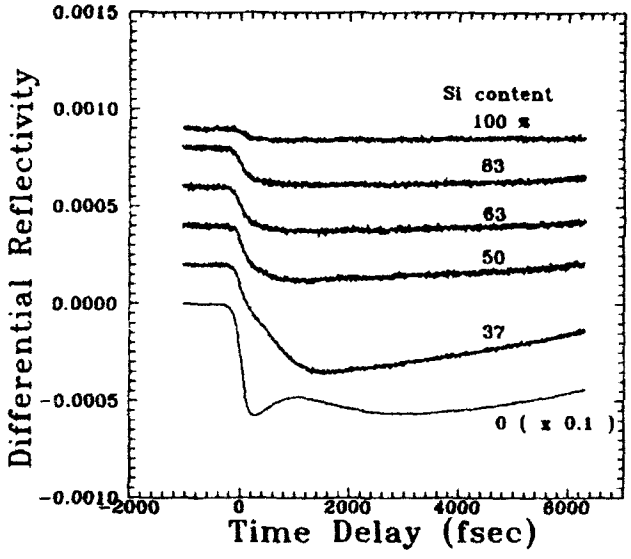


Fig.1 Femtosecond reflectivity response of the Si_xGe_{1-x} alloys. The numbers represent the Si content in each alloy. The reflectivities at timedelay zero are displaced for clarity.

from Si-like band structure (X-valley CB minimum) to Ge-like band structure (L valley CB minimum), which brings about a rapid increase in the pump absorption coefficient α at 2.0 eV. Secondly, at the same crossover composition, a two-component picosecond reflectivity response appears, evident as a delayed fall time in the x=0.37 sample, and as a prominent secondary reflectivity minimum for pure Ge. The correlation of this effect with the transition to Ge-like band structure suggests that the two must be related.

We believe that the two-component response occurs because, as Ge content approaches unity, direct transitions across the L valley increasingly dominate the pump absorption, resulting in a large concentration of electrons and holes around the L point. For electrons, this is the CB minimum. Consequently, little further electron dynamics are expected in the first few picoseconds. L-point holes, on the other hand, carry excess > 1 eV, which exceeds the band gap. They can relax via two processes: 1) impact ionization, in which the L-point hole scatters to Γ via creation of an electron-hole pair, and 2) hole-phonon scattering, causing transfer of the hole from L to Γ without creating an additional electron-hole pair. Both processes may contribute to the delayed reflectivity decrease observed in Ge-like samples. Impact ionization is a source of delayed, secondary carrier generation. Furthermore, the L $\rightarrow$ Γ hole transfer decreases the hole effective mass m_h^* which enhances the Drude term $4\pi Ne^2/m_h^*$ in the evolving dielectric function, thus also contributing to a delayed reflectivity decrease.

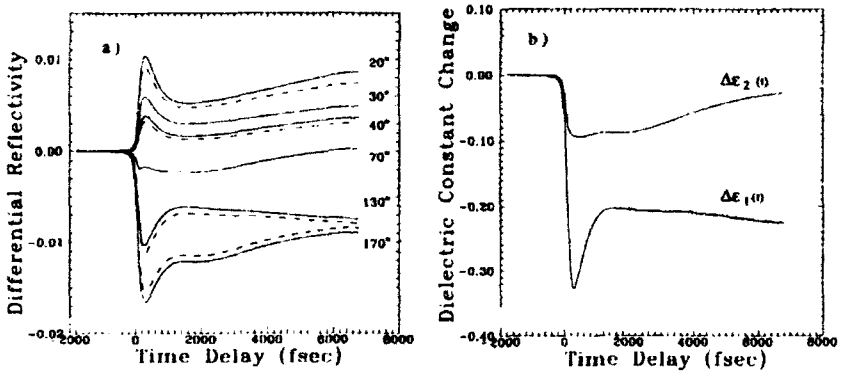


Fig.2 Examples of oblique incident ellipsometric probe measurement using pure Ge as a sample.
a)Pump induced reflectivity changes for different compensator angles. The angles next to each curve represent the compensator angle settings. Measured results are shown as solid curves. Calculated responses using the time-resolved dielectric constants in b) are shown as dotted curves. Probe incident angle was 67°. Polarizer and analyzer were set at 0° and 45°, respectively.
b)Calculated dielectric constant change using the results of compensator angle at 30° and 70° in a).

Fig.2a) shows an example of complete characterization of the femtosecond response of pure Ge using approximately the same pump energy as in Fig.1, but a more oblique probe incident angle (Θ_{probe} = 67°) and a variety of ellipsometric settings. For the measured responses (solid line) shown the polarizer and analyzer were set at 0° and 45°, respectively, and the compensator angle was varied as indicated in the figure whose definition can be found in reference [6]. The pumping condition is the same for all of the responses. For a uniform, optically thick sample, any two of these responses can be used to determine uniquely both the real $\Delta\varepsilon_1(t)$ and imaginary $\Delta\varepsilon_2(t)$ parts of the evolving dielectric constant at the probe wavelength. Fig.2b) shows $\Delta\varepsilon_1(t)$ and $\Delta\varepsilon_2(t)$ extracted from the measured responses at $\Theta_{compensator}$ = 30° and 70°. Clearly the real and imaginary components undergo changes of comparable magnitude, showing that the Drude dielectric term from the electron-hole plasma and the interband absorption term both contribute strongly to the observed response, as expected for above gap probe photons. The extracted $\Delta\varepsilon_1(t)$ and $\Delta\varepsilon_2(t)$ can be used to calculate the response for other ellipsometer settings as a check of internal consistency. The dashed curves in Fig.2a) represent such calculations, and demonstrate a high degree of internal consistency in the data.

Fig.3 shows several examples of the sensitivity of the femtosecond response of Si_xGe_{1-x} to sample conditions. Curve a) in Fig.3 shows the measured response of a strained, optically thin (80nm thick) sample of $Si_{0.75}Ge_{0.25}$ grown by MBE on Si[100] under the same conditions and in the same chamber, as the thicker samples described above.

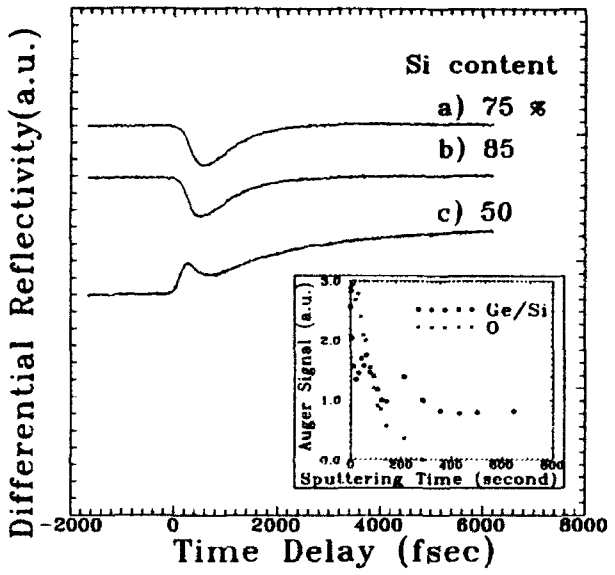


Fig.3 Reflectivity responses of strained and oxidized samples. a)MBE-grown 80nm thick $Si_{0.75}Ge_{0.25}$. b)RPCVD-grown 80nm thick $Si_{0.85}Ge_{0.15}$. c) $Si_{0.5}Ge_{0.5}$ after exposed to ambient air for several months. The inset shows the Auger depth profile analysis result. The black dots represent the ratio of Ge to Si Auger signal.

The recovery time is dramatically shortened compared to the relaxed sample. Undoubtedly this effect is related to the high defect density present in the strained region around the $Si_{0.75}Ge_{0.25}$/Si interface, which lies well within the optical pumping and probing depths. The faster recovery can be attributed to trapping of carriers in defect states, leading to accelerated recombination. In order to check the reproducibility of these effects, a second $Si_{0.85}Ge_{0.15}$ sample of the same thickness and similar composition was grown on Si[100] by a different method - remote plasma chemical vapor deposition (RPCVD). Curve b) in Fig.3 shows that its response is virtually indistinguishable from that of the MBE-grown sample. This high degree of reproducibility suggests that the faster recovery is caused by inherent features of the strained interface, independent of details of the growth procedure or conditions.

Curve c) in Fig.3 shows the measured femtosecond response of the $Si_{0.5}Ge_{0.5}$ sample used to acquire the Fig.1 data, after the sample had been exposed to ambient air at room temperature for several months. Clearly the character of the femtosecond response has changed completely. Not only has the sign and magnitude of the initial reflectivity response changed, but a two-component response is now clearly evident, suggesting Ge enrichment near the surface, possibly caused by Ge diffusion driven by the surface oxidation process[4]. As a test of this hypothesis, we performed an Auger depth profile analysis[5] of Ge, Si, and Oxygen concentration in the oxidized sample, shown in the inset. Enhancements in Ge and

oxygen concentration near the surface are clearly evident. Nevertheless, we cannot at present rule out the possibility that the high surface Ge concentration was present in the as-grown sample.

These examples illustrate that subtle changes in sample properties sometimes significantly alter the femtosecond response of Si_xGe_{1-x} samples, thus demonstrating the possible diagnostic power of a femtosecond ellipsometer.

ACKNOWLEDGMENTS

This research was supported in part by the Science and Technology Center Program of the National Science Foundation (NSF grant CHE-890210), and in part by the Air Force Office of Scientific Research (Contract F49620-89-C-0044) and the Robert A. Welch Foundation (Grant F-1038). We are grateful to S. Banerjee and A. Tasch for helpful discussions, and for providing some of the Si_xGe_{1-x} samples used in this work.

REFERENCES

1. D. E. Aspnes, W. E. Quinn, and S. Gregory, Appl. Phys. Lett. 57, 2707 (1990).
2. G. C. Cho, W. Kutt, and H. Kurz, Phys. Rev. Lett. 65, 764 (1990); A. Esser, W. Kutt, M. Strahnen, G. Maidon and H. Kurz, Appl. Surf. Sci. 46, 446 (1990).
3. M. Strahnen, W. Kutt and H. Kurz, in VME-bus in Research: Proceedings of an International Conference, Zurich, Switzerland, edited by C. Eck and C. Parkman (North-Holland, Amsterdam, 1988) , p.69.
4. J. Eugene, F. K. LeGoues, V. P. Kesan, S. S. Iyer, and F. M. d'Heurle, Appl. Phys. Lett. 59, 78 (1991).
5. H. K. Liou, P. Mei, U. Gennser, and E. S. Yang, Appl. Phys. Lett. 59, 1200 (1991).
6. R. Azzam and N. Bashra, in Ellipsometry and Polarized Light, New York, North Holland, 1977, Chapter 3.

Measurement of Femtosecond Ionization Dynamics of Atmospheric Density Gases by Spectral Blueshifting

Wm. M. Wood, C. W. Siders, and M. C. Downer
Physics Department, University of Texas at Austin, Austin, Texas 78712
(Received 5 August 1991)

The new plasma diagnostic technique of spectral blueshifting of femtosecond pulses is used for the first time to analyze quantitatively the ionization of noble gases under the influence of intense, femtosecond illumination. The two processes of strong-field tunneling ionization and electron impact ionization are found to play competing roles on these time scales.

PACS numbers: 52.40.Nk, 32.80.−t, 34.80.Dp, 52.25.Jm

Highly ionized plasmas approaching atmospheric density are a potential future source of coherent x rays [1], and a potential medium for charged-particle acceleration [2]. *Ionization by intense, femtosecond pulses* holds promise for precise control of initial plasma conditions (temperature, ionization state, density) [3] which are critical to these applications. At the same time, new, quantitative, experimental diagnostics compatible with the high gas density and ultrafast time scale are needed to measure the ionization and subsequent plasma dynamics which give rise to these conditions. A number of authors have shown, both experimentally and theoretically, that a laser pulse which rapidly ionizes a gas experiences a frequency blueshift caused by the creation of a free-electron plasma [4,5]:

$$\Delta\omega = -\frac{\omega_0}{c}\int_0^{\cdot} \frac{\partial n}{\partial t}(l)\,dl\,. \tag{1}$$

Here, ω_0 is the angular frequency of the light, l is the longitudinal distance over which the interaction occurs; $n=(1-\omega_p^2/\omega_0^2)^{1/2}$ is the index of refraction of the medium through which the pulse travels, and is found using the Drude model (ω_p is the plasma frequency.) In experiments which used pulse durations long with respect to ion-ion collision times [4], plasma expansion and recombination following the ionization contributed to and complicated the phase modulation induced on the pulse. In a previous publication [6], we showed for the first time that *femtosecond* pulses, tightly focused to intensities above the ionization threshold, experience a "pure" blueshift, i.e., with no trace of components redshifted from the original pulse spectrum, indicating qualitatively that the phase modulation on the laser pulse is caused entirely by ionization. Furthermore, complicating nonlinear optical interactions with neutral gas, which can induce spectral *broadening* (supercontinuum) and self-focusing on femtosecond pulses focused more loosely (below the ionization threshold) in dense ($p>40$ atm) gases [7], were suppressed at near atmospheric pressures because of the lower density and rapid plasma growth early in the tightly focused pulse [6]. This preliminary result suggested that femtosecond time-resolved measurement and analysis of the spectral shifts could measure detailed growth dynamics of the ionization front over a broad range of gas pressures and gas species, distinct from the subsequent plasma dynamics and optical nonlinearities of the neutral gas. In a wider context, such frequency upshifts might also quantitatively diagnose plasma density oscillations in plasma-based particle accelerator schemes [8].

The purpose of this Letter is to demonstrate, for the first time, the quantitative use of this plasma diagnostic technique in analyzing the breakdown of noble gases caused by intense, femtosecond illumination. This will be done in three stages: (1) We report systematic observations of the self-blueshifting of laser pulses after ionizing 1–5-atm pressure samples of He, Ne, Ar, Kr, and Xe, which reveal a universal, reproducible pattern in the shape of the blueshifted spectra. Specifically, with increasing laser intensity, gas pressure, and atomic number, the self-blueshifted spectra develop from a near replica of the incident pulse spectrum into a complex structure consisting of two spectral peaks: a narrow peak shifted between 5 and 10 nm towards the blue from the original spectrum, and a broad peak shifted further towards the blue whose position and width depend strongly on the gas pressure, gas species, and the laser-pulse energy. (2) We report time-resolved spectral shifts of a weak probing pulse which show different temporal evolution for each of these two spectral features. (3) Finally, we propose a quantitative, *ab initio* model which relates these two spectral features to two competing ionization mechanisms: collisionless tunneling ionization, predicted to dominate early in the ionizing pulse profile, and electron-impact ionization, predicted to dominate in the maximum of the pulse profile.

In our experiments, 100-fs laser pulses with center wavelength of 620 nm and energies up to 0.4 mJ are focused at $f/5$ to a peak intensity of $10^{16\pm0.3}$ W/cm^2 into a glass cell containing 1–5 atm of He, Ne, Ar, Kr, or Xe. All of the light transmitted through the focal region is collected and analyzed by a spectrometer. Typically, less than 1% of the pulse energy is lost in ionizing the gas. Autocorrelation measurements [6(b)] after the interaction region show only −10% temporal broadening. Figure 1(a) shows the measured spectrum of the ionizing pulse after breakdown in 5 atm Kr. As pulse energy increases, the spectrum blueshifts with little change in shape until the pulse energy reaches 0.1 mJ [$\log_{10}$(intensity) = 15.0], after which the position of the peak changes very little. As the energy increases further, a shoulder

© 1991 The American Physical Society

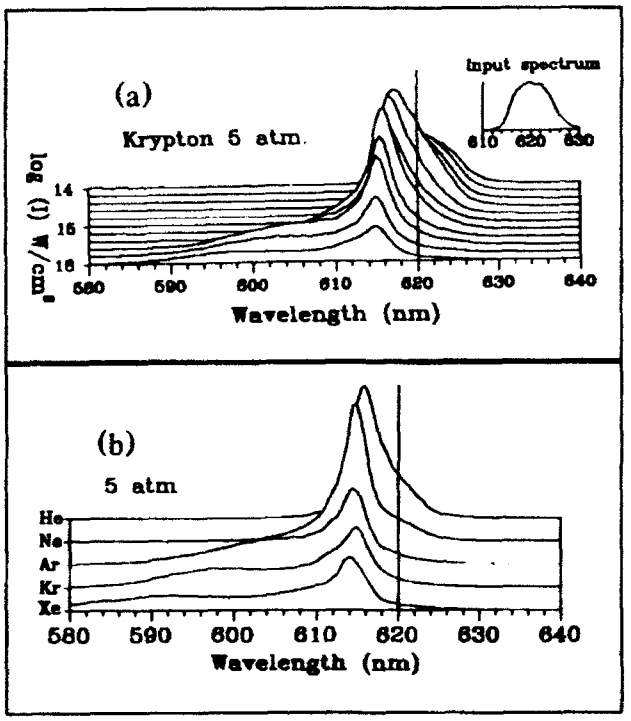


FIG. 1. (a) Spectra after interaction with 5-atm pressure Kr as a function of pulse energy. Pulse energy is increasing towards the bottom in steps of $\times 10^{0.2}$. (b) Spectra after interaction of 0.25-mJ pulses in 5-atm pressure of each of the noble gases studied.

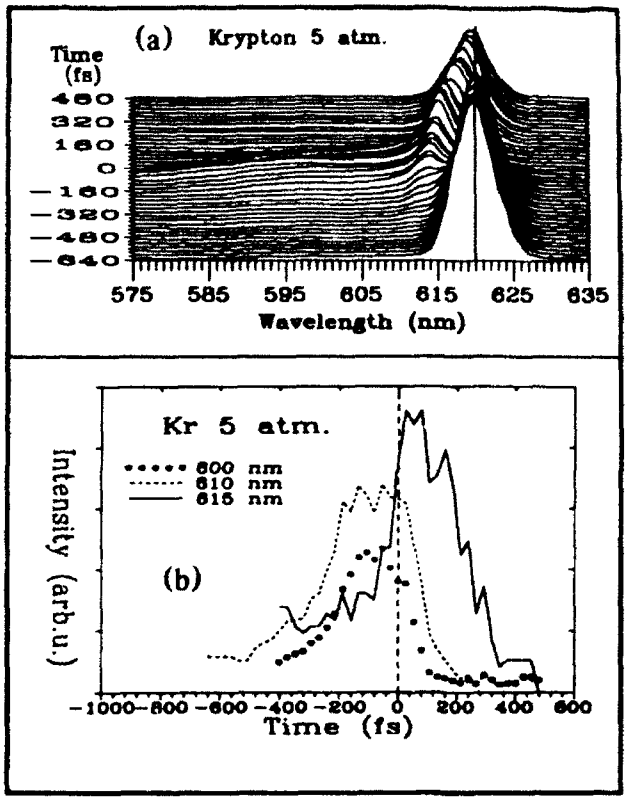


FIG. 2. (a) Time-resolved spectra in krypton at 5 atm using pump pulse energy of 0.22 mJ. There are 27 fs between spectra; time increases towards the top of the figure, with negative times corresponding to the probe pulse arriving before the pump pulse. Coincidence of the pump and probe pulses is indicated by an emboldened spectrum. (b) Energies at 600, 610, and 615 nm as functions of time for time-resolved spectra in 5 atm Kr.

appears on the blue side of the spectrum, then broadens and shifts further towards the blue. Figure 1(b) shows self-shifted spectra after ionization of 5 atm of each of the noble gases using 0.25-mJ pulses. The center of the unshifted pulse spectrum is indicated by a vertical line, showing that a blueshift occurs in all cases. However, the shape of the blueshifted spectrum depends strongly on gas species. In Ar, Kr, and Xe, the narrow, less-shifted peak and the broad, blue shoulder are clearly discernible, the latter becoming a separate peak in Kr and Xe. In He and Ne, the blue shoulder is absent, and a different feature —a "red shoulder" corresponding approximately to the unshifted spectrum—appears showing that some of the initial pulse energy remains unshifted. Although the amplitude and width of each of these spectral features depend on the focal profile, chirp, and other details of the ionizing pulse, numerous measurements have confirmed that the qualitative trends shown in Fig. 1 and described above are universal, reproducible features of femtosecond ionization of the noble gases. Modest defocusing of the transmitted pulse and a visible breakdown "spark" (i.e., recombination luminescence) always accompany, and are precisely correlated with, the onset of the blueshift, consistent with their common origin in the rapid formation of a reduced index plasma [6].

Time-resolved pump-probe experiments show that each of these spectral features also has a characteristic temporal evolution within the pump pulse. To obtain time-resolved blueshifted spectra a weak probe pulse, derived from the pump with a beamsplitter, was polarized orthogonally to the ionizing pulse, then copropagated through the focal region before ($\Delta t < 0$), coincident with, or after ($\Delta t > 0$) the pump pulse. The probe was separated from the pump after the interaction region by a polarization analyzer, and its spectrum recorded for different Δt. Figure 2(a) shows a series of time-resolved probe spectra resulting from ionization of 5 atm Kr by a pump pulse centered at $\Delta t = 0$. Coincidence was determined to within ± 20 fs by slightly turning the pump polarization, and then adjusting an optical delay line to maximize the contrast of the resulting pump-probe interference fringes. The probe spectra around $\Delta t = 0$ exhibit the same dual structure seen in the self-shifted spectrum [Fig. 1(a)]. However, close examination shows that the less-shifted peak and the broad bluer shoulder evolve differently in time. To show this contrast quantitatively, we have plotted in Fig. 2(b) the area under the spectra in small regions around 615, 610, and 600 nm. The values at 610 and 600 nm display the same temporal behavior, with maxima at approximately 100 fs before the maximum of the ionizing pulse, indicating that the process giving rise

to the blue shoulder occurs in the early part of the pulse. The values plotted for 615 nm show a maximum approximately 50 fs after the maximum of the pump pulse, indicating that the process causing the narrow, less-shifted peak occurs near or slightly after the maximum of the ionizing pulse. A similar time development is observed for the two corresponding blueshifted features for Kr and Ar. The single blueshifted peak for He and Ne, on the other hand, evolves early in the pump profile, with little or no probe blueshift observed for $\Delta t > 0$. Again, quantitative details of the temporal evolution depend on focal profile, but the qualitative trends are observed reproducibly [9].

We now propose a simple, *ab initio* ionization model which accounts for (1) the two blueshifted features observed in Ar, Kr, and Xe, (2) their respective temporal behavior, and (3) the "unshifted" portions of the spectra and the *absence* of the blue shoulder observed in He and Ne. In the presence of a strong light field, the ionization of the atoms can be modeled using the strong-field tunneling theory due to Keldysh or Ammosov, Delone, and Krainov [10], which yield similar results for the lower stages of ionization. The coupled equations governing the densities N_i of atoms of a particular ionization state i are written in the following way:

$$\frac{dN_i}{dt} = (P_{i-1}N_{i-1} - P_iN_i) + (N_e\sigma_{i-1}v_eN_{i-1} - N_e\sigma_iv_eN_i)\,. \tag{2}$$

The first parenthetical term on the right describes the rate of growth of the ith ionization stage, where P_k is the probability per unit time [10] for electrons to tunnel from k-times ionized parent ions. The second term describes collisional ionization via electron impact, where N_e is the free-electron density, v_e is the rms electron velocity, and σ is the cross section [11] for the process. Intermediate resonant states can be ignored under our conditions [12]. The density $N_e(t)$ of free electrons and the resulting index of refraction $n(t)$ calculated from Eq. (2) yields a calculated blueshift upon substitution into Eq. (1). The model using only the tunneling ionization terms P_k adequately describes the observed blueshifting in He and Ne. Because tunneling ionization occurs almost entirely during the first half of the laser pulse, only the leading edge of the pulse becomes blueshifted; the trailing part of the pulse remains unshifted, and appears as the spectral "red" shoulder observed in the He and Ne gas breakdown.

At Ar, Kr, and Xe pressures > 1 atm and with maximum pump intensity, the red unshifted shoulder disappears, and the *entire* pulse spectrum is shifted [Fig. 1(b)]. This observation implies that *ionization occurs during* and after the peak of the ionizing pulse, and can be explained by including electron-impact ionization [the second parenthetical term of Eq. (2)]. Collisional ionization is strongest near the peak of the laser pulse because (1) the cross sections are maximal for electron energies in the range 100–500 eV [11], very close to the electron quiver energy at the peak (370 eV for 10^{16} W/cm^2); (2) N_e is large due to the strong-field ionization early in the pulse; and (3) the rms velocity v of free electrons is maximal at the peak of the pulse. *Stated from a different* viewpoint, for the first few charge states, the thresholds for semiclassical barrier suppression ionization (BSI) [13] are less than those for collisional ionization (defined as the intensity at which the quiver energy equals the low-field ionization energy), while for subsequent ionization stages the collisional threshold is reached before the barrier suppression threshold [13]. Estimates of typical ionizing collision times in both He and Ne, where observations imply little or no collisional effect, yield τ

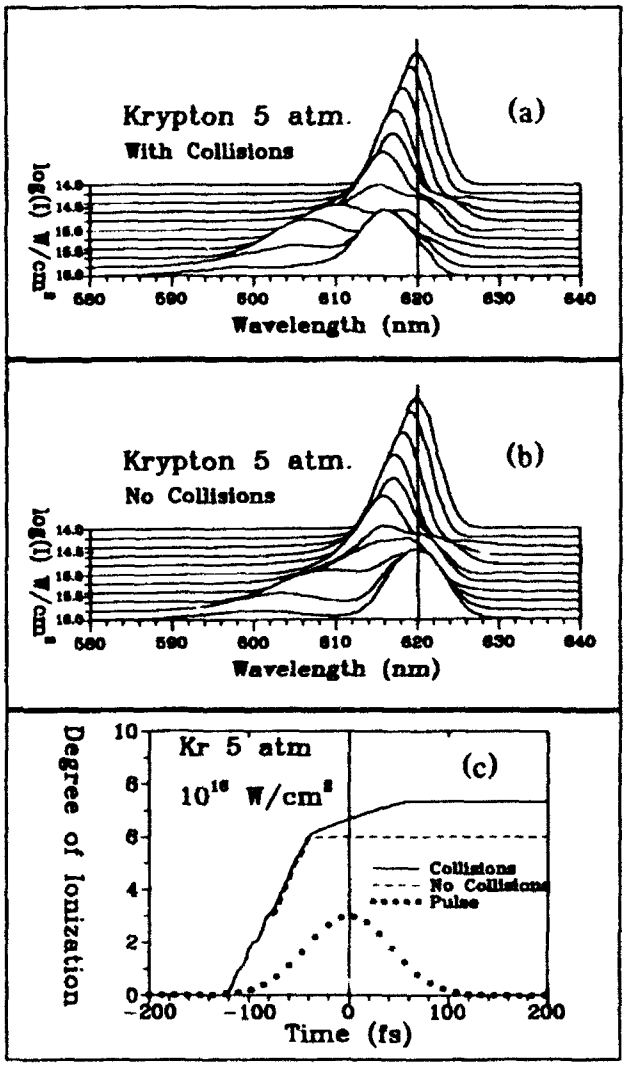


FIG. 3. (a) Calculated self-shifted spectra for ionization of 5 atm Kr by 100-fs pulses of varying peak intensities, using Ammosov strong-field and collisional ionization terms. (b) Calculated spectra for ionization of 5 atm Kr without collisions in the model. (c) Degree of ionization vs time calculated both with and without collisions in 5-atm Kr gas peak intensity of 10^{16} W/cm^2.

$=(N_e\sigma v)^{-1}\sim 200$ fs, significantly *longer* than the light pulse duration. In Ar, Kr, and Xe at 5-atm pressure, however, these collision times are significantly *shorter* than the pulse duration ($\tau < 30$ fs), resulting in significant ionization near the peak of the pulse, and a shifting of the *entire* pulse spectrum.

Figure 3(a) shows the calculated spectra after breakdown of 5 atm Kr, for peak intensities ranging from 10^{14} to 10^{16} W/cm^2, calculated using Eq. (2) with collisions and a simplified cylindrical interaction region with constant transverse intensity [14]. Figure 3(b) shows the spectra calculated *without* collision ionization: Significant unshifted energy remains in these spectra as peak intensity increases [Fig. 3(b)]. *With* collisions in the model, the *entire* pulse spectrum is shifted, following the same trend as observed in the data. The broad, blue shoulder observed in the data for Ar, Kr, and Xe corresponds to strong-field ionization (and thus has the same origin as the blueshifts observed in He and Ne), while the narrow, less-shifted peak corresponds to slower collisional impact ionization. The calculated temporal behavior of the two features further corroborates this interpretation. Figure 3(c) shows the calculated degree of ionization as a function of time in 5 atm Kr illuminated by a 100-fs pulse with a peak intensity of 10^{16} W/cm^2, both with and without collisions. Collisional ionization, like the smaller blueshift which it causes, occurs during and after the peak of the pulse, and is slower than strong-field ionization, which causes the larger blueshift early in the pulse. This temporal behavior is also observed in Xe at lower pressures, and in 5 atm Ar. Because of the tight focus and the early onset of ionization in the pulse profile, the calculated blueshift is negligibly affected by including self-phase modulation caused by n_2 of the neutral gas, ions, and electrons, as shown in detail elsewhere [6(b), 15].

In conclusion, the model predicts a high rate of ionization early in the laser pulse due to collisionless strong-field ionization, followed by collisional ionization at a lower rate near the peak of the laser pulse, in good agreement with the data. A more detailed report of the experiments and model will be published separately [16].

This research was supported by the National Science Foundation (Grant No. DMR8858388), the Robert A. Welch Foundation (Grant No. F-1038), and the Air Force Office of Scientific Research (Contract No. F49620-89-C-0044).

[1] W. W. Jones and A. W. Ali, J. Appl. Phys. **48**, 3118 (1977); J. Peyraud and N. Peyraud, J. Appl. Phys. **43**, 2993 (1972).

[2] T. Tajima and J. M. Dawson, Phys. Rev. Lett. **43**, 297 (1979); P. Sprangle *et al.*, Appl. Phys. Lett. **53**, 2146 (1988).

[3] N. H. Burnett and P. B. Corkum, J. Opt. Soc. Am. B **6**, 1195 (1989); B. Penetrante and J. N. Bardsley, Phys. Rev. A **43**, 3100 (1991).

[4] E. Yablonovitch, Phys. Rev. A **10**, 1888 (1974).

[5] S. C. Wilks, J. M. Dawson, and W. B. Mori, Phys. Rev. Lett. **61**, 337 (1988).

[6] (a) Wm. M. Wood, G. B. Focht, and M. C. Downer, Opt. Lett. **13**, 984 (1988); (b) M. C. Downer *et al.*, in *Proceedings of the Tenth International Conference on Spectral Line Shapes*, edited by L. Frommhold and J. W. Keto, AIP Conf. Proc. No. 216 (AIP, New York, 1990).

[7] P. B. Corkum, C. Rolland, and T. Srinivasan-Rao, Phys. Rev. Lett. **57**, 2268 (1986); J. H. Glownia *et al.*, J. Opt. Soc. Am. B **3**, 1573 (1986); H. J. Lehmeier, W. Leupacher, and A. Penzkofer, Opt. Commun. **56**, 67 (1985).

[8] E. Esarey, A. Ting, and P. Sprangle, Phys. Rev. A **42**, 3526 (1990); S. C. Wilks *et al.*, Phys. Rev. Lett. **62**, 2600 (1989).

[9] Note that the nonlinear index $n_2 > 0$ of the neutral gases or ions, if significantly present, would induce a rising index (redshift) in the leading edge of the pulse and a falling index (blueshift) in the trailing edge, contrary to the observed blueshift in the leading edge, which must therefore be caused solely by ionization.

[10] L. V. Keldysh, Zh. Eksp. Teor. Fiz. **48**, 874 (1965) [Sov. Phys. JETP **21**, 1307 (1965)]; M. V. Ammosov, N. B. Delone, and V. P. Krainov, Zh. Eksp. Teor. Fiz. **91**, 2008 (1986) [Sov. Phys. JETP **64**, 1191 (1986)].

[11] H. Tawara and T. Kato, At. Data Nucl. Data Tables **36**, 167–353 (1987). Impact-ionization cross sections for intensely irradiated atoms and ions are probably larger than the values for ground-state atoms and ions given in these tables and used in our calculations.

[12] P. B. Corkum, N. H. Burnett, and F. Brunel, Phys. Rev. Lett. **62**, 1259 (1989).

[13] S. Augst, D. D. Meyerhofer, D. Strickland, and S. L. Chin, J. Opt. Soc. Am. B **8**, 858 (1991). For example, using the BSI model, we obtain the following threshold intensities for the first seven ionization stages in Kr: $\log_{10}(I_{BSI}) = 14.19$, 14.54, 14.92, 15.28, 15.45, 15.56, 16.09 W/cm^2, $\log_{10}(I_{coll}) = 14.59$, 14.83, 15.01, 15.16, 15.26, 15.32, 15.49 W/cm^2.

[14] Calculations which include convolution over realistic transverse intensity profiles have also been done and yield the same qualitative spectral features as shown in Figs. 3(a) and 3(b), typically with peaks of somewhat different widths and amplitudes. Since, however, the geometry of the interaction region is not quantitatively characterized because of defocusing, calculations using the simpler geometry are presented to illustrate the essential physics.

[15] The ionization-induced index change $\Delta n_{plasma} \cong -2\pi Ne^2/m \sim -0.02$ for 1-atm triply ionized plasma exceeds even the hypothetical nonlinear index $n_2^{neutral} I$ which would be induced by light of intensity $I = 10^{16}$ W/cm^2 interacting with neutral He, Ne, Kr, and Ar (for Xe, $n_2^{neutral} I \cong \Delta n_{plasma}$), using measured $n_2^{neutral}$ values (Lehmeier *et al.*, Ref. [7]). Taking into account the disappearance of the neutral gas early in the pulse, and the smaller n_2^{ion} values of the ions, $n_2^{ion} I$ is small compared to Δn_{plasma} at the peak of our pulse in all the noble gases. As for the free electrons, $n_2^{electron}$ becomes significant only for relativistic quiver velocities ($I \sim 10^{18}$ W/cm^2).

[16] Wm. M. Wood, C. W. Siders, and M. C. Downer (to be published); Wm. M. Wood, Ph.D. dissertation, University of Texas at Austin, 1991 (unpublished).

Optical properties of liquid carbon measured by femtosecond spectroscopy

D. H. Reitze,* H. Ahn, and M. C. Downer
Physics Department, The University of Texas at Austin, Austin, Texas, 78712
(Received 2 August 1991)

A comprehensive report of femtosecond time-resolved reflectivity and transmission of graphite and diamond following optical excitation above critical melting fluences F_m of 0.13 and 0.63 J/cm^2, respectively, is presented. Normal- and oblique-incidence reflectivity has been measured with 100-fs resolution at wavelengths ranging from 700 to 310 nm. Within 1 ps following excitation above F_m, probe reflectance increases sharply at visible frequencies, remains nearly unchanged at near-ultraviolet frquencies, and depends weakly on excitation fluence. These optical changes are interpreted as an ultrafast melting transition from crystalline graphite or diamond to a common, more reflective liquid state. During the first picosecond following excitation, electron and lattice temperatures substantially equilibrate, and the lattice melts, before heat conducts out of the absorbing volume or the surface hydrodynamically expands. A Drude model of the reflectance spectrum 1 ps after excitation reveals a strongly damped plasma (plasma frequency-relaxation time product $\omega_p\tau\sim 1$), in contrast to liquid silicon ($\omega_p\tau\sim 5$). Inferred electron mean free paths approach the average interatomic spacing (2 Å), implying electron localization. Optically determined dc resistivities up to $625\pm 75\ \mu\Omega$ cm agree with measurements at kilobar ambient pressure, but significantly exceed resistivities measured and calculated at low pressure. Thus, the attribution "metal" is questionable for fluid carbon under these conditions. The results demonstrate that femtosecond lasers can extend condensed-matter thermophysics measurements to temperature-pressure regimes inaccessible by other methods.

I. INTRODUCTION

Scientific debate over the properties of the elusive liquid state of carbon, nature's most refractory material, dates back to Ludwig's[1] 1902 report of increased resistance in carbon rods melted at high pressure. Extensive recent research[2-20] on high temperature-pressure phases of carbon, which has been reviewed by Bundy,[2] reflects not only the wide-ranging importance of the problem in condensed matter physics,[2] astrophysics,[16] and geology,[17-19] but the continuing controversy over the basic properties of the liquid state. For example, much recent debate has focused on whether liquid carbon is metallic or insulating, and on how its electrical conductivity depends on temperature and pressure.[4-14] Much of the controversy stems from the difficulty of creating and studying a reproducible sample of liquid carbon in the laboratory, particularly producing identical samples and conditions by different methods. Since the melting temperature (~ 5000 K) exceeds that of any containment vessel, and since the solid sublimes at low ambient pressure, long-term confinement at elevated temperature and pressure is impossible. Consequently experiments must rely on transient melting and probing, using methods such as shock waves,[15] pulsed ohmic heating[4] in high pressure cells,[14] and pulsed laser heating.[3,5,8-13]

Recent pulsed laser melting experiments[5,8,12,13] were the focus of a particularly sharp disagreement because of the apparent contradiction in measured properties. On the one hand, 30-ns laser pulses incident on highly oriented pyrolytic graphite (HOPG) created melt depths which Steinbeck *et al.*[5] could explain only by assuming large, metallic thermal conductivities in the melt. This result agreed qualitatively with metallic electrical resistivity ($\rho\sim 50\pm 20\ \mu\Omega$ cm) measured by microsecond pulsed ohmic heating of pyrolytic graphite fibers at low pressure.[4] On the other hand, in time-resolved optical reflectivity measurements of HOPG melted and probed by 20-ps laser pulses, Malvezzi *et al.*[8] observed *decreased* surface reflectivity following excitation above a critical melting fluence, which they interpreted as the formation of an insulating liquid phase. This interpretation was supported by an optoelectronic switching experiment[9] in which picosecond laser-melted HOPG functioned as the photoconductive gap material in a current transmission circuit, but challenged by authors[4] who argued that material ablation obscured the liquid surface within the 20 ps pulse duration, causing a reflectivity decrease unrelated to liquid carbon.

In an earlier Rapid Communication,[10] we briefly reported the first *femtosecond* time-resolved reflectivity experiments on HOPG. The current paper is a comprehensive report of these experiments, augmented with expanded data and analysis. Femtosecond time resolution provides the ultimate "inertial confinement" of a laser-melted sample, because there is no time in the first 1–2 ps, following femtosecond excitation for atoms to move away from the surface. Thus the optical integrity of the surface is preserved, and the question of hydrodynamic surface expansion, or ablation, does not arise during this time interval. At the same time, considerable evidence suggests that structural disordering, or melting, of the lattice[21,22] and equlibration of electron and lattice temperatures[23] progress substantially toward steady-state during this time interval. Thus a critical time window exists during which the optically smooth surface of a quasi-

 ©1992 The American Physical Society

steady-state liquid can be probed. Indeed the femtosecond experiments showed sharply *increased* reflectivity for several picoseconds following excitation above the critical melting fluence, which decayed to the much lower values observed in ps experiments[8] only after time delays of 10 ps or more. This pattern closely matched the reflectivity response of the much better understood silicon when irradiated at several times the melting fluence by a femtosecond pulse, where it is generally attributed to ultrafast melting into a metallic liquid followed by the slower process of surface expansion.[21,22] Thus these earlier fs experiments defined the time scale during which the optical properties of the unobscured liquid should be measured following melting by a femtosecond pulse.

In this paper the measurement and analysis of the optical properties of femtosecond-laser-melted carbon during the initial increased reflectivity stage is carried out in much greater detail. Specifically, time-resolved reflectance data at a range of probe wavelengths ($700 \gtrsim \lambda \gtrsim 310$ nm), oblique incidence angles, and pump fluences is reported. Furthermore a dielectric function based on the Drude model is fit to the data. From the model, Drude parameters—conduction electron density and collision frequency—are extracted. There are three motivations for such detail. First, careful reflectance measurements have historically provided one of the most important experimental keys to the underlying electronic properties of liquid metals.[24] Second, the recent appearance of molecular dynamics simulations[6] and a density functional theory[7] of the liquid state of carbon provide testable predictions of optical properties which can be compared directly to our data. Third, the low-frequency limit of the optical dielectric function can be related to previous measurements[4,14] and calculations[6,7] of the dc electrical resistivity of liquid carbon. Our results show much higher resistivity ($\rho \sim 600\ \mu\Omega$ cm) than ohmically heated carbon at low pressure[4] or liquid silicon and germanium,[25] but agree closely with measured resistivity ($\rho \sim 1000\ \mu\Omega$ cm) of carbon melted by ohmic heating at high pressure.[14] Additional new points of the current paper are time-resolved reflectance data of femtosecond-laser-melted diamond, which closely corroborate the graphite results, an analysis of the morphology of the damage spot produced above the critical melting fluence and its relationship to ablation, and a quantitative model of the delayed reflectivity decrease as a surface hydrodynamic expansion.

The paper is organized as follows. Section II discusses experimental methods. Section III presents experimental results of three kinds: (a) postmortem analysis of sample damage morphology, (b) self-reflectivity and -transmission of 90-fs pulses, and (c) pump-and-probe reflectivity and transmission measurements of HOPG and diamond at several probe wavelengths and polarizations. Section IV discusses the thermodynamic conditions which exist in the first picoseconds after intense femtosecond excitation: the rate of hydrodynamic surface expansion, the estimated extent of electronic and structural equilibration in the melt prior to surface expansion, and the estimated magnitude and time evolution of the transient temperature and pressure. Finally in Sec. V, we analyze the optical properties of liquid carbon by applying a Drude model to femtosecond reflectance measurements at $\Delta t \sim 1$ ps, after excitation, and relate this analysis to other measurements[4,14] and theories[6,7] of the liquid state of carbon.

II EXPERIMENTAL METHODS

Experiments were performed with samples of highly oriented pyrolytic graphite (HOPG) and 250 μm thick, type IIa (100) diamond obtained from Dubbledee Diamond Corporation. The basic material properties of both the HOPG and diamond are described in detail elsewhere.[5,26,27] Throughout this work, many of the measurements were repeated under identical conditions on a reference silicon sample, which, because of its structural similarity and extensive previous studies of its femtosecond melting dynamics[21,22,28] and equilibrium liquid state (see Ref. 43), served both as a calibration and an aid in interpreting results. All measurements were made in air. The surfaces of the HOPG samples (normal to the graphite c axis) were prepared by using tape to peel off the first few layers immediately before each set of measurements to expose a fresh surface. The diamond surfaces were prepared by standard optical polishing techniques. Laser pulses of 90 fs duration, centered at 620 nm, and up to 0.2 mJ energy were provided by a colliding pulse mode-locked (CPM) ring dye laser[29] followed by a four stage Nd:YAG pumped optical dye amplifier system[30] operating at a 10-Hz repetition rate.

When freshly prepared surfaces were excited by a 90-fs, 620-nm pulse above a critical fluence F_m, a damage spot appeared. The fluence of incident pump pulses was calibrated by measuring their energy with a calibrated photodiode, while measuring the focused spot size at the sample surface by monitoring transmission through a pinhole (mounted in the same plane as the sample) translated across the focal spot. At each of several fluences thus calibrated, a sampling of irradiated spots was examined under Nomarski and scanning electron microscopes, and F_m was defined as the minimum fluence at which any change in surface morphology attributable to the laser pulse (e.g., amorphization, cratering) was observed.[31] Measured values of F_m are reported and discussed in the next section. F_m of a reference silicon wafer, measured by the same procedure, agreed with the value 0.1 J/cm^2 reported by other investigators for 90-fs, 620-nm pulses.[21,22,29]

At fluences $F > F_m$, the diameter and depth of the craters created by irradiation pulses was measured as a function of fluence in HOPG, diamond, and the reference silicon sample. The average radius of the damage spot $r_{d,m}$ at each fluence for a common focal spot size was determined by examining a statistical sampling of damage spots under a Nomarski microscope. Damage crater profiles were mapped with approximately 30-Å resolution using an Alpha-Step depth profiler. These depth profiles help in estimating the amount of material removed from the surface, and thus in evaluating the role of ablation in the experiments.

Two types of surface reflectance measurements were

performed: (1) self-reflectance of single 90-fs, 620-nm, near-normally incident pulses as a function of fluence, and (2) pump-probe measurements using a 90-fs, 620-nm, normally incident pump pulse and 90-fs probe pulses of varying wavelength ($310 < \lambda_{probe} < 700$ nm), polarization, and incidence angle. In addition, with diamond, transmission measurements were performed. Self-reflectivity measurements were performed on HOPG by focusing pulses to 50-μm diameter at a 10° angle of incidence onto the sample surface, collecting and imaging[21] the reflected pulses, then detecting only the center of the pulse profile, which was selectivity apertured in the image plane. This procedure avoided spatial averaging over a transverse fluence distribution. Thirty shots were taken at each fluence, and a matched reference photodiode monitored the pulse energy incident on the sample to provide normalization against shot-to-shot energy fluctuations, permitting reliable measurement of reflectance changes smaller than 1%. The sample was translated 150 μm after each laser shot to ensure that each shot interrogated a fresh region of the surface. A computer controlled the acquisition and shot-to-shot subtraction of the signal and reference voltages and the movement of the sample raster stage. The onset of surface morphology changes provided in situ calibration of absolute laser fluence.

Pump-and-probe measurements monitored the temporal evolution of reflectivity changes induced by surface melting. For these experiments, a beamsplitter divided the amplified pulses into pump and probe pulses. The pump beam was sent through a computer controlled optical delay line and focused to a spot diameter of 60 μm (HOPG) or 30 μm (diamond) at normal incidence onto the sample surface. The probe beam was attenuated to well below F_m. Time-resolved reflectivity and transmission of diamond was probed only at near normal incidence with 620-nm pulses, because of the limited amount of sample available. Normal incidence reflectivity of HOPG was probed not only at 620 nm, but at different wavelengths generated either by frequency doubling in a thin, phase matched potassium dihydrogen phosphate (KDP) crystal (to produce 310-nm pulses) or by white light continuum[32] in a jet of ethylene glycol (to generate pulses at $700 < \lambda < 500$ nm). Band-pass filters selected the appropriate wavelength for detection. In addition reflectance was probed at an oblique incidence angle ($\theta = 50°$) with s and p polarizations. Steeper incidence angles, though valuable for characterizing the dielectric function precisely,[25] were avoided because of the much greater difficulty of fully overlapping the pump and probe spots on the sample, and thus in obtaining quantitatively accurate absolute reflectance values. The probe beam was focused to a spot diameter of approximately 30 μm (HOPG) or 15 μm (diamond) in the center of the pump spot. Acquisition and shot-to-shot normalization were accomplished as described above. As an additional normalization, each spot was probed twice: first without, then with the pump, accomplished by blocking, then transmitting alternate pump pulses with an electronically controlled shutter. The difference between these two signals, each separately normalized to the reference photodiode signal, measured the change in reflectivity (HOPG, diamond) and transmission (diamond). Multiple laser shots at each time delay were then averaged. This procedure corrected for possible surface nonuniformities as the sample was rastered, and significantly improved signal-to-noise ratio, reproducibility, and accuracy of the measured *absolute* values of surface reflectivity. With the reference silicon sample, this procedure yielded absolute reflectance within 5% of independently measured values (see Ref. 43).[21,22,25] Such measurement accuracy is critical to extraction of accurate Drude parameters, and thus to conclusions regarding the nature of the liquid state of carbon. The temporal zero-delay (t_0) between pump and probe was calibrated to within ±15 fs by monitoring the well-characterized reflectivity response[21,22,28] of the melting transition of the reference silicon sample.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. Postmortem analysis of sample damage

Using the procedure described above, we find the critical damage fluence F_m of HOPG to be 0.13 ± 0.02 J/cm^2 for single 620-nm, 90-fs optical pulses, very close to the value of 0.14 J/cm^2 determined by Malvezzi *et al.*[8] using 20-ps, 532-nm pulses. From known equilibrium material properties of HOPG,[5,26] we can calculate the final, equilibrated surface temperature T_f reached slightly below F_m (independent of the dynamical pathway by which it is reached) by equating absorbed optical energy to the integrated heat capacity:

$$(1-R)F_m\alpha = \rho \int_{300\ \mathrm{K}}^{T_f} C_p(T)dT \ . \tag{1}$$

We assume temperature-independent density $\rho = 2.25$ g/cm^3 and linear absorption ($\alpha^{-1} = 300$ Å), which is nearly the same for 620 and 532 nm.[26] A functional form of temperature-dependent heat capacity $C_p(T)$ below the melting temperature is given by Eq. (9) of Ref. 5, and pump pulse reflectivity $R \sim 0.3$ at F_m has been measured directly (see next section). The time constant for vertical heat diffusion out of the absorption volume is approximately 500 ps,[23,26] and can be neglected on the time scale (~ 1 ps)[23] on which an equilibrium temperature is established. We obtain $T_f \sim 6000$ K for both 20-ps and 90-fs pulses, 20–50% higher than the high pressure melting temperature,[2] of HOPG. However, the actual temperature is somewhat lower because the interband absorption saturates[23] as fluence approaches F_m, reducing the value of α by 30–50% in Eq. (1). This simple equilibrium analysis shows the HOPG indeed reaches the melting temperature at F_m, and suggests that melting occurs and causes the damage spot to form at $F > F_m$.

The damage threshold for diamond was measured to be 0.63 ± 0.15 J/cm^2 for 620-nm, 90-fs pulses. The fivefold higher value can be attributed to the requirement for simultaneous absorption of at least three 2.0-eV photons to cross the indirect energy gap E_g of 5.48 eV in diamond.[27] Self-reflectivity and transmission measurements presented below indeed confirm that strongly nonlinear absorption is present at fluences just below F_m, although

the final surface temperature T_f is therefore difficult to compute accurately because α depends strongly on the pulse energy and intensity profile.

Figure 1 presents measurements of the average damage spot radius as a function of the fluence of a normally incident 90-fs pump pulse focused to a spot radius of 30 μm on HOPG, or 15 μm on diamond. At F_m, a small damage spot (only a few microns in radius) appears in the center of the irradiated area. For $F > F_m$, the damage radius increases, at first rapidly, then more slowly with increasing fluence. For thermal melting by a pulse with Gaussian spatial profile $F = F_0 \exp(-r^2/\rho_0^2)$, the damage spot radius should increase with fluence as[33] $r_{dam} = \rho_0[\ln(F_0/F_m)]^{1/2}$, which is the fraction of the pulse cross-sectional profile in which $F > F_m$. The curves through the data points in Fig. 1 are fits of this formula to the measured damage radii using $\rho_0 = 30$ μm (HOPG) or 15 μm (diamond), which agree with the independently measured pump spot radius. From Fig. 1 we can determine that for $F > 0.4$ J/cm^2 (HOPG) or 1.5 J/cm^2 (diamond), the damage spot encloses $> 95\%$ of the energy of a Gaussian profile probe pulse focused as described above. However, below this fluence, any fluence dependence to the probe reflectivity must be attributed partly to the changing size of the melted region.

Figures 2(a) and 2(b) present measured depth profiles across the center of representative damage spots induced by 90-fs, 620-nm pulses in HOPG and silicon, respectively. Profiles of regions irradiated below F_m show no evidence of craters in either material. At 2.0 F_m (0.26 J/cm^2) in HOPG, a typical crater with a depth of ~ 300 Å is shown. In silicon, although amorphization, a signature of prior melting and rapid recrystallization, is clearly seen under a microscope in the irradiated region,[33] the depth profile still shows insignificant cratering at 2.0 F_m (0.2 J/cm^2). At 4.0 F_m, a typically deeper and wider crater is shown for HOPG, whereas the crater shown for silicon at 5.0 F_m is still small by comparison. The plot of average crater depth versus fluence (normalized to F_m) in Fig. 2(c) summarizes measurements of numerous damage spots in HOPG (filled squares) and silicon (open squares). For HOPG, *measurable craters are observed at all fluences above the damage threshold,* indicating that material has been lifted off the surface. The depths increase monotonically with increasing fluence, approaching 1000 Å at 12 F_m. In silicon, on the other hand, measurable craters are observed only for $F > 2.5F_m$, and increase in depth up to about $6F_m$. At fluences above about $10F_m$ in silicon, surface mounds appear, which may be caused by freezing of the molten silicon after being subjected to hydrodynamic shock from the laser pulse.[31] The higher threshold for crater formation in silicon can be attributed quantitatively[21] to the additional energy required to heat the stable liquid to the vaporization temperature, and to provide the latent heat of vaporization. Thus for $F > 2.5F_m$, the liquid is above the vaporization temperature. Most carbon phase diagrams,[2] on the other hand, show that the liquid phase exists in equilibrium only at pressures above 100 atm, and is thus thermodynamically unstable at atmospheric pressure, where it tends to sublime. Consequently, material loss to vaporization is expected as soon as the liquid phase is formed at F_m. With femtosecond excitation at $F > F_m$, material loss is probably enhanced compared to longer pulsed excitation, because of the initially elevated pressure (see Sec. IV).

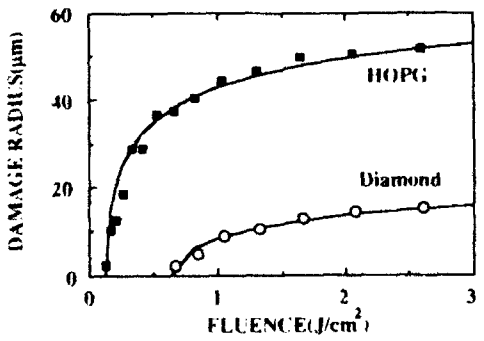


FIG. 1. Average radius of damaged area as a function of pump fluence for HOPG (solid squares) and diamond (open squares) after photoexcitation by 90-fs, 620-nm laser pulses focused to $1/e$ spot radius of 30 μm (HOPG) or 7.5 μm (diamond).

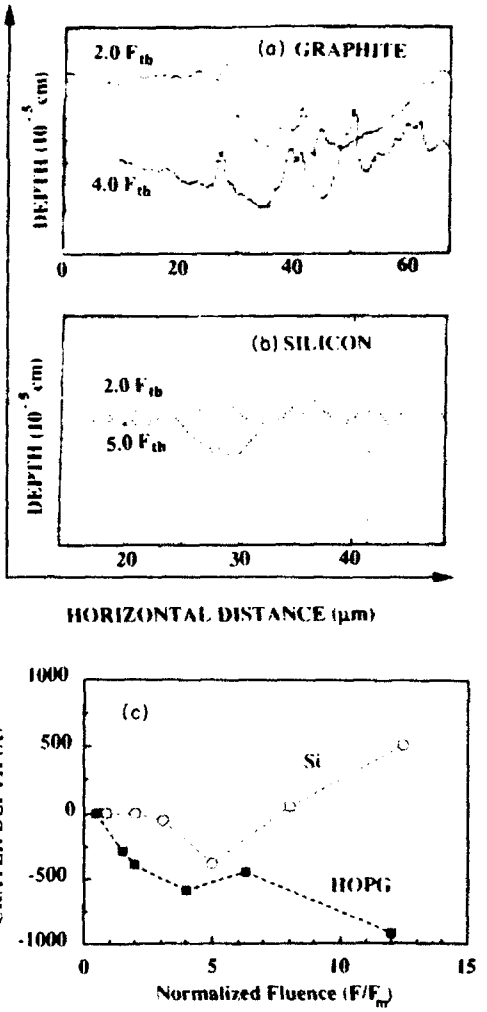


FIG. 2. Depth profiles of representative damage craters on (a) HOPG and (b) silicon for two fluences F above the critical melting fluence F_m. (c) Average crater depth as a function of F/F_m for HOPG (solid squares) and silicon (open squares).

B. Self-reflectivity and transmission

1. Graphite

Figure 3(a) shows the near-normal incidence self-reflectivity of the central part of the profile of 620-nm, 90-fs pulses from HOPG as a function of fluence normalized to F_m. Below F_m, the measured self-reflectivity is fluence independent and indistinguishable from the value of 0.3 measured for unexcited HOPG. Precisely at F_m, however, self-reflectivity begins to rise, reaching a values above 0.5 at $2F_m$. This result contrasts with the self-reflectivity response observed by Malvezzi *et al.*[8] using 532-nm, 20-ps pulses. where self-reflectivity *decreases* above F_m.

2. Diamond

Figure 3(b) shows the self-reflectivity (R) and self-transmission (T) of near-normally incident 620-nm, 90-fs pulses in diamond as a function of normalized fluence. In this case the entire reflected and transmitted pulse energies were monitored, and the plotted R and T values include the fraction of pulse energy which was multiply reflected between the faces of the 250-μm-thick sample. At very low fluences ($F<0.01F_m$), $R+T=1$, signifying the absence of absorption. In the range $0.01F_m<F<F_m$, R remains nearly fluence independent, while T drops sharply to less than $\frac{1}{2}$ of its low fluence value, a clear signature of multiphoton absorption. A fit of the data to a three-photon absorption model[11] yields a three-photon absorption coefficient $\gamma\sim 25$ cm$^3/TW^2$, although the data is not sufficiently accurate to rule out the simultaneous presence of higher order absorption nonlineraities. Precisely at F_m, self-reflectivity begins to increase sharply, just as observed in HOPG, while self-transmission continues to decrease.

The *precise* correlation between the self-reflectivity increase and the damage threshold in both samples suggests that the increase is caused by the initial dynamics of the melting process which occur during the pump pulse, as discussed further in Sec. IV. Measurements of the related process of *two*-photon absorption of 4.0-eV femtosecond pulses in diamond have been presented elsewhere.[34]

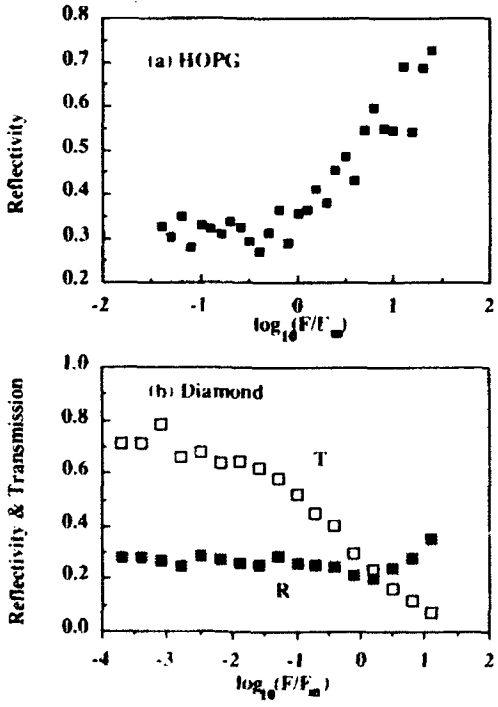


FIG. 3. (a) Self-reflectivity of 90-fs, 620-nm pulses normally incident on HOPG as a function of fluence F normalized to the critical melting fluence F_m. (b) Fluence dependent self-reflectivity and transmission of 90-fs, 620-nm pulses normally incident on diamond.

C. Femtosecond pump-probe measurements

1. Graphite

a. 620-nm wavelength probe, the first 25 ps. Figure 4 presents the time-resolved reflectivity of 620-nm probe pulses from HOPG for the first 25 ps following excitation above F_m as the pump fluence increases from 0.5 J/cm^2 to 20 J/cm^2. This time interval is long enough to show all of the major features of the reflectivity response with relatively coarse time resolution. Beyond $\Delta t=25$ ps, the visible reflectivity remains nearly constant for approximately a nanosecond, as also observed by Malvezzi *et al.*[8]

Figure 4(a) shows three reflectivity responses corresponding to a common pump fluence ($F=0.5$ J/cm^2), but to three different probe polarizations: s polarized at $\theta=50°$, near normal incidence, and p polarized at $\theta=50°$. The measured reflectivities for $\Delta t<0$ agree with Fresnel equation predictions when the published optical constants[26] $n=2.6$, $k=1.5$ for unexcited graphite at 620 nm are used. For $\Delta t\geq 0$, the inequality $R_s(\theta=50°)>R_p(\theta=10°)>R_p(\theta=50°)$ is maintained at all time delays. At $\Delta t=0$, reflectivity increases sharply with a pulse width limited rise time to a value roughly 50% higher than the initial value. There is a slight "overshoot" in the reflectivity, evident as only a single data point on this coarse time scale, which recovers in a fraction of a picosecond. There follows a slower decay to the initial value in ~ 8 ps, then a continuing drop to a final steady state value in 20 to 40 ps. From the three measured reflectivities, the complex index of refraction $n+ik$ can be determined uniquely at any time delay from Fresnel's equations, provided that the surface is assumed to remain perfectly discrete. Thus at 0.3 ps$<\Delta t<1$ ps, we obtain $n=1.35$, $k=1.93$, showing increased metallic character. At time delays greater than several picoseconds, however, the assumption of a discrete interface may no longer be valid, as discussed further in Sec. IV.

Figures 4(b) and 4(c) show the reflectivity response probed at near normal incidence at progressively higher pump fluences of 2.0 and 3.9 J/cm^2, respectively. Again reflectivity increases sharply at $\Delta t=0$, reaching a maximum value which increases slightly with pump fluence. The subsequent drop in reflectivity depends strongly on pump fluence. The reflectivity decays more rapidly, and

the final reflectivity value becomes smaller, as pump fluence increases. This trend continues at higher fluences, as shown by the oblique incidence probe reflectivity data at 20 J/cm^2 pump fluence in Fig. 4(d) (data for a near normally incident probe is not shown in this figure). The general pattern of a sharp reflectivity increase followed by a slower, but fluence-dependent decrease closely resembles the reflectivity response of silicon[21,28] when melted by femtosecond pulses above the vaporization fluence $F > 2.5F_m - 0.25$ J/cm^2. The main difference is that in silicon the reflectivity decay is not observed for pump fluences $F_m < F < 2.5F_m$ between the melting and vaporization fluences, whereas in HOPG, the decay is observed at all fluences $F > F_m$. Thus the reflectivity decay correlates with the appearance of damage craters described above, suggesting that hydrodynamic surface expansion causes the reflectivity decay.

It is helpful to fit the observed reflectivity responses in Figs. 4(a)–4(d) phenomenologically to a function $R(\Delta t)$, which convolves a single exponential response function $r(t) = \exp[-(t-t')/\tau_a] - C$ with the temporal intensity envelopes $P_{\text{pump}}(t)$ and $P_{\text{probe}}(t)$ of the pump and the probe pulses:

$$R(\Delta t) = \int_{-\infty}^{\infty} P_{\text{probe}}(t+\Delta t) \int_{-\infty}^{t} P_{\text{pump}}(t') r(t') dt' dt \ . \tag{2}$$

τ_a is the relaxation time constant, and the modified baseline C represents the final reflectivity value. The pump and probe intensity envelopes are conveniently represented as $\text{sech}^2(t/\tau_p)$. Examples of fits to Eq. (2), using the measured pulse width $\tau_p = 90$ fs, are shown by the curves superimposed on the data in Figs. 4(b) and 4(c) where time constants $\tau_a = 6$ and 4 ps, respectively, were used. Fits of equal quality are obtained for other fluences. In Fig. 5 these decay time constants obtained in the same manner from femtosecond reflectivity data on silicon photoexcited at $F > 0.25$ J/cm^2 are also plotted (solid circles). For similar fluences, similar time constants are obtained in HOPG and silicon. The solid line is a theoretical curve which predicts a characteristic time constant for hydrodynamic surface expansion, which decreases with laser fluence in a manner very similar to τ_a. This model is discussed further in Sec. IV.

Equation (2) can also be used to calculate the reflectivity response $R(\Delta t)$ which would be observed in an experiment with longer pump and probe pulses. The grey-tone curves in Figs. 4(b) and 4(c) were obtained using the same response function $r(t)$, but $\tau_p = 20$ ps, corresponding to the pulse duration used by Malvezzi *et al.*[5] Clearly the initial reflectivity increase cannot be observed at all with this time resolution. In addition, the fluence dependence of the reflectivity drop is largely masked. Consequently the calculated response shows only a

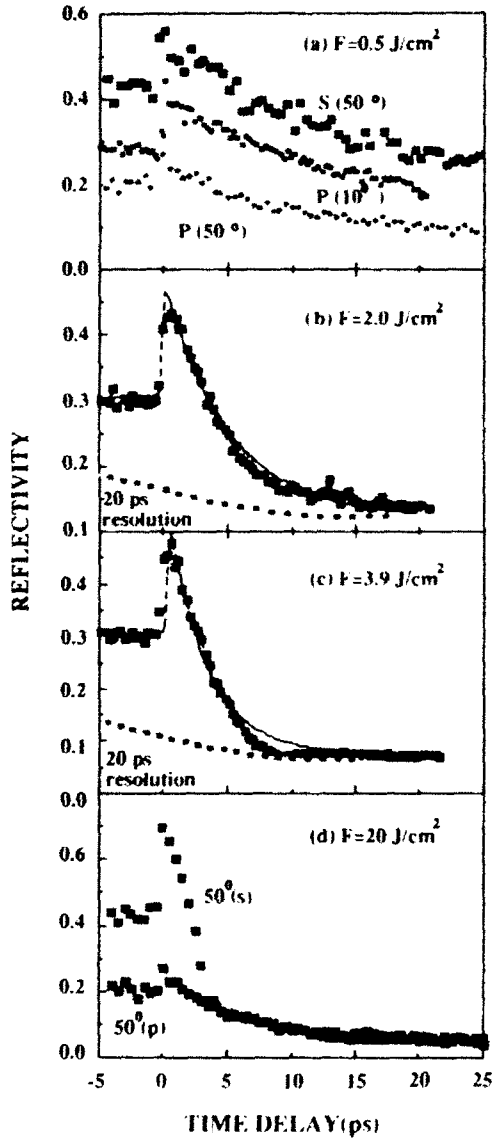


FIG. 4. Femtosecond time-resolved reflectivity of HOPG. (a) Excitation at pump fluence $F = 0.5$ J/cm^2, probe at 620 nm, comparing 50° incident angle s and p polarized reflectivity, and near-normal incidence reflectivity. (b) Near-normal incidence 620-nm reflectivity at $F = 2.0$ J/cm^2; dashed line shows convolution of data with 20-ps pump and probe pulses. (c) Same with $F = 3.9$ J/cm^2. (d) Excitation at 20 J/cm^2, probe at 620 nm, comparing 50° incident angle s and p polarized reflectivity. Vertical scales on all panels denote absolute reflectivity.

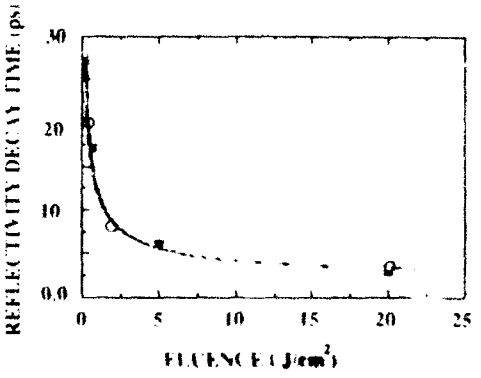


FIG. 5. Fluence dependence of the decay time constant of the reflectivity response of HOPG (open circles) and silicon (solid circles). The solid line is a theoretical fit to a $F^{-1/2}$ dependence, as discussed in the text.

monotonic reflectivity drop following photoexcitation, consistent with the picosecond data of Malvezzi *et al.*[8]

One additional feature of the Fig. 4 data is noteworthy. Maximum *s*-polarized reflectivity, observed at $\Delta t \sim 1$ ps, increases form 0.58 at 0.5 J/cm^2 [Fig. 4(a)] to 0.69 at 20 J/cm^2 [Fig. 4(d)], a trend also observed in normal incidence reflectivity. Maximum *p*-polarized reflectivity, on the other hand, *decreases* from 0.3 (0.5 J/cm^2) to 0.26 (20 J/cm^2). The selective suppression of *p*-polarized reflectivity at such a high fluence as 20 J/cm^2 probably results from a significant degree of ionization, and the formation of a short scale length plasma density gradient at the surface. Consequently resonance absorption selectively attenuates *p*-polarized light.[35] Similar effects have been observed and analyzed by Milchberg *et al.*[36] Since *s*- and *p*-polarized reflectivity increase about equally at lower fluences [e.g., Fig. 4(a)], however, we can assume that the surface is discrete at early time delays.

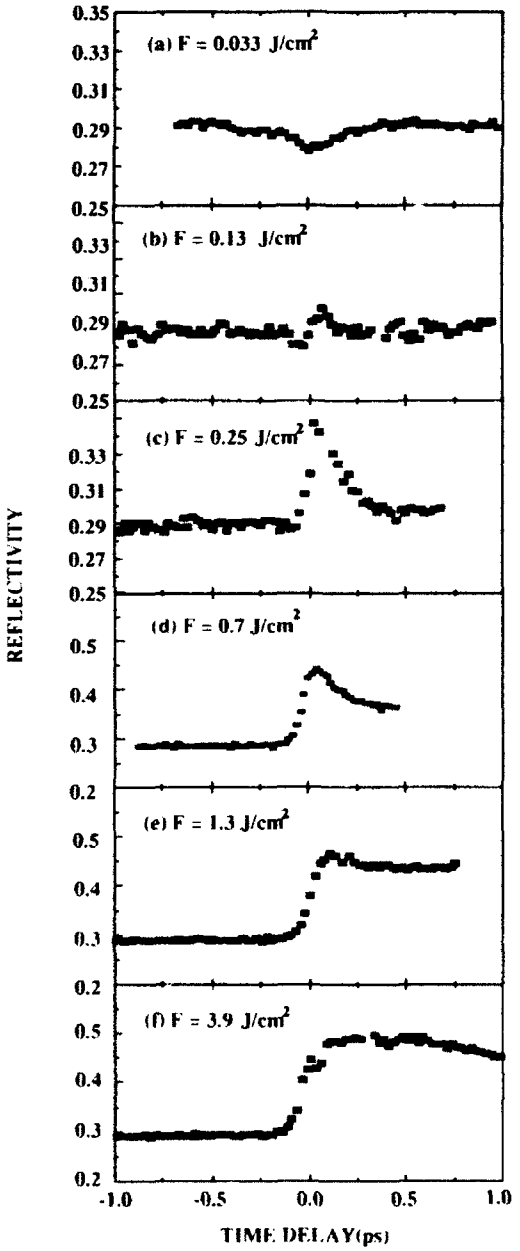


FIG. 6. The first picosecond of the time-resolved reflectivity response of HOPG probed by 90-fs, 620-nm pulses at near normal incidence. (a) Excitation at pump fluence F below critical melting fluence F_m; (b) $F = F_m$; (c)–(f) increasing fluences above F_m. Vertical scales denote absolute reflectivity.

b. 620-nm probe wavelength, the first picosecond. Fine time-scale data over the first picosecond documents the initial stage of photoexcitation and melting dynamics. Figure 6 shows the first picosecond of the reflectivity response probed at near-normal incidence as the pump fluence is varied from 0.033 to 3.9 J/cm^2. Just below the threshold [$F = 0.033$ J/cm^2, Fig. 2(a)], a small decrease in reflectivity is observed during the pump pulse, which recovers in ~ 200 fs. This response is caused by the generation, relaxation, and recombination of an electron-hole plasma, and is in good quantitative agreement with the extensive reflectivity measurements below F_m reported by Seibert *et al.*[23] When the pump fluence is increased to exactly 0.13 J/cm^2, a nearly flat response is seen, as shown in Fig. 2(b). As soon as the fluence is increased above F_m, a sharp reflectivity *increase* is observed [$F = 0.25$ J/cm^2, Fig. 2(c)], followed by a fast (~ 200 fs) recovery and a small residual reflectivity increase, which persists for over 10 ps. For the data in Figs. 2(b) and 2(c), the damage spots were smaller than the probe spot. Consequently, the measured response averages the actual response below and above F_m to some extent. At higher fluences, however, essentially the entire probe pulse sees material excited above F_m. At $F = 0.7$ J/cm^2 and 1.3 J/cm^2 [Figs. 2(d) and 2(e)], the fast transient is still evident, but the persistent reflectivity component increases sharply in magnitude to over 0.4. At still higher pump fluence [$F = 3.9$ J/cm^2, Fig. 2(f)], the fast transient is no longer evident, and a slight further increase to $R \sim 0.5$ occurs.

Two important features of the initial reflectivity response above F_m are noteworthy. First, the rise times of reflectivity signals in Figs. 6(c)–6(e) are all pulse width limited. This contrasts with a resolvable delay of ~ 200 fs in the rise time of the reflectivity of silicon when photoexcited just above F_m.[22] This comparison suggests that the initial response of HOPG above F_m is caused by an overdense electron-hole plasma, which rapidly equilibrates, whereas the electron-hole plasma in silicon fails to reach critical density just above F_m, yielding a response time governed by the melting dynamics. Second, the fast transient observed above F_m, though opposite in sign, recovers on the same time scale as the transient observed below F_m [Fig. 6(a) and Ref. 23]. This comparison suggests that the former, like the latter,[23] is also caused by equilibration of hot electrons with the lattice, as discussed further in Sec. IV.

c. Other probe wavelengths. A reflectivity spectrum over as wide a range of wavelengths as possible is key to modeling the underlying electronic structure of liquid carbon. Figure 7 shows the time-resolved reflectivity of 310-nm, probe pulses from HOPG following excitation below F_m [Fig. 7(a)] and at two fluences above F_m [Figs. 7(b) and 7(c)]. The response differs substantially from that at 620 nm. Following sample excitation at 0.11

[Fig. 7(a)], ultraviolet reflectivity first decreases for the first 100 fs, then recovers within 2 ps to a steady-state value 4% higher that the initial reflectivity, which persists for ~500 ps. A detailed interpretation of this response has been presented elsewhere.[23] The initial decrease is caused by the rapid generation and energy relaxation of electrons and holes, while the slowly rising reflectivity is caused by the thermal renormalization of the graphite π energy bands,[23] and provides a measure of the time required for full electron-lattice equilibration. The ultraviolet reflectivity response above F_m differs significantly from the response when excited below F_m, providing further evidence that an irreversible phase transition occurs at F_m. At $F = 0.7$ J/cm^2, no reflectivity increase whatsoever is observed with the ultraviolet probe, in sharp contrast to the 620-nm, probe [Figs. 4(a) and 6(d)]. Instead, reflectivity remains unchanged for the first picosecond, then monotonically decreases within 20 ps to a steady-state value ($R = 0.05$), which persists for at least 300 ps. When pump fluence is increased, the reflectivity drops more rapidly, and to a lower final value, the same trends observed with the 620-nm, probe. In addition, when the pump fluence is increased to 4 J/cm^2, a slight transient reflectivity increase (barely above the noise level) is observed during the first ~0.5 ps, as shown on the magnified scale of Fig. 7(b). This transient evolves on approximately the same time scale as the reflectivity transients shown in Figs. 6(a)–6(e) and 7(a). We attribute this to the initially nonequilibrium electron-hole plasma which at this fluence has become sufficiently dense to be detected at 310 nm.

Figure 8 shows fluence dependent reflectivity measurements made at fixed time delays of 0.7 and 20 ps, using a white light continuum probe, where wavelengths of 550 nm [Fig. 8(a)], 620 nm [Fig. 8(b)], and 700 nm [Fig. 8(c)] were selected for detection. A delay of 0.7 ps probes the high reflectivity portion of the evolution after the initial transient 200-fs features, while a delay of 20 ps probes the approach to a low steady-state reflectivity. In both cases, no change in reflectivity occurs until the critical melting threshold is reached. At $\Delta t = 0.7$ ps, the reflectivity begins to rise precisely at the critical melting fluence at *all* probe wavelengths, reaching a final value of $\sim 1.5R_0$ at $10F_m$. In addition, the reflectivity scales equally with

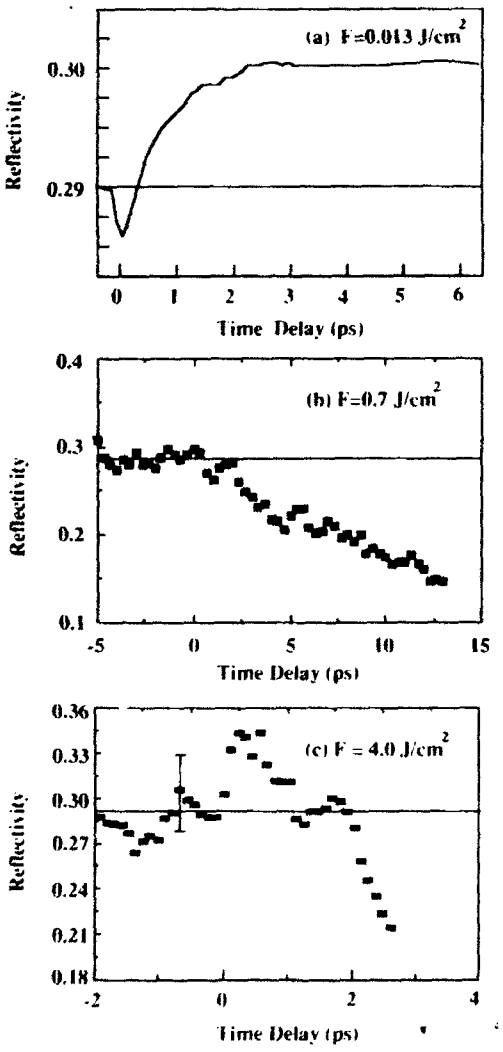


FIG. 7. Time-resolved reflectivity response of HOPG probed by 90-fs, 310-nm pulses at near normal incidence for three fluences F of the 90-fs, 620-nm pump pulse: (a) $F = 0.1F_m$, (b) $F = 5F_m$, (c) $F = 30F_m$.

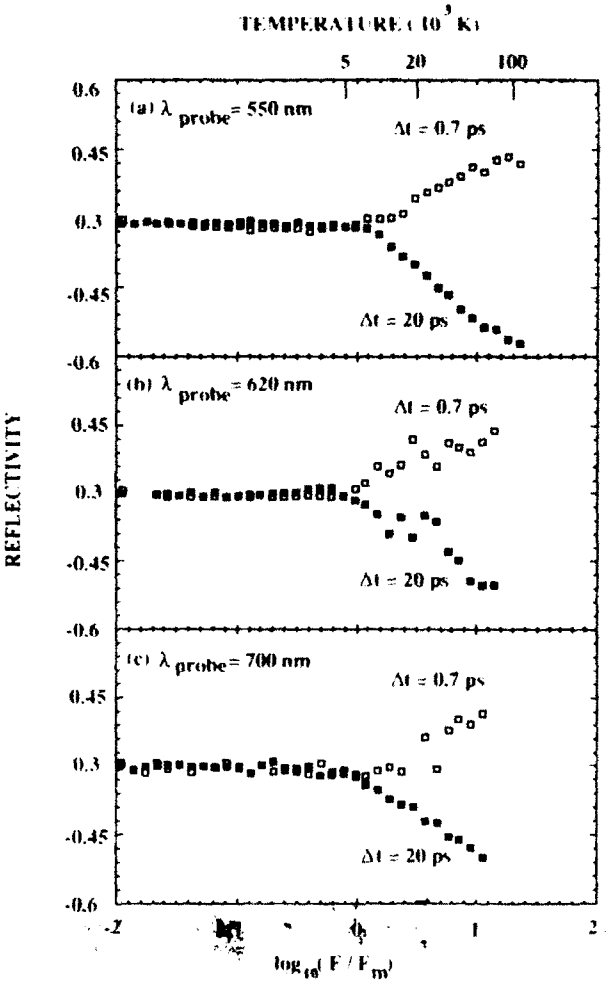


FIG. 8. Reflectivity of HOPG as a function of normalized pump fluence F/F_m for probe delays of 0.7 and 20 ps, for probe wavelengths (a) 550 nm, (b) 620 nm, (c) 700 nm.

fluence in each case. Spot size effects contribute to the increasing magnitude of the reflectivity in the fluence range $F_m < F < 3F_m$. Above $3F_m$, however, the increasing reflectivity is caused entirely by changing properties of the material as its temperature increases. The scale at the top of Fig. 8(a) gives a rough estimate ($\pm 50\%$) of sample temperature at $\Delta t = 0.7$ ps derived by an extension of Eq. (1), as discussed further in Sec. IV. At a given fluence (temperature) reflectivity is slightly higher at longer wavelengths. At $\Delta t = 20$ ps, for all wavelengths, the reflectivity drops below the background HOPG value precisely at the critical melting fluence, and decreases monotonically with fluence. A very slight rise in reflectivity is observed below the critical melting fluence at 620 nm, as observed in previous studies[23] below F_m, and is related to electron-hole carrier generation.

2. *Diamond*

Below F_m very different femtosecond responses are expected in graphite and diamond, because of the very different electronic band structures. Above F_m, however, we expect to measure similar optical properties, if the two samples melt and equilibrate to a common liquid phase on the time scale of the measurements.

Figure 9 shows the time-resolved reflectivity and transmission of 620-nm probe pulses from diamond following excitation just below F_m [Fig. 9(a)] and well above F_m [Figs. 9(b) and 9(c)]. For these measurements, the probe is incident at a large enough angle ($\sim 20°$) that the front surface reflection has been spatially separated from the back surface reflection. Below F_m [Fig. 9(a)], no change in front surface reflectivity was detectable. However, transmission drops to 0.45 immediately upon photoexcitation, with negligible recovery over tens of picoseconds. We believe this response to be free carrier absorption in an electron-hole plasma generated by three-photon absorption. The recovery rate is severely limited compared to graphite because of the large energy gap which separates electrons and holes following their relaxation to band edges. Above F_m, we observe on a coarse time scale [Fig. 9(c)] a reflectivity response qualitatively very similar to the graphite responses shown in Figs. 4: a sharp initial increase followed by a slower decay over ~ 10 ps. The lower peak reflectivity value (~ 0.3) compared to HOPG excited at $10F_m$ (~ 0.45) is partly an artifact resulting from the coarse time spacing of data points used to conserve sample area in taking the Fig. 9(c) data. When repeated at the same fluence ($F = 9F_m$) with finer time resolution over a shorter time interval [Fig. 9(c)], the data indeed show that diamond reflectivity increases from 0.17 to approximately 0.4, very close to, although still slightly lower than, the value observed with HOPG. This small discrepancy may be a real difference in the reflectivity of the liquid phases, perhaps because the liquid initially retains the higher density of diamond (3.5 g/cm^3) compared to graphite (2.2 g/cm^3).

The corresponding transmission response of diamond excited at $3F_m$ (solid squares) and $6F_m$ (open squares) is shown in Fig. 9(b). Transmission decreases abruptly at $\Delta t = 0$ from its unexcited value (0.7) to less than 0.2. On the time scale (25 ps) on which reflectivity undergoes large changes, no further change in transmission is observed. The slightly higher final transmission value at $3F_m$ results from the probe spot being slightly larger than the damage spot. A similar transmission response is observed with silicon-on-sapphire upon melting with a femtosecond pulse.

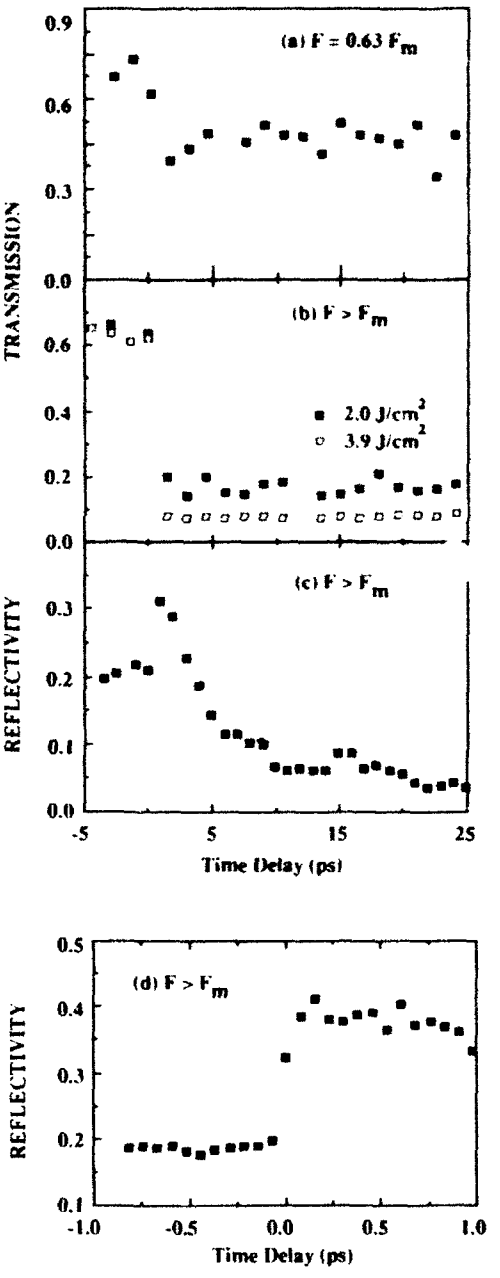


FIG. 9. Femtosecond time-resolved front surface reflectivity (R) and transmission (T) of diamond, probed at near normal incidence by 90-fs, 620-nm pulses. (a) Pump fluence $F = 0.6F_m$; (b) T at $F = 3F_m$ (solid squares) and $6F_m$ (open squares); (c) R at $F = 9F_m$ on coarse time scale; (d) detail of first picosecond of reflectivity response at $F = 9F_m$.

IV. DISCUSSION

A. Hydrodynamic surface expansion

In order to obtain meaningful optical data on the liquid phase, it is critical to collect data before surface optical properties deteriorate because of hydrodynamic expansion. Femtosecond experiments remove all ambiguity that the surface can be probed before the surface expands significantly. The consistent observation of an early reflectivity increase (unresolvable in ps experiments[8]) followed by a fluence dependent decrease is therefore simply and naturally explained as initial creation of a highly reflective condensed phase, which is subsequently obscured by surface expansion. Indeed there is general agreement[8,10,37] that at fluences $F > 5F_m$, ablation attenuates surface reflectivity within several picoseconds of photoexcitation.

Nevertheless, some authors[8,37] have argued that during a 20-ps melting pulse in the limited fluence range $F_m < F < 5F_m$, insufficient material can evaporate to affect surface reflectivity significantly, and that reduced reflectivity in this fluence range must therefore be ascribed to intrinsic (electrically insulating) properties of the condensed surface. Analogously, in our previous paper,[10] we did not rule out the possibility that the delayed reflectivity decrease for $F_m < F < 5F_m$ observed in fs experiments could be caused by a delayed metal-to-insulator transition within the liquid, which maintained a discrete, unobscured interface.

Several cogent arguments can now be advanced against this interpretation of the fs experiments. First, if the reflectivity decrease corresponded to an unobscured insulating material developing, a transmission recovery on the same time scale as the reflectivity decrease would be expected. Yet no such recovery is observed in the diamond transmission measurements at any pump fluence [Fig. 9(b)]. Instead transmission drops sharply upon photoexcitation to a value which remains constant for over 25 ps. On the other hand, this pattern is consistent with the early formation of a glossy, reflective liquid and subsequent surface expansion. Such transmission measurements were not available in the previous report on graphite.[10] Second, the careful damage morphology analysis now shows that delayed reflectivity decreases following fs melting in both graphite and silicon are *precisely correlated* with appearance of a damage crater, and thus with removal of material from the surface at some time after irradiation. Specifically, in silicon, damage craters and delayed reflectivity decreases[21] both are observed only for $F > 2.5F_m$, the fluence required to heat the liquid to the boiling temperature and provide the latent heat of vaporization; correspondingly in graphite, damage craters and a delayed reflectivity increase both occur for $F > F_m$, suggesting that liquid carbon spontaneously ablates at atmospheric pressure. Third, the time constants of the reflectivity decrease (Fig. 5), have nearly the same values and fluence dependence for femtosecond-irradiated carbon and silicon, suggesting a common ablative mechanism. Finally, no theoretical basis exists for structural or electronic equilibration within the liquid on a 10–30-ps time scale. On the contrary, recent molecular dynamics simulations of the liquid state of carbon[6] show that the liquid reaches steady state structural and electronic properties within ~ 1 ps of heating to the melting temperature.

In order to reconcile the observation of a strong reflectivity decrease within ~ 10 ps with the very small evaporation rate expected[8,37] at $F < 5F_m$, we must reconsider the microscopic mechanism for the reflectivity decrease. The "evaporation" model described in Refs. 8 and 37 in the context of ps irradiation depicts the escape of individual atoms from the liquid to form an absorptive vapor above the liquid surface, which is assumed to retain an atomically sharp interface. Indeed optical absorption from the expected vapor density at $\Delta t \sim 10$ ps is clearly inadequate to account for the observed reflectivity decrease at $F < 5F_m$. On the other hand, recent studies of metal surfaces irradiated by intense, femtosecond pulses[35,36] show that strong reflectivity decreases within a few picoseconds are caused by hydrodynamic expansion of the surface as a whole, because of the high transient pressure created by femtosecond irradiation (Sec. III C below). The initially atomically abrupt interface develops a short scale length density gradient shortly following excitation, similar to the hydrodynamic expansion of a surface experiencing a strong shock wave.[38] When the gradient scale length L approaches a significant fraction of an optical wavelength λ, surface reflectivity can no longer be described by Fresnel's equations, which are based on boundary conditions for an abrupt interface, but instead requires numerical treatment via the Helmholtz wave equation.[36] Typical calculations[35,36] show strong reflectivity decreases for $L > 0.2\lambda$, with complicated dependence on polarization, wavelength, material, and the precise form of the density gradient.

In order to estimate the time τ for such a density gradient to form under our experimental conditions, we use the expression for hydrodynamic expansion velocity of a surface experiencing a strong shock wave:[38]

$$\langle v \rangle = \frac{2}{\gamma - 1} \left| \frac{ZkT}{M} \right|^{1/2}, \tag{3}$$

where $\langle v \rangle$ is the average particle velocity in the normal direction, $\gamma = C_p / C_v$ is the ratio of electronic heat capacities at constant pressure and volume, Z is the effective ion charge, T is the surface (electron) temperature, and M is the mass of a carbon atom. Estimating $\gamma \sim 1.6$ and $Z \sim 4$ for carbon at high temperatures,[39] we obtain $\langle v \rangle \sim 2[k_B T(\text{eV})]^{1/2} \times 10^6$ cm/s, close to the speed of sound. To a fair approximation, $k_B T \sim 0.5$ eV at $F = F_m$, and $k_B T \propto F$ at higher fluences. Thus at $F = 4F_m$, for example, a scale length $L = 0.5\lambda = 310$ nm develops in a time $\tau = L / \langle v \rangle \sim 10$ ps, consistent with the time scale of the reflectivity decrease in Fig. 4(a). To obtain a fit to the fluence dependence of the exponential decay constants τ plotted in Fig. 5, we assume that surface reflectivity is reduced by e^{-1} when an L develops over a fraction $0.1 < f < 1$ of the probe wavelength λ, where f can be treated as a fitting parameter which depends on polarization, wavelength, material, and the form of the density

gradient. The time for the necessary gradient to form is then $\tau = f\lambda/\langle v \rangle$. As the fluence (and therefore T) increases, $\langle v \rangle$ increases as $(k_B T)^{1/2}$, resulting in a reduced gradient formation time. The solid (graphite) and dashed (silicon) theoretical curves in Fig. 5 are fits to the data obtained using $f = 0.6$ (graphite) or 0.15 (silicon), corresponding to $L = 3720$ Å (graphite) or 930 Å (silicon).

Thus both the magnitude and fluence dependence of the reflectivity decay time constant τ in both silicon and graphite (Fig. 5) are consistent with the one-dimensional expansion rate of condensed material surface into a vacuum after being pressurized and heated by the absorption of a femtosecond pump pulse. Significantly, the fluence dependence of τ for both $F_m < F < 5F_m$ and $F > 5F_m$ fit naturally to a *single model*, further confirming that the delayed reflectivity decreases in both fluence ranges of fs excitation are caused by a common mechanism of surface hydrodynamic expansion. Thus if the influence of hydrodynamic expansion is to be ruled out unambiguously, the optical properties of a condensed phase driven to elevated temperature and pressure by a femtosecond pulse must be probed within a critical time window of no more than a few picoseconds.

B. Electron-lattice and structural equilibration

A key question then becomes: to what extent has the photoexcited carbon equilibrated within the available time window? Several lines of evidence suggest that electron and lattice temperatures substantially equilibrate within ~ 1 ps following photoexcitation of the carbon sample. First, femtosecond spectroscopy of graphite excited *below* F_m shows[23] that photoexcited carriers relax almost completely within 1 ps. The graphite lattice temperature is measured directly by the rising UV reflectivity response in Fig. 7(a),[23] which confirms that, below F_m, the lattice reaches its maximum temperature within 2 ps, and has nearly reached the maximum within 1 ps. Similar or shorter electron-lattice equilibration times have been measured for many metals.[40] As the lattice melts, the increased disorder should further accelerate electron-lattice equilibration. as observed, for example, in comparisons of carrier lifetimes in crystalline,[41] amorphous,[42] and liquid[25] silicon.

Second, the subpicosecond reflectivity transient observed upon photoexciting graphite *above* F_m [Figs. 6(b)–6(e) and 7(c)] recovers on a similar time scale to the transient below F_m. In our previous paper,[10] we showed that this reflectivity "overshoot" at 620-nm, probe wavelength [Figs. 6(b)–6(e)] is explained by the instantaneous generation of an overdense electron-hole plasma ($N_{e\text{-}h} > 10^{22}$ cm^{-3}) upon photoexciting HOPG at $F \geq F_m$. With an ultraviolet probe [Figs. 7(b) and 7(c)], higher pump fluence ($\sim 10F_m$) and carrier density ($N_{e\text{-}h} > 4 \times 10^{22}$ cm^{-3}) is required to see this overshoot, as expected from the Drude dielectric function,[10] and confirms the interpretation of the overshoot as a short-lived electron hole plasma. The common subpicosecond recovery time of these reflectivity transients both below and above F_m supports the interpretation that electron and lattice subsystems equilibrate to nearly a common temperature within 1 ps.

Third, molecular dynamics (MD) simulation of the liquid state of carbon by Galli *et al.*[6] show that structural equilibration is also very rapid: the structural and electronic properties of an initial amorphous carbon structure heated to 5000 K in the electronic ground state reach a steady-state liquid structure in well under 1 ps. Simulations which begin by heating a crystalline graphite or diamond structure might yield a somewhat longer equilibration time, and would be of great interest in interpreting femtosecond experiments. Nevertheless our observation of very similar subpicosecond reflectivity evolution of graphite and diamond samples excited at $F > F_m$ [Figs. 6 and 9(c)] strongly suggests that both have equilibrated to a common liquid phase at $\Delta t \sim 1$ ps.

Finally, in a well-studied material as silicon, subpicosecond reflectivity transients are also observed,[28] but reach steady-state values within 1 ps[21,22] which are fully consistent with the independently measured optical properties of fully equilibrated, over-melted silicon.[25,43] In fact, femtosecond and picosecond experiments at $F_m < F < 2.5F_m$ show a constant reflectivity for 0.5 ps $< \Delta t <$ 1 ns or more.[21,25] Femtosecond surface second harmonic measurements[22] show that long-range crystalline order is lost in 0.5 ps, demonstrating that the measured reflectivity is indeed that of a liquid phase.

Available evidence thus supports the conclusion that optical properties probed at $\Delta t \sim 1$ ps are those of a liquid state of carbon with substantially equilibrated electron and lattice subsystems. At very high fluences $F \gg F_m$, the term "solid density plasma"[35,36] may be more appropriate than "liquid," in recognition of the higher ionization states[39] which accompany higher temperature and pressure conditions. However, there is no sharp demarcation between the two regimes, and both may be encompassed by the term "fluid."

C. Temperature and pressure following femtosecond excitation

For $F > F_m$, the quasiequilibrium temperature T_f of the liquid established by $\Delta t \sim 1$ ps can, in principle, be calculated by a generalization of Eq. (1) which takes into account variations in the absorption coefficient $\alpha(t)$ during the pump pulse, and the heat capacity $C_{\text{liquid}}(T)$ of the liquid phase. In practice, the variations in $\alpha(t)$ are quite complicated and $C_{\text{liquid}}(T)$ can only be estimated, thus precluding any more than a rough estimate of T_f of the liquid. The significant saturation (decrease in α) which occurs at fluences approaching F_m for HOPG has already been noted. On the other hand, above F_m, the onset of melting and metallic properties probably increases α later in the pump pulse. Thus as a compromise the linear absorption depth $\alpha^{-1} = 300$ Å of HOPG provides an approximate basis for estimating T_f. On the short time scale of $\Delta t \sim 1$ ps, volume expansion can be neglected ($\rho = \rho_0 = 2.25$ g/cm^3 for HOPG), and the heat capacity C_v at constant volume should be used. The specific heat $(3/2)Nk_B$ of a classical monatomic gas provides a basis for estimating for $C_{\text{liquid}}(T)$. Internal vibrational modes in the liquid would, of course, increase this value toward the Dulong-Petit value of $3Nk_B$. The mo

sured pump self-reflectivities $R(F_m)$ in Fig. 3(a) can be used directly in Eq. (1). Combining these approximations, the temperature estimates on the upper horizontal axis of Fig. 8 were obtained, and indicate that temperatures as high as 100 000 K can be reached under our experimental conditions. The expected accuracy of this scale, however, is no better than ±50% in view of the above approximations.

When the sample is heated and probed on a time scale shorter than the characteristic time for thermal expansion, transient pressure exists because the sample is inertially confined to a constant volume. The magnitude of the pressure depends on the degree of equilibration within the phonon system. As an upper limit which assumes an equilibrated phonon system, the pressure p of a solid heated to $k_B T < 10$ eV (assuming $T_e = T_i = T$) at the constant density ρ_s of the low temperature solid is determined primarily by anharmonic ionic motion, and is described by the approximate formula $p \sim p_i \sim 3N\gamma k_B T$, where γ is the Gruneisen parameter.[44] More precisely, total pressure p also includes contributions from Thomas-Fermi electron pressure, and a correction for chemical bonding, but for $k_B T < 10$ eV and $\rho \sim \rho_s$, the ionic contribution p_i strongly dominates.[44] The dashed curves in Fig. 10 plot $p_{\rm ion}(T)$ for liquid carbon for fixed densities corresponding to graphite (2.25 g/cm^3) or diamond (3.5 g/cm^3). The solid curves show $p(T)$ corrected for the electronic and bonding contributions [derived from Eqs. (81) and (106) of Ref. 44]. For completeness, the curves have been extended to $T < T_m$, using the corresponding expression for $p_{\rm ion}(\rho, T)$ for a high temperature solid [Eq. 34(b) of Ref. 44]. Figure 10 shows that transient pressures of approximately 1 megabar can exist at $T > T_m$. The time scale $\tau_{\rm rel}$ for release of the pressure is estimated from the ratio of the heated (absorption) depth α^{-1} to the speed of sound c_s. For HOPG, we estimate $\tau_{\rm rel} \sim \alpha^{-1}/c_s \sim 5$ ps, using $\alpha^{-1} \sim 500$ Å and $c_s \sim 10^6$ cm/s. This estimate confirms that solid density, and thus elevated pressure, still exist at $\Delta t \sim 1$ ps, and simultaneously supports the argument that the reflectivity decrease observed at $\Delta t > 5$ ps corresponds to the expansion of initially pressurized material. Rothenberg[45] has measured such a pressure release time in picosecond-irradiated solid silicon using time-resolved photothermal deflection spectroscopy.

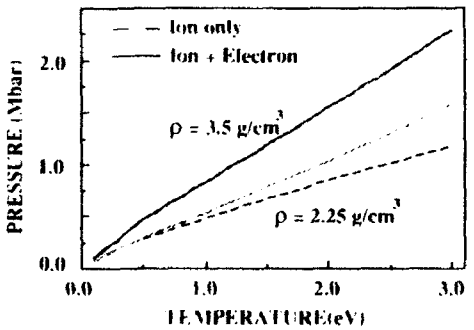


FIG. 10. Effective pressure of carbon sample following constant volume heating at density of graphite (2.25 g/cm^3) or diamond (3.5 g/cm^3), calculated using approximate equation of state formulas described in Ref. 44. Dashed curves include the dominant ionic contribution only; solid curves include both ionic and Thomas-Fermi electronic contributions to pressure.

V. DRUDE DIELECTRIC FUNCTION OF LIQUID CARBON

We now analyze in detail the measured optical properties of the fluid state. This analysis is compared directly to optical properties predicted by molecular dynamics simulations, and to pulsed electrical measurements of dc conductivity of liquid carbon.

The dielectric function of most liquid metals is well described by a Drude function:[46]

$$\epsilon(\omega) = \epsilon_{1c} - \frac{\omega_p^2 \tau^2}{1+\omega^2\tau^2} + i\epsilon_{2c} + i\frac{\omega_p^2 \tau}{\omega(1+\omega^2\tau^2)} . \quad (4)$$

Success in fitting the optical properties of the related column IV liquid metals Si,[25,43] Ge,[47] and Sn,[47] with this model especially motivates its application to liquid carbon. Usually the polarizability (ϵ_{1c}) and residual interband absorption (ϵ_{2c}) of the ionic cores can be neglected ($\epsilon_{1c} = 1$, $\epsilon_{2c} = 0$), leaving the free electron plasma frequency $\omega_p^2 = 4\pi n_e e^2/m$ (n_e = electron density, m = free electron mass) and relaxation time τ as the dominant parameters which determine surface reflectivity. We shall follow this simplifying assumption initially, then later examine the effect of introducing finite core dielectric parameters. Full determination of the Drude parameters ω_p and τ requires at least two independent measurements; measurements at steep incidence angle usually provide the greatest precision in determining $\epsilon(\omega)$. Our measurements can be summarized in two categories: (1) the normal incidence reflectivity spectrum $R(h\nu)$ over the range $1.77 \le h\nu \le 4.0$ eV (see Figs. 6–8) and (2) oblique incidence ($\theta = 50°$) s- and p-polarized reflectivity at $h\nu = 2.0$ eV [e.g., Figs. 4(a) and 4(d)]. Measurements at $\Delta t \sim 1$ ps at each of several pump fluences are used for the analysis. Any choice of Drude parameters must be consistent with all measurements for a given pump fluence.

Figure 11 plots the normal incidence reflectivity spectrum at $\Delta t \sim 1$ ps for two fluences (large solid squares: 0.5 J/cm^2; open squares: 2.0 J/cm^2), as gleaned from the data in Figs. 6–8. This data is compared to the measured reflectivity spectra of solid graphite (solid curve), diamond (crosses), and liquid silicon (open circles). The measured reflectivity decreases monotonically with increasing photon energy $h\nu$, and equals or exceeds the reflectivity of the corresponding solids throughout the probed frequency range. At visible frequencies the measured reflectivity increases with pump fluence up to a maximum of about 0.5, although at $h\nu = 4.0$ eV, reflectivity was independent of fluence within experimental error. The first column of Fig. 12 plots the measured s- and p-polarized oblique incidence reflectivity at $h\nu = 2.0$ eV, $\Delta t \sim 1$ ps, at $F = 0.5$ J/cm^2, compared to the normal incidence reflectivity at the same probe wavelength.

The remaining curves in Figs. 11 and 12 are theoretical models of the reflectivity spectrum, which we now discuss

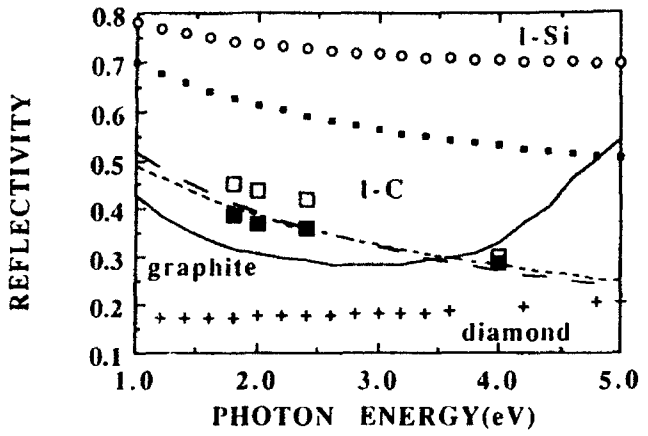


FIG. 11. Normal incidence reflectivity spectra of crystalline diamond (Ref. 27) (crosses); crystalline pyrolytic graphite (Ref. 26) (solid curve); carbon measured at $\Delta t \sim 1$ ps after a femtosecond melting of graphite with pulse of 0.5 J/cm^2 (large solid squares) or 2.0 J/cm^2 (large open squares); liquid carbon calculated using Drude parameters $\omega_p = 3.5\times10^{16}$ s^{-1}, $\tau = 0.25\times10^{-16}$ s (dotted curve) and $\omega_p = 1.6\times10^{16}$ s^{-1}, $\tau = 0.9\times10^{-16}$ s (dashed curve); liquid carbon calculated from molecular dynamics simulation at atmospheric pressure (Ref. 6) (small solid squares); liquid silicon measured by ellipsometric methods (Ref. 25) (open circles).

further. The small solid squares in Fig. 11 represent a reflectivity spectrum $R(\omega)$ extracted from the conductivity function $\sigma(\omega)$ derived by Galli *et al.*[6] from a molecular dynamics simulation of liquid carbon [see Fig. 2 of Ref. 6(a)] for the following conditions: low pressure $P \sim 1$ atm, $T_e = T_i = 5000$ K, density $\rho = 2.0$ g/cm^3. To obtain this $R(\omega)$, $\mathrm{Im}\sigma(\omega)$ was derived by Kramers-Kronig transformation of $\mathrm{Re}\sigma(\omega)$ given by Galli *et al.*;[6] then $\epsilon(\omega)$, expressed in terms of complex conductivity,

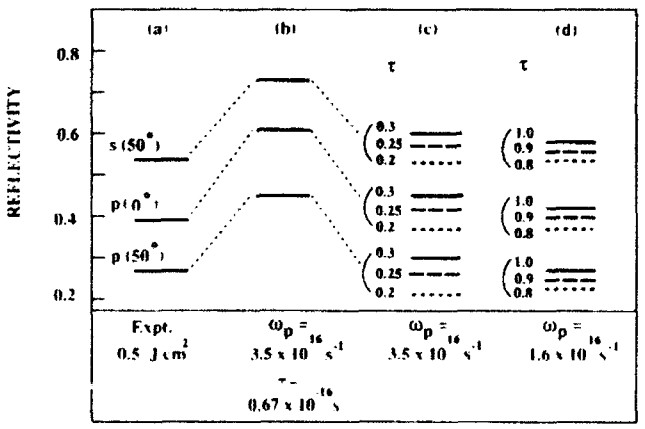


FIG. 12. Oblique ($\theta = 50°$) incidence, s- and p-polarized, and normal ($\theta = 0°$) incidence reflectivity of liquid carbon at photon energy 2 eV. (a) measured at $\Delta t \sim 1$ ps after a femtosecond melting of graphite with pulse of 0.5 J/cm^2; (b) calculated from molecular dynamics simulation at atmospheric pressure (Ref. 6); (c) calculated using Drude parameters $\omega_p = 3.5\times10^{16}$ s^{-1}, $\tau = 0.2, 0.25, 0.3\times10^{-16}$ s; (d) calculated using Drude parameters $\omega_p = 1.6\times10^{16}$ s^{-1}, $\tau = 0.8, 0.9, 1.0\times10^{-16}$ s.

was used to derive $R(\omega)$. Alternatively, a nearly identical $R(\omega)$ was obtained over the frequency range of our data by fitting $\mathrm{Re}\sigma(\omega)$ to a Drude function $\omega_p^2\tau/4\pi(1+\omega^2\tau^2)$, then substituting the resulting best fit Drude parameters $\omega_p = 3.5\times10^{16}$ s^{-1} and $\tau = 0.67\times10^{-16}$ s^{-1}, into the dielectric function Eq. (4) with $\epsilon_{1c} = 1$, $\epsilon_{2c} = 0$. Physically, the best fit ω_p corresponds to $n_e = 3.8\times10^{23}$ cm^{-3} (assuming a free electron mass), very close to four times the atomic density $n_a = 1\times10^{23}$ cm^{-3} corresponding to $\rho = 2.26$ g/cm^3, implying a nearly quadrivalent metal. The best fit τ implies a mean-free path $l = v_F\tau = 1.8$ Å (v_F = Fermi velocity $= 2.6\times10^8$ cm/s), close to the average interatomic spacing $n_a^{-1/3} = 2.1$ Å. The second column of Fig. 12 shows the oblique incidence reflectivity calculated from Fresnel's equations, using the same dielectric function.

The calculated reflectivity is approximately 50% higher than the observed reflectivity (Figs. 11 and 12), well outside of experimental inaccuracy. We therefore consider possible causes of this significant discrepancy. One possibility is that core dielectric terms must be included along with the best fit Drude parameters. However, introducing a core polarizability as large as that of solid graphite ($\epsilon_{1c} = 4.5$) together with the best-fit Drude parameters cited above yields $R(2\text{ eV}) = 0.58$ and $R(4\text{ eV}) = 0.45$, still well above the observed values. We have investigated systematically combinations of core dielectric terms in the ranges $1 \le \epsilon_{1c} \le 4.5$ and $0 \le \epsilon_{2c} \le 4.0$, and find generally that core dielectric terms cannot explain the discrepancy shown. We must conclude that the Drude parameters themselves are different from the values extracted from the simulations of Galli *et al.*,[6] or that the Drude function itself does not adequately describe the properties of carbon 1 ps after photoexcitation. Close agreement should not necessarily be expected, because our experiment probes a sample at high pressure, and densities close to those of the initial solids (graphite: 2.26 g/cm^3; diamond: 3.5 g/cm^3). The discrepancy provides motivation for further simulations of the electronic properties of liquid carbon under conditions closer to those of the experiment. In fact, a molecular dynamics simulation at high pressure has been reported,[6(b)] but to our knowledge electrical properties were not derived. In the current absence of such simulations, we consider what revisions in the dielectric function can better explain the reflectivity data.

The dotted curve in Fig. 11 shows a greatly improved fit to $R(\omega)$ at $F = 0.5$ J/cm^2 obtained by reducing the value of τ from 0.67 to 0.25×10^{-16} s, while retaining other parameters ($\omega_p = 3.5\times10^{16}$ s^{-1}, $\epsilon_{1c} = 1$, $\epsilon_{2c} = 0$) corresponding to a simple nearly quadrivalent Drude metal. The third column of Fig. 12 shows that the oblique incidence reflectivity at $F = 0.5$ J/cm^2 is also explained almost perfectly by this dielectric function, and that slightly larger (0.3×10^{-16} s) or smaller (0.2×10^{-16} s) values of τ significantly worsen the fit. Nevertheless this fit is not unique. The dashed curve in Fig. 11 and the fourth column of Fig. 12 show that a fit of nearly equal quality is obtained by decreasing ω_p to 1.6×10^{16} s^{-1}, corresponding to a monovalent ($n_e = 1\times10^{23}$ cm^{-3}) metal, and increasing τ to 0.9×10^{-16} s. In fact a family of

(ω_p,τ) pairs in between these two extremes, plotted as a line in the (ω_p,τ) plane in Fig. 13(a), provide satisfactory fits to our reflectance data for $F=0.5$ and 2 J/cm^2. Fits of similar quality are obtained for the data at other fluences. Thus the individual Drude parameters are not uniquely determined by our current data. On the other hand, the line in Fig. 13(a) determines the *product* $\omega_p\tau$ and the *dc resistivity* $\rho_{dc}=4\pi/\omega_p^2\tau$ to within relatively narrow ranges, as shown in Fig. 13(b) (e.g., $1<\omega_p\tau<1.6$ and $400<\rho_{dc}<500$ $\mu\Omega$ cm for 0.5 J/cm^2). Finally Fig. 13(c) shows the dependence of the optically determined

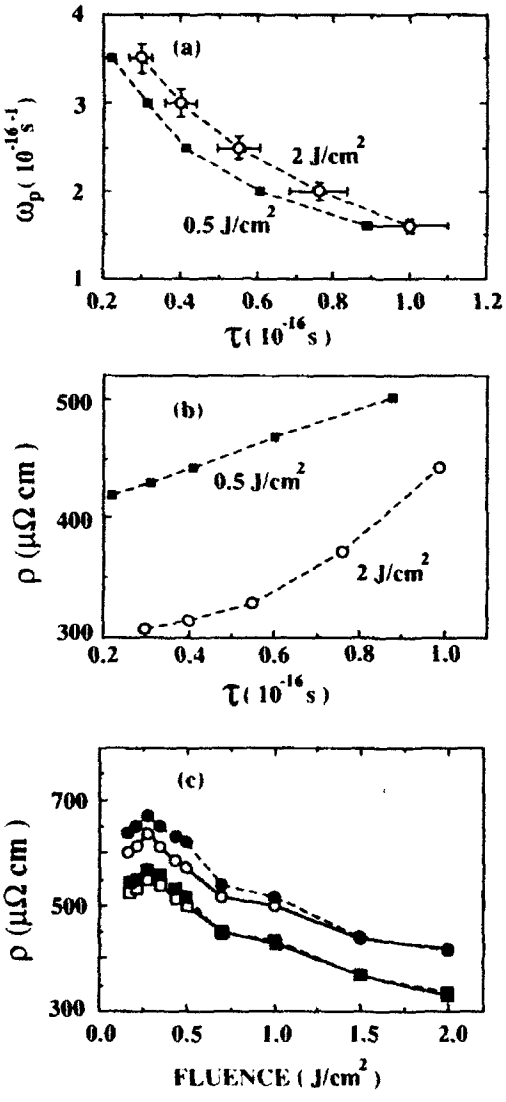


FIG. 13. (a) Family of Drude parameters ω_p and τ which provide satisfactory fits to carbon reflectance data at $\Delta t\sim 1$ ps after femtosecond melting pulse at 0.5 and 2.0 J/cm^2; (b) dc resistivity values $4\pi/\omega_p^2\tau$ corresponding to family of Drude parameters in (a); (c) dependence of optically determined resistivity on pump fluence, showing four different data analyses assuming ω_p constant at either 1.6×10^{16} s^{-1} (open circles) or 3.5×10^{16} s^{-1} (open squares) and τ variable with increasing fluence; or, conversely, τ constant at either 0.3×10^{-16} s (solid squares) or 1.0×10^{-16} s (solid circles), and ω_p variable with increasing fluence.

ρ_{dc} on pump fluence. Because ω_p and τ may individually vary with fluence in a range of possible ways, four plots of ρ_{dc} illustrating the extreme cases are shown in Fig. 13(c): ω_p constant (at either 1.6 or 3.5×10^{16} s^{-1}) and τ variable with increasing fluence; or, conversely, τ constant (at either 0.3 or 1.0×10^{-16} s), and ω_p variable with increasing fluence. In all cases, ρ_{dc} decreases from 625 ± 75 $\mu\Omega$ cm slightly above the melting fluence to 375 ± 75 $\mu\Omega$ cm at tenfold higher fluence. For $0.13<F<0.3$ J/cm^2, the reflectance data was first corrected for the incomplete overlap of the melted area with the probe spot, using the data in Fig. 1. The slight increase in ρ for $F<0.25$ J/cm^2 falls within experimental error. The trend in ρ contrasts with the usual increase of ρ with temperature in metals, but follows the usual decrease of ρ with pressure, as discussed further below.

Table I compares our optically derived Drude parameters and dc resistivity with optical-derived parameters of the other column IV liquid metals (measured at atmospheric pressure slightly above their respective melting points), and with pulsed electrical measurements and theoretical simulations for liquid carbon. Our product $\omega_p\tau\sim1$ is significantly smaller than for the other column IV liquid metals, indicating an unusually strongly damped electron plasma and strong electron localization. Similarly, the optically determined ρ_{dc} is significantly higher than those of the other metals. The *l*-C values are consistent qualitatively with the trends of decreasing $\omega_p\tau$ and increasing ρ_{dc} as atomic size decreases in column IV. The unusually large size of the changes between *l*-Si and *l*-C is consistent with the much higher temperature required to melt carbon, which in turn strongly reduces the electron collision time τ. Nevertheless, the unusually small numerical values of τ ($0.3<\tau<1.0\times10^{-16}$ s) for *l*-C compared to optically determined τ values in *l*-Si (2.1×10^{-16} s), *l*-Ge (2.5×10^{-16} s), and *l*-Sn (3.5×10^{-16} s) warrant further comment, since they imply electron mean free paths $l\sim v_F\tau$ in the range $0.8<l<1.7$ Å, which are actually somewhat *smaller* than the average interatomic spacing ($N^{-1/3}\sim2$ Å). $N^{-1/3}$ should set an approximate lower limit on the path between electron scattering events, leading to "saturation" of ρ_{dc} and τ with increasing temperature,[48] as has been observed in aluminum under short pulse heating conditions.[49] A possible physical mechanism yielding l slightly less than $N^{-1/3}$ might be collisions within localized carbon molecules containing small angle bonds. Alternatively, a residual population of anomalously hot electrons with $v>v_F$ would have mean-free paths larger than the range indicated above. On the other hand, in view of the phenomenological nature of the Drude model, the precise value of τ should probably not be interpreted too literally, as long as $v_F\tau$ saturates in the vicinity of $N^{-1/3}$. Indeed the fluence dependence shown in Fig. 13(c) is fully consistent with a saturated resistivity above the melting fluence, since resistivity does not increase as fluence increases above F_m (with the possible exception of a slight increase in a narrow range just above F_m). In fact, rather than leveling off, resistivity actually *decreases* with approximately $F^{-0.6}$ dependence. The decrease in ρ prob-

TABLE I. Summary of measurements and calculations of Drude parameters and dc resistivity in column IV liquid metals.

Liquid	Pressure (bars)	Temperature (K)	$\omega_p\tau$	ρ_{dc} ($\mu\Omega$ cm)	Method	Reference
Sn	1	1130	8	60	cw ellipsometry	Hodgson (Ref. 47)
Ge	1	1270	6	75	cw ellipsometry	Hodgson (Ref. 47)
Si	1	1800	5.6	80	cw ellipsometry	Shvarev *et al.* (Ref. 43)
	1	1800	5.3	85	Picosecond reflectance	Li and Fauchet (Ref. 25)
C	$\sim 0.3\times 10^6$	~ 6000	1–1.6	625 ± 75	Femtosecond reflectance	This work
	4×10^3	~ 5000		1000	Ohmic heating	Gathers *et al.* [Ref. 14(a)]
	1	~ 4500		30–70	Ohmic heating	Heremans *et al.* (Ref. 4)
	1	5000	2.3	140 ± 28	Molecular dynamics simulation	Galli *et al.* [Ref. 6(a)]

ably has multiple causes, including increasing ionization (i.e., increasing ω_p, even at constant τ), and an approach to high temperature plasma behavior, in which resistivity is increasingly determined by Coulomb scattering,[50] and scales as $T^{-3/2}$.

The range of Drude parameters cited above corresponds to a range of possible electronic and molecular structures. The smaller ω_p (1.6×10^{16} s^{-1}) corresponds to a monovalent metal, implying that three of the four valence electrons of carbon are involved in covalent bonding. Indeed the molecular dynamics simulation of Galli *et al.*,[6] although consistent with a quadrivalent electronic structure at low pressure, nevertheless shows prevalent clusters and chainlike molecular structures in the liquid state. On the other hand, the other column IV liquid metals are all quadrivalent metals.[25,47] The current results provide no firm basis for distinguishing between these two pictures, or an intermediate version. Molecular dynamics simulations of the electronic properties at elevated pressure, together with extended femtosecond ellipsometric measurements at very steep probe incidence angles and vacuum ultraviolet probe frequencies will narrow the uncertainty considerably, and possibly detect variations from Drude-like optical properties. Whichever of these interpretations proves ultimately correct, the current experiments show clearly that liquid carbon is a resistively saturated material, for which the "free electron" Drude model and the attribution "metal" are at the limits of their validity.

Comparison of our optically determined resistivity with pulsed electrical heating measurements of liquid carbon (see Table I) also provides physical insight. The measurements which most closely approximate our pressure-temperature conditions involve pulsed ohmic heating of vitreous[14(c)] or glassy carbon[14(d)] to temperatures up to 6000 K under 2–4 kbar rare gas pressure. Melt resistivities $\rho\sim 1000$ $\mu\Omega$ cm were reported from these measurements, which is very consistent with our optically determined values just above F_m. In fact the somewhat smaller values obtained in our experiment are fully consistent with the higher sample compression under femtosecond excitation, which suppresses the amplitude of ionic vibrational motion responsible for electron scattering and resistivity. Indeed clear evidence of volume expansion was obtained in the pulsed ohmic heating measurements,[14(c)] and was taken into account in determining the resistivity. On the other hand, recent measurements[4] and calculations[6,7] at ambient atmospheric pressure have yielded much smaller resistivities than obtained in our experiment. Heremans *et al.*[4] reported resistivity $30<\rho<70$ $\mu\Omega$ cm based on resistance measurements of ohmically heated pyrolytic graphite fibers at atmospheric pressure. Two theoretical simulations performed after this experiment yielded resistivities of 140 $\mu\Omega$ cm,[6] and 80 $\mu\Omega$ cm[7]. Volume expansion upon melting was neglected in analyzing the measured resistance, which, if present to the degree observed in Ref. 14(c), could raise the reported resistivity by as much as a factor of 2. Nevertheless, even with this correction the values would still fall significantly below ours. The lower sample pressure compared to our conditions is expected to yield higher, rather than lower resistivity (in the absence of phase transitions). Thus other causes of the discrepancy with our measurements must be sought. One possibility is that the nature of the material is very different at atmospheric pressure. Indeed, a number of independent experiments in which graphite was melted in atmospheres of pressurized inert gases have agreed that the solid-vapor-liquid triple point occurs at ~ 110 atm. pressure.[3(c),51] Thus according to these investigators, an equilibrated liquid phase of carbon does not even exist at atmospheric pressure. It is therefore conceivable that atmospheric pressure experiments measure a qualitatively different phase of carbon than high pressure experiments. Reconciliation of the various pulsed ohmic heating experiments is complicated by their use of different starting forms of carbon, and by the lack (to our knowledge) of measurements on a single type of starting sample at pressures both above and below 110 atm. In addition, the reported electrical resistivity of the heated phase sometimes depends on prior history (e.g., heat treatment) of the starting solid. These ambiguities would be clarified by performing ohmic heating experiments on well-characterized pyrolytic graphite fibers, glassy carbon, vitreous carbon, and perhaps other solid carbon forms at systematically varied pressures ranging from 1 bar to several kbar. In summary, our optically determined resistivity agrees very well with pulsed electrical measurements reported for carbon samples melted at pressures well above the widely accepted triple point pressure of

110 atm,[14] but significantly exceeds reported measurements[4] and calculations[6,7] of resistivity for low pressure samples.

VI. CONCLUSIONS

Femtosecond melting experiments measure the optical and electronic properties of refractory condensed materials in quantitative detail at extremes of pressure and temperature inaccessible by other experimental methods. We have presented a comprehensive study of the optical properties of liquid carbon produced by femtosecond excitation of pyrolytic graphite and diamond above critical melting fluences of 0.13 J/cm^2 and 0.6 J/cm^2, respectively. All evidence suggests that both samples melt, reach nearly common electron-lattice temperature and steady state visible reflectance higher than the starting solids within 1 ps, before the release of internal pressure and hydrodynamic expansion distort the optical properties of the surface. A careful analysis of fluence-dependent reflectance at $\Delta t > 1$ ps and fluence-dependent damage morphology shows that surface expansion strongly attenuates surface reflectivity within a few picoseconds of femtosecond irradiation at all fluences above the critical melting fluence. A Drude model fit of the measured reflectance spectrum at $\Delta t \sim 1$ ps shows liquid carbon to be a resistively saturated metal under femtosecond melting conditions, consistent with dc resistivity measurements made on carbon samples melted by pulsed ohmic heating in an inert, pressurized atmosphere.

This work points to a number of areas for additional research. The pressure dependence of the electrical properties of liquid carbon should be clarified by theoretical simulations at high pressure and by pulsed electrical measurements on well-characterized samples at systematically varied pressures between 1 bar and several kbar. Femtosecond experiments can be improved by using probe pulses at vacuum ultraviolet wavelengths and steep incidence angles, in order to probe the plasma edge more accurately and determine individual Drude parameters more precisely than has been possible in this work. Significant deviations from Drude-like behavior might also be detected. The degree of equilibration following femtosecond excitation can be further clarified by dynamical simulations which start with graphite and/or diamond crystalline structures, and by more extensive comparison of the femtosecond time-resolved reflectance of less refractory materials with independently measured optical properties of the equilibrated liquid phases. Finally, in a larger context, femtosecond spectroscopy provides a promising method for optically characterizing a wide range of condensed, refractory materials in previously inaccessible pressure and temperature regimes characteristic of planetary and stellar interiors.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (Grant No. DMR-8858388), the Robert A. Welch Foundation (Grant F-1038), and the Air Force Office of Scientific Research (Contract F49620-89-C-0044). We acknowledge helpful discussions with J. C. Thompson and N. Bloembergen.

*Present address: Bellcore, Red Bank, NJ 07701.

[1]A. Ludwig, Z. Electrochem. **8**, 273 (1902).

[2](a) F. P. Bundy, Physica (Amsterdam) **156A**, 169 (1989); (b) J. Geophys. Res. **85**, 6930 (1980).

[3](a) A. G. Whittaker, P. L. Kitner, L. S. Nelson, and N. Richardson (unpublished); (b) M. A. Scheindlin, Mater. Res. Soc. Symp. Proc. **22**, 33 (1984); (c) N. A. Gokchen, E. T. Chang, T. M. Poston, and D. J. Spencer, High Temp. Sci. **8**, 81 (1976).

[4]J. Heremans, C. H. Olk, G. L. Eesley, J. Steinbeck, and G. Dresselhaus, Phys. Rev. Lett. **60**, 452 (1988).

[5]J. Steinbeck, G. Braunstein, M. S. Dresselhaus, T. Venkatesan, and D. C. Jacobson, J. Appl. Phys. **58**, 4374 (1985); **64**, 1802 (1988).

[6](a) G. Galli, R. M. Martin, R. Car, and M. Parrinello, Phys. Rev. Lett. **63**, 988 (1989); (b) Science **250**, 1547 (1990); (c) Phys. Rev. B **42**, 7470 (1990).

[7]M. W. C. Dharma-wardana and Francois Perrot, Phys. Rev. Lett. **65**, 76 (1990).

[8]A. M. Malvezzi, N. Bloembergen, and C. Y. Huang, Phys. Rev. Lett. **57**, 146 (1986).

[9]E. A. Chauchard, C. E. Lee, and C. Y. Huang, Appl. Phys. Lett. **50**, 812 (1987).

[10]D. H. Reitze, X. Wang, H. Ahn, and M. C. Downer, Phys. Rev. B **40**, 11 986 (1989).

[11](a) D. H. Reitze, H. Ahn, X. Wang, and M. C. Downer, in *Ultrafast Phenomena VII*, edited by C. B. Harris, E. P. Ippen, G. A. Mourou, and A. Zewail (Springer-Verlag, Berlin, 1990); (b) D. H. Reitze, Ph.D. dissertation, University of Texas at Austin, 1990.

[12]T. Venkatesan, D. C. Jacobson, J. M. Gibson, B. S. Elman, G. Braunstein, M. S. Dresselhaus, and G. Dresselhaus, Phys. Rev. Lett. **53**, 360 (1984).

[13]J. S. Speck, J. Steinbeck, and M. S. Dresselhaus, J. Mater. Res. **5**, 980 (1990).

[14](a) A. C. Mitchell, J. W. Shaner, and R. N. Keller, Physica **139**, 386 (1986); (b) G. R. Gathers, J. W. Shaner, and R. L. Brier, Rev. Sci. Instrum. **47**, 471 (1976); (c) G. R. Gathers, J. W. Shaner, and D. A. Young (unpublished); (d) J. W. Shaner, Bull. Am. Phys. Soc. **32**, 607 (1987).

[15](a) J. W. Shaner, J. M. Brown, C. A. Swenson, and R. G. McQueen, J. Phys. C **8**, 235 (1984); (b) J. Kleinman, R. B. Heimann, D. Hawken, and N. M. Salansky, J. Appl. Phys. **56**, 1440 (1984).

[16]M. Ross, Nature (London) **292**, 435 (1981).

[17]D. A. Young and R. Groover, in *Shock Waves in Condensed Matter*, edited by S. C. Schmidt and M. C. Holmes (Elsevier, New York, 1988), p. 131.

[18]M. S. Weathers and W. A. Bassett, Phys. Chem. Miner. **15**, 105 (1987).

[19]J. S. Dickey, W. A. Bassett, J. M. Bird, and M. S. Weathers, Geology **11**, 219 (1983).

[20]A. Ferraz and N. H. March, Phys. Chem. Liq. **8**, 289 (1979).

[21]M. C. Downer, R. L. Fork, and C. V. Shank, J. Opt. Soc. Am. B **2**, 595 (1985).

[22]H. W. K. Tom, G. D. Aumiller, and C. H. Brito-Cruz, Phys. Rev. Lett. **60**, 1438 (1988).

[23]K. Seibert, G. C. Cho, W. Kütt, H. Kurz, D. H. Reitze, J. I

Dadap, H. Ahn, M. C. Downer, and A. M. Malvezzi, Phys. Rev. B **42**, 2842 (1990).

[24]N. H. March, *Liquid Metals* (Cambridge University Press, Cambridge, England, 1990).

[25]K. D. Li and P. Fauchet, Appl. Phys. Lett. **51**, 1747 (1987).

[26](a) B. T. Kelly, *The Physics of Graphite* (Applied Science, London, 1980); (b) L. G. Johnson and G. Dresselhaus, Phys. Rev. B **7**, 2275 (1973); (c) E. A. Taft and H. R. Philipp, Phys. Rev. **138**, 197 (1965).

[27](a) G. S. Painter, D. E. Ellis, and A. R. Lubinsky, Phys. Rev. B **4**, 3610 (1971); (b) H. R. Philipp and E. A. Taft, Phys. Rev. **136**, A1445 (1964); (c) R. A. Roberts and W. C. Walker, *ibid.* **161**, 730 (1967).

[28]C. V. Shank, R. Yen, and C. V. Hirlimann, Phys. Rev. Lett. **50**, 454 (1983).

[29]R. L. Fork, B. I. Greene, and C. V. Shank, Appl. Phys. Lett. **38**, 671 (1981).

[30]W. M. Wood, G. Focht, and M. C. Downer, Opt. Lett. **13**, 984 (1988).

[31](a) M. F. Becker, R. M. Walser, Y. K. Jhee, and D. Y. Sheng, SPIE **322**, 93 (1982); (b) L. W. Boyd, S. C. Moss, T. F. Boggess, and A. L. Smirl, Appl. Phys. Lett. **45**, 80 (1984); (c) A. M. Malvezzi, H. Kurz, N. Bloembergen, in *Energy Beam-Solid Interactions and Transient Thermal Processing*, edited by D. K. Biegelsen, G. A. Rozgonyi, and C. V. Shank (Materials Research Society, Pittsburgh, 1985).

[32]R. L. Fork, C. V. Shank, C. Hirlimann, R. Yen, and W. J. Tomlinson, Opt. Lett. **8**, 1 (1983).

[33]J. M. Liu, Opt. Lett. **7**, 196 (1980).

[34]J. I. Dadap, D. H. Reitze, G. B. Focht, and M. C. Downer, Opt. Lett. **16**, 499 (1991).

[35]R. Fedosejevs, R. Ottmann, R. Sigel, G. Kühnle, S. Szatmari, and F. P. Schäfer, Phys. Rev. Lett. **64**, 1250 (1990).

[36](a) H. M. Milchberg and R. R. Freeman, J. Opt. Soc. Am. B **6**, 1351 (1989); (b) H. M. Milchberg, R. R. Freeman, S. C. Davey, and R. M. More, Phys. Rev. Lett. **61**, 2364 (1988).

[37]N. Bloembergen, in *Beam-Solid Interactions and Phase Transformations*, edited by H. Kurz, G. L. Olsen, and J. M. Poate (Materials Research Society, Pittsburgh, 1986), p. 3.

[38]Y. B. Zeldovich and Y. P. Raizer, *Physics of Shock Waves and High Temperature Hydrodynamic Phenomena* (Academic, New York, 1966).

[39]R. M. More, Adv. At. Mol. Phys. **21**, 305 (1985).

[40]S. D. Brorson, A. Kazeroonian, J. S. Moodera, D. W. Face, T. K. Cheng, E. P. Ippen, M. S. Dresselhaus, and G. Dresselhaus, Phys. Rev. Lett. **64**, 2172 (1990).

[41]F. E. Doany and D. Grischkowsky, Appl. Phys. Lett. **52**, 36 (1988).

[42]P. M. Fauchet and D. Hulin, J. Opt. Soc. Am. B **6**, 1024 (1989).

[43]K. M. Shvarev, B. A. Baum, and P. V. Gel'd, Fiz. Tverd. Tela (Leningrad) **16**, 3246 (1974) [Sov. Phys. Solid State **16**, 2111 (1975)].

[44]R. M. More, K. H. Warren, D. A. Young, and G. B. Zimmerman, Phys. Fluids **31**, 3059 (1988).

[45]J. E. Rothenberg, Opt. Lett. **13**, 713 (1988).

[46]T. E. Faber, *Introduction to the Theory of Liquid Metals* (Cambridge University Press, Cambridge, England, 1972).

[47](a) J. N. Hodgson, Philos. Mag. **6**, 509 (1961); (b) N. R. Cohins, *ibid.* **25**, 817 (1972).

[48]N. F. Mott, *Metal-Insulator Transitions* (Taylor and Francis, London, 1974).

[49]H. M. Milchberg, R. R. Freeman, S. C. Davey, and R. M. More, Phys. Rev. Lett. **61**, 2364 (1988).

[50]L. Spitzer, Jr., *Physics of Fully Ionized Gases* (Wiley, New York, 1962).

[51](a) G. J. Schoessow, Phys. Rev. Lett. **21**, 738 (1968); (b) N. S. Fateeva and L. F. Vereshchagin, Pis'ma Zh. Eksp. Teor. Fiz. **13**, 157 (1971) [JETP Lett. **13**, 110 (1971)]; (c) D. M. Haaland (unpublished).

Femtosecond time-resolved reflectivity of hydrodynamically expanding metal surfaces

X. Y. Wang and M. C. Downer

Department of Physics, University of Texas at Austin, Austin, Texas 78712

Received June 30, 1992

We present femtosecond time-resolved *p*- and *s*-polarized reflectivity measurements of silver, aluminum, and tungsten surfaces that are hydrodynamically expanding in response to pulsed laser excitation. A numerical model of the experimental results shows that (1) electron heating and surface expansion influence the reflectivity on distinguishable time scales and (2) covalency, described by non-Drude dielectric constants, is necessary to explain the time-resolved reflectivity of the tungsten target.

Intense femtosecond excitation of solid targets has been used to produce and measure the properties of solid density plasmas[1,2] and pressurized liquids[3] that resemble states of matter found in planetary and stellar interiors. With sufficiency short pulses, the surface can be fully heated and optically probed before the surface hydrodynamically expands. In several recent studies,[1,2] the ultrafast reflectivity response of highly excited metals was observed with a single, intense laser pulse serving both as target excitation and as optical probe. In these studies, the initial stages of hydrodynamic expansion occurred during the pulse. Consequently, competing processes of electron heating, melting, and hydrodynamic surface expansion, which influence the reflectivity to different degrees and at different times, could not be clearly resolved.

In this Letter we present femtosecond time-resolved experiments in which a 90-fs visible pump pulse excites either a free-electron metal (Ag, Al) or *d*-band metal (W) target at intensities of as much as 5×10^{13} W/cm^2. The front surface reflectivity $R_{p,s}(\Delta t)$ of a 90-fs, obliquely incident *p*- or *s*-polarized probe pulse of the same center wavelength is then monitored as the pump–probe time delay Δt is varied from -1 to $+20$ ps. The probe $R_{p,s}(\Delta t)$ data are analyzed at each Δt by numerically solving the Helmholtz wave equations.[4] The results show that (1) electron heating, which occurs instantaneously on surface excitation, can be temporally distinguished from the delayed surface hydrodynamic expansion and (2) non-Drude background dielectric constants are needed to explain $R_{p,s}(\Delta t)$ of the expanding non-free-electron metal W target, which indicates a strong influence of covalent bonding in the liquid and the expansion region. Through such results we aim to understand quantitatively the optical properties of metal surfaces during the initial stage of plasma expansion following short-pulse excitation, which in turn may aid in testing hydrodynamic codes used to model inertial confinement fusion and short-pulse x-ray generation.

The experiments were performed in air immediately after the samples were polished, and the initial optical properties were determined with a He–Ne laser. A colliding-pulse mode-locked dye laser and a four-stage Nd:YAG-pumped dye amplifier generated 0.5-mJ, 620-nm, 90-fs pulses at a 10-Hz repetition rate. The amplified pulses are split into a pump beam, a 1000-fold weaker probe beam, and a reference beam, which is used for normalizing the effect of pulse energy fluctuations. The probe is incident at 70° from the surface normal and is focused to a 30 μm $\times$ 60 μm spot on the sample. This steep angle provides good sensitivity to changes in the surface dielectric function and density gradient. The *p*-polarized pump is incident at 45° and focused to a 60 μm $\times$ 80 μm spot centered around the probe spot. The ~5-ns pedestal of amplified spontaneous emission was measured to be less than 4% of the pump-pulse energy, well below the threshold for preplasma formation. The sample was translated after each shot to ensure that each pump pulse excited an undamaged area. Each data point was averaged over 20 shots.

$R_{p,s}(\Delta t)$ data for the free-electron metals Ag and Al excited at 10^{13} W/cm^2 are shown in the main plot and the inset of Fig. 1(a), respectively. $R_p(\Delta t)$ decreases with the pulse-width-limited fall time from an initial value of 0.88 (Ag) or 0.79 (Al) to a final value of ~0.35 after excitation. After reaching a minimum, $R_p(\Delta t)$ partially and more slowly recovers to approximately 60% of the original value within 15 ps. For $\Delta t > 15$ ps, $R_p(\Delta t)$ levels off or, with higher pump intensity, begins to decrease. $R_s(\Delta t)$ of identically excited Ag and Al, on the other hand, changes only slightly during the interval $0 < \Delta t < 15$ ps. Moreover its near-unity value shows that nonspecular scattering is weak during this time interval. For pump intensities of $>10^{13}$ W/cm^2, $R_s(\Delta t)$ decreases gradually for $\Delta t > 10$ ps, which suggests the onset of nonspecular scattering as plasma nonuniformities develop, but still only slightly for $\Delta t < 10$ ps.

$R_{p,s}(\Delta t)$ data for the *d*-band metal W identically excited and probed are shown in Fig. 1(b). The inset shows $R_p(\Delta t)$ for excitation at 5×10^{12} W/cm^2. In this case, solid W simply melts in <1 ps, without ablating, into a liquid that is more reflective than the

0146-9592/92/201450-03$5.00/0 © 1992 Optical Society of America

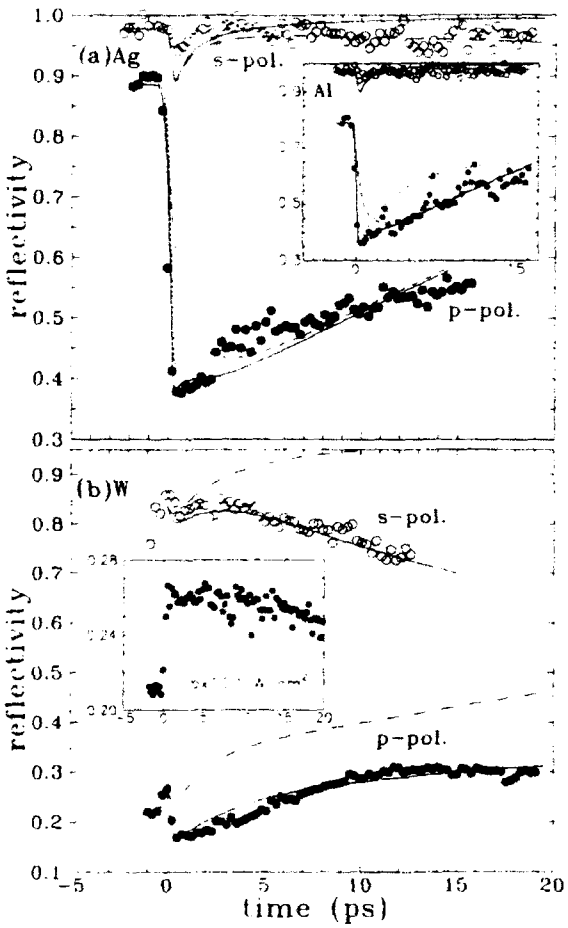


Fig. 1. Time-resolved reflectivity of (a) Ag and Al (inset) and (b) W surfaces pumped by *p*-polarized, 90-fs, 620-nm pulses incident at 45° from the normal at 10^{13} W/cm^2 and probed by *s*-polarized (open circles) or *p*-polarized (filled circles) pulses (90 fs, 620 nm) incident at 70°. Curves represent theoretical models based on numerical solutions of Helmholtz equations, as follows: (a) Dotted curves, electron heating included, surface expansion neglected; dashed curves, electron heating neglected, surface expansion included; solid curves, electron heating and surface expansion both included; long-dashed curves, covalent model (*s* polarized). (b) Dashed curves, free-electron model; solid curves, covalent model; inset: *p*-polarized data at 5×10^{12} W/cm^2.

solid, so R_p remains nearly constant over 20 ps. When the excitation was increased to 10^{13} W/cm^2 [Fig. 1(b), main plot], however, an initial R_p increase to the liquid value (~0.26) is followed by a drop and recovery, as observed in Al and Ag. R_s also increases by approximately 10% immediately on excitation then decays slightly over the 15-ps period of observation.

We now analyze $R_{p,s}(\Delta t)$ for Ag, Al, and W by numerically solving the Helmholtz equations[4] at each Δt using a model that includes (1) heating and subsequent cooling of the electrons and (2) development of a plasma gradient that expands at velocity $c_{exp} \sim c_{sound}$ normal to the metal surface. A space- and time-dependent Drude dielectric function $\epsilon(x) = \epsilon_{10}(x) + i\epsilon_{20}(x) - \omega_p^2(x)/\omega[\omega + i\nu(x)]$ describes the surface expansion region at each time Δt after excitation, where ν is the electron collision frequency, $\omega_p = (n_e e^2/\epsilon_0 m)^{1/2}$ is the plasma frequency, $-\omega_p^2/\omega(\omega + i\nu)$ is the free-electron contribution to the dielectric function that is due to intraband transitions, while $\epsilon_{10} - 1$ and ϵ_{20} describe the covalent electron contribution from interband transitions.

The optical properties of liquid Ag and Al are well explained by a free-electron Drude dielectric function at visible frequencies.[5] Thus, as a first approximation, we set $\epsilon_{10} = 1$ and $\epsilon_{20} = 0$, so that the x dependence of $\epsilon(x)$ in the expansion region is contained entirely in the x dependence of the atomic density $N(x)$, as $\omega_p^2(x)$ and $\nu(x)$ are linearly proportional to $N(x)$. The dotted curves in Fig. 1(a) were calculated based on electron heating and cooling only, without hydrodynamic surface expansion, as follows. First, the time evolution of electron temperature $T_e(t)$ was calculated *a priori* by solving the coupled diffusion equations.[6] As a lower limit, we used room-temperature electron–phonon coupling constants of 3.6×10^{17} W/m^{3}K for Al and 3.6×10^{16} W/m^{3}K for Ag.[7] The calculated peak T_e is insensitive to a threefold increase of the coupling constant. In solving the equation, a Gaussian laser temporal profile provided the heat generation term that determines the rise time of T_e, and the T_e dependence of the electron heat capacity and chemical potential was included[8] to ensure accurate description of the transition from degenerate to Maxwellian electrons. Room-temperature thermal conductivities were used. Ballistic heat transport was neglected here because of the much shorter electron mean free path at high T_e. These calculations yielded peak electron temperatures of 10, 14.3, and 9.41 eV for Al, Ag and W, respectively, at 10^{13} W/cm^2. Second, $T_e(t)$ was related to $\epsilon(t)$, and therefore to $R_{p,s}(\Delta t)$, by using[9] $\nu(t) = Cn_i Z^2 T_e(t)/\omega_p^2$ (n_i is the ion density, Z is the effective ionization number), which is valid for $kT_e < 20$ eV. The parameter C involves the structure factor and other weak temperature-dependent factors. However, C can be regarded as a temperature-independent constant and was the sole parameter used to fit the R_p minima for Al and Ag at several excitation levels. With ω_p/ω equal to the liquid-state values [4.52 (Ag) and 7.31 (Al)][5] on melting and $Z = 3$ (Al) or 1 (Ag), the dotted curves in Fig. 1 are generated, which correspond to maximum collision frequencies of $\nu_{Al}/\omega = 1.88$ and $\nu_{Ag}/\omega = 0.87$ for 10^{13} W/cm^2 excitation. The observed R_p minima at several other excitation levels of Al and Ag were also consistent with this electron heating analysis to within 10%. Finally, the fitted value $\nu_{Al} = 5.7 \times 10^{15}$ s^{-1} is close to the corresponding value $\nu = 6.1 \pm 0.4 \times 10^{15}$ s^{-1} derived from Ref. 1 for Al at $T_e = 10$ eV. These independent checks all confirm that the rapid drop in R_p and the R_p minima are determined almost entirely by electron heating. However, the calculated electron cooling shown by the dotted curves fails to explain the slow observed recovery of R_p without invoking unreasonably small phonon coupling constants.

On the other hand, the recovery of R_p can be explained well by introducing hydrodynamic surface expansion at approximately c_{sound}. The dashed curves in Fig. 1(a) show the calculated $R_p(\Delta t)$ values

of Al and Ag by considering expansion alone, without electron heating. This calculation assumed that the atomic density gradient function in the expansion region follows the Riemann solution of the hydrodynamic equations[10]: $N(x) = N_0(3/4 - x/4v_{exp}t)^3$ for $-v_{exp}t < x < 3v_{exp}t$. Expansion velocities $v_{ex} = 1.2 \times 10^6$ cm/s (Al) and 0.8×10^6 cm/s (Ag), assumed constant over the initial 15 ps of expansion, were then determined from the best fit to the observed recovery and agree well with the acoustic velocities independently calculated[2] with the peak electron temperatures cited above, $v_0 = (Zk_BT_e/m_i)^{1/2} = 1.16$ and 0.36×10^6 cm/s, respectively. Nevertheless, this model alone cannot explain the pulse-width-limited fall time of R_p. However, when both increased collisionality and plasma expansion are taken into account with the same calculated $\nu(t)/\omega$ values and fitted v_{exp} parameter cited above for the dotted and dashed curves, a satisfying fit to $R_p(t)$ of Ag and Al is obtained for the entire interval $0 < t < 15$ ps, as shown by the solid curves in Fig. 1(a). The final curves were insensitive to the recovery dynamics of $T_e(t)$. A small drop in $R_s(t)$ near $t = 0$ is also expected, but with magnitude close to the experimental noise level.

In the case of W, if we make the free-electron assumption (i.e., $\epsilon_{10} = 1$, $\epsilon_{20} = 0$) for the liquid and expanding material, we find that the postmelting reflectivity drop and recovery shown in Fig. 1(b) cannot be explained by a surface expansion model like that described above. For example, the dashed curves in Fig. 1(b) show our best electron heating/surface expansion model under the constraints $\epsilon_{10} = 1$ and $\epsilon_{20} = 0$ but disagree strongly with both the $R_p(t)$ and $R_s(t)$ data. For these curves $\nu(0)/\omega = 3.65$ and $\omega_p/\omega = 9.07$ were used in order to satisfy the initial reflectivity values of molten W. However, no better fit could be obtained despite wide variations in the choice of parameters. Under our conditions, recombination is negligible[11] in the expansion region over the initial 15 ps.

To obtain a satisfactory fit of the W data, covalent background dielectric constants $\epsilon_{10} > 1$ and $\epsilon_{20} > 0$ must be introduced in both the liquid and expansion regions. The solid curve in Fig. 1(b) shows a satisfying example of such a fit to both $R_p(t)$ and $R_s(t)$ data, with parameters as follows. For the liquid surface, $\epsilon_{10} = 7.65$, $\epsilon_{20} = 12.3$, $\nu/\omega = 0.7$, and $\omega_p/\omega = 4.3$ were introduced because they are consistent simultaneously with the initial maximum liquid W reflectivity and the known resistivity[12] of W near the melting point. The fit of subsequent evolution then assumed this liquid as an initial condition. Continued heating of the liquid drives ν/ω to a maximum value of 1.8 at the R_p minimum, then electrons cool in ~3 ps. The observed R_p minima imply that the structure parameters for the three metals are in the ratio $C_{Al}:C_{Ag}:C_W = 0.99:1.1:1.0$ with reasonable liquid-state parameters $Z = 1.7$ and the effective optical mass $m_e^* = 2.0m_e$ [$m_e^*/m_e \approx 1$ for Al and Ag (Ref. 5)]. In the expansion region ($x > -v_0t$), ϵ_{10} and ϵ_{20} decrease with material density, approaching the vacuum values ($\epsilon_{10} = 1$, $\epsilon_{20} = 0$) far from the surface. We made the simplifying assumption that $(\epsilon_{10} - 1)$ and ϵ_{20} decrease linearly with density toward the vacuum values. The best fit, shown by the solid curve in Fig. 1(b), was then obtained with $v_{exp} = 0.33 \times 10^6$ cm/s, which is the same as the calculated value v_0. More rigorous models of the density dependence of ϵ_{10} and ϵ_{20} would necessitate modest changes in all parameters. Nevertheless, this simplified model demonstrates the influence of covalency both in the liquid and the expansion region of W. Similar conclusions are reached from experiments with other covalently bonded materials (e.g., carbon, silicon).

The slight deviation of liquid Ag and Al (Ref. 5) optical properties from a free-electron dielectric function can also be incorporated easily into the model. Optical data[5] justify $\epsilon_{10} = 1.3$, $\epsilon_{20} = 1.2$, and $\nu/\omega = 0.085$ for Ag and $\epsilon_{10} = 1.2$, $\epsilon_{20} = 1.1$, and $\nu/\omega = 0.31$ for Al at 620 nm, which yield the dashed curves for $R_s(t)$ in Fig. 1(a). The $R_p(t)$ fits hardly deviate from the previous ones, but the subunity values of $R_s(t)$ are better explained. For all the best $R_{p,s}$ fits, the data begin to deviate from the model for $t > 15$ ps. This discrepancy may be caused by the onset of nonspecular scattering, as shown by recent time-resolved measurements[13] from femtosecond photoexcited GaAs that show the scattering rising sharply for $\Delta t > 15$ ps. Thus the surface expansion model based on the Helmholtz equations appears valid only in a limited time interval $0 < \Delta t < 15$ ps.

This research was supported by the U.S. Air Force Office of Scientific Research (contract F4962089C-004), the U.S. Office of Naval Research (contract N0001-88-K-0663), the Robert A. Welch Foundation (grant F-1038), and the National Science Foundation (grant DMR-8858388).

References

1. H. M. Milchberg, R. R. Freeman, S. C. Davey, and R. M. More, Phys. Rev. Lett. **61**, 2364 (1988).
2. R. Fedosejevs, R. Ottmann, R. Sigel, G. Kühnle, S. Szatmari, and F. P. Schäfer, Appl. Phys. B **50**, 79 (1990).
3. D. H. Reitze, H. Ahn, and M. C. Downer, Phys. Rev. B **45**, 2677 (1992).
4. H. M. Milchberg and R. R. Freeman, J. Opt. Soc. Am. B **6**, 1351 (1989).
5. J. C. Miller, Philos. Mag. **20** 1151 (1969).
6. J. G. Fujimoto, J. M. Liu, E. P. Ippen, and N. Bloembergen, Phys. Rev. Lett. **53**, 1837 (1984).
7. P. B. Allen, Phys. Rev. Lett. **59**, 1460 (1987); R. H. M. Groeneveld, R. Sprik, and A. Lagendijk, Phys. Rev. Lett. **64**, 784 (1990).
8. The density of electron states is cited in W. A. Harrison, *Electronic Structure and the Properties of Solids* (Freeman, San Francisco, Calif., 1980), p. 524.
9. S. Kato, R. Kawakami, and F. Mima, Phys. Rev. A **43**, 3560 (1991); see Eq. (24).
10. R. M. More, K. H. Warren, D. A. Young, and G. B. Zimmerman, Phys. Fluids **31**, 3059 (1988).
11. B. Chichkov and T. Peter, Z. Phys. D **17**, 123 (1990).
12. R. C. Weast, ed., *CRC Handbook of Chemistry and Physics* (CRC Press, Boca Raton, Fla., 1985).
13. P. N. Saeta, Ph.D. dissertation (Harvard University, Cambridge, Mass., 1992).

Ionization-induced frequency shifts in intense femtosecond laser pulses

B. M. Penetrante and J. N. Bardsley

High Temperature Physics Division, Lawrence Livermore National Laboratory, Livermore, California 94550

W. M. Wood, C. W. Siders, and M. C. Downer

Department of Physics, University of Texas at Austin, Austin, Texas 78712

Received January 24, 1992

Electromagnetic plasma computer simulations are used to analyze the frequency shifts caused by the ionization of atmospheric-density noble gases during interaction with intense femtosecond laser pulses, the results are presented and compared with experimental data. The simulations trace the temporal evolution of plasma growth during the femtosecond ionizing pulse and calculate the resulting self-induced blue shift of the ionizing pulse spectrum. Variations with pulse intensity, gas pressure, and gas species are calculated. The relative contributions of strong-field ionization and electron-impact ionization on the frequency shifts are discussed. The simulations provide qualitative explanations of most of the features observed experimentally in the blue-shifted spectra. The technique of spectral blue shifting intense femtosecond laser pulses provides a new diagnostic tool for studying strong-field ionization and laser-induced breakdown in dense plasmas.

1. INTRODUCTION

Ultrashort-pulse lasers with intensities greater than 10^{16} W/cm^2 have been clearly demonstrated. These lasers make possible the ionization of small regions of gas in a time of the order of the pulse duration. The rapid creation of the plasma by strong-field ionization is accompanied by a fast decrease in the index of refraction of the medium. The incident laser beam is thus transmitted through a region of rapidly varying refractive index, which causes a frequency upshift in the laser pulse. Such field-ionization-induced blue shifts were observed in the experiments of Wood and co-workers[1–3] with 100-fs, 620-nm, 10^{16} W/cm^2 laser pulses.

Even for moderate laser intensities (10^{11}–10^{12} W/cm^2) it is known that the laser frequency can shift as a result of the variation of the refractive index of the forming plasma. Yablonovitch and Bloembergen observed spectral broadening and frequency shifts in laser-breakdown plasmas.[4–8] In their case plasma formation is caused by avalanche ionization resulting from electron impact. More recently, because of the availability of lasers with intensities of 10^{15} W/cm^2 and above, Wilks *et al.*[9] and Gildenburg *et al.*[10] suggested that the mechanism of strong-field ionization of the atoms of a gas can be used for effective control of the frequency upshifts of the laser radiation. These field-ionization-induced blue shifts were verified in the experiments of Wood and co-workers[1–3] in the noble gases. Yablonovitch[11] and Downer *et al.*[12] also explain that such blue shifts are implied in the above-threshold ionization experiments of Freeman *et al.*[13] For a different frequency range, Joshi *et al.*[14] demonstrated the frequency upshifting of microwave radiation by laser-ionized plasmas.

The frequency shift can be approximated by the Drude model[11] as

$$\Delta\omega = -\frac{\omega_0}{c}\int^{L}\frac{\partial n(x,t)}{\partial t}\,dx,$$

where ω_0 is the original frequency, c is the speed of light in vacuum, L is the interaction length, $n(x,t)$ is the index of refraction along the propagation direction x,

$$n^2 = 1 - \frac{\omega_p^2}{(\omega_0^2 + \gamma^2)},$$

ω_p is the plasma frequency, and γ is the collisional frequency. This model provides a simple scaling formula for the frequency shift of the transmitted laser beam. According to this model, large frequency shifts can be obtained by simply increasing the interaction length or the gas density. The creation of intense fields can currently be accomplished only through tight focusing of the laser beam; therefore long interaction lengths may be difficult to achieve because of diffraction effects. Spatial gradients of the induced index serve to limit this interaction length further by defocusing the laser beam. The high gas pressures at which the critical density of free electrons is achieved would limit the efficiency of the frequency upshift process, because a significant portion of the laser beam would be reflected. High gas pressures also tend to increase defocusing because of the higher plasma density and resulting higher index gradients in the interaction region.

The initial motivation for this investigation came from the experiments of Wood and co-workers.[1–3] These experiments have shown that the magnitude of the blue shift, as well as the shape and time development of the blue-shifted spectra, depends strongly and reproducibly on noble-gas species, gas pressure, and ionizing laser intensity, as is discussed in Section 3. Our main objective in the numerical simulations is to examine whether the ... or

0740-3224/92/112032-09$05.00 © 1992 Optical Society of America

experimental trends can be explained in the context of appropriate theories of ionization in intense light fields. To this end, special emphasis is placed on two points. First, the tunneling formula of Ammosov *et al.*[15] is used to evaluate field-ionization rates, in contrast to earlier preliminary analyses[2] of the blue shifts based on the Drude model with Keldysh[16] field-ionization rates. Since recent ion-yield experiments[17,18] show field ionization rates that are in discord with the Keldysh formula and in better agreement with the tunneling formula of Ammosov *et al.*,[15] it is important that estimates of the frequency shifts be evaluated by using the latter formula. Second, the possible contribution of collisional ionization to the observed blue shifts is evaluated. Recent studies[3] of the pressure, the intensity, the species, and the time dependences of the blue shifts provided evidence that electron-impact ionization begins to supplement field ionization in the heavy noble gases at the highest light intensity (10^{16} W/cm^2) and pressure [5 atm (3.8 kTorr)] studied experimentally. A Drude model that combined field- and impact-ionization rates yielded qualitative agreement with the data.[3] Here, using a more rigorous procedure in which the wave propagation is solved consistently with electron dynamics, we reexamine the relative significance of these ionization processes.

We use a one-dimensional electromagnetic plasma simulation code to study the ionization-induced self-frequency shifts of the femtosecond ionizing laser pulses. A description of the simulation model is given in Section 2. In Section 3 we present some of the experimental results and discuss the reason for invoking the significance of the collisional ionization process in explaining the observed frequency shifts. A more comprehensive presentation of the experimental results is being prepared.[19] The numerical results are then presented in Section 4. The conclusions are given in Section 5.

2. SIMULATION MODEL

The model solves Maxwell's equations for the fields and the continuity equations for the electron density and current. The pulse lengths are short enough that ion motion is negligible. We consider a linearly polarized electromagnetic wave traveling along x; there are no variations in y or z ($k_y = k_z = 0$). The transverse electromagnetic wave components are $E_y(x,t)$ and $B_z(x,t)$. The source wave is incident from left to right on a gas cell, for which we specify a gas density uniform along x for a given interaction length L. The propagation of the wave is simulated by integration of Maxwell's equation:

$$\frac{\partial \mathbf{E}}{\partial t} = c\nabla \times \mathbf{B} - 4\pi \mathbf{J}, \qquad \frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E}.$$

Our previous simulations show that for laser intensities of 10^{16}–10^{17} W/cm^2 the dynamics of the plasma electrons is dominated by the transverse motion.[20] Furthermore, for plasma densities greater than 10^{18} cm^{-3} the spatial profile of the electron density, $n_e(x,t)$, is well constrained by the space-charge forces. Thus the electron density is determined only by ionization, while the interaction of the field with the plasma is determined only by the transverse current density, $J_y(x,t)$. Let $N_j(x,t)$ be the number density of ions with charge state j. Assuming a stepwise ionization process,

$$\frac{\partial N_0}{\partial t} = -W_1 N_0,$$

$$\frac{\partial N_j}{\partial t} = W_j N_{j-1} - W_{j+1} N_j,$$

$$\frac{\partial N_{j\max}}{\partial t} = W_{j\max} N_{j\max-1},$$

$$n_e = \sum_{j=1}^{j\max} j N_j,$$

where W_j is the total ionization rate for the production of charge state j. We take both field ionization and electron-impact ionization into account. The fluid equation for the current density is

$$\frac{\partial J_y}{\partial t} = -\frac{e^2}{m} n_e E_y.$$

The Keldysh formula[16] for the ionization rate is commonly used to describe high-intensity laser–atom interactions. An important feature of the Keldysh theory is that the ionization rate is determined more by the properties of the outgoing electron than by the details of the atomic structure. The full form of the Keldysh formula for the ionization rate is

$$W = A\omega\left[\frac{U_i(0)}{\hbar\omega}\right]^{1/2}\left[\frac{\gamma}{(1+\gamma^2)^{1/2}}\right]^{3/2} S\left[\gamma, \frac{U_i(\omega)}{\hbar\omega}\right] \times \exp\left\{-\frac{2U_i(\omega)}{\hbar\omega}\left[\sinh^{-1}\gamma - \gamma\frac{(1+\gamma^2)^{1/2}}{1+2\gamma^2}\right]\right\},$$

where γ is the tunneling parameter, given by

$$\gamma = \omega\frac{[2\quad U_i(0)]^{1/2}}{E},$$

E is the electric field stre n, ω is the laser frequency, and $U_i(0)$ is the field-free ionization potential. The other terms in the formula are

$$U_i(\omega) = U_i(0)\left(1 + \frac{1}{2\gamma^2}\right),$$

$$S\left[\gamma, \frac{U_i(\omega)}{\hbar\omega}\right] = \sum_{m=0}^{\infty}\left(\exp\left\{-\left[2\left\langle\frac{U_i(\omega)}{\hbar\omega}+1\right\rangle - \frac{U_i(\omega)}{\hbar\omega}+m\right]\left[\sinh^{-1}\gamma - \frac{\gamma}{(1+\gamma^2)^{1/2}}\right]\right\}\right) \times \Phi\left(\left\{\frac{2\gamma}{(1+\gamma^2)^{1/2}}\left[\left\langle\frac{U_i(\omega)}{\hbar\omega}+1\right\rangle - \frac{U_i(\omega)}{\hbar\omega}+m\right]\right\}^{1/2}\right),$$

$$\Phi(x) = \int_0^x \exp\quad - x^2)dt.$$

The factor A in the Keldysh formula is of order unity and accounts for a weak dependence on the details of the atom.

In most of our calculations we choose the tunneling formula of Ammosov *et al.*[15] to calculate the field-ionization

rates. These rates are much higher than those calculated with the Keldysh formula.[16] Recent ion-yield experiments[17,18] show that the Ammosov *et al.* formula is in better agreement with data. The Ammosov *et al.* field-ionization rate is given by

$$W = 1.61\omega_{au}\frac{Z^2}{n_{eff}^{4.5}}\left(10.87\frac{Z^2}{n_{eff}^4}\frac{E_{au}}{E}\right)^{2n_{eff}-1.5} \times \exp\left(-\frac{2}{3}\frac{Z^3}{n_{eff}^3}\frac{E_{au}}{E}\right),$$

where ω_{au} is the atomic unit of frequency ($=4.1\times10^{16}$ s^{-1}). The effective quantum number is obtained by equating the ionization potential U_i of the critical electron orbit with $(Z^2/n_{eff}^2)U_H$,

$$n_{eff} = \frac{Z}{(U_i/U_H)^{1/2}},$$

where U_H is the ionization potential of H ($=13.6$ eV) and Z refers to the residual charge seen by this electron. The formulation of Ammosov *et al.* also used in the analysis in Ref. 3.

A number of empirical formulas for the electron-impact ionization cross sections have been proposed. The formula derived by Lotz[21] is widely used and seems to be in good agreement with measured data.[22] The cross section for impact ionizing an ion in its electronic ground state in charge state i to produce a ground-state ion in charge state $i+1$ is

$$\sigma_{i\to i+1} = \sum_{j=1}\frac{a_{ij}q_{ij}}{E_e P_{ij}}\ln\frac{E_e}{P_{ij}}\left\{1 - b_{ij}\exp\left[-c_{ij}\left(\frac{E_e}{P_{ij}}-1\right)\right]\right\}.$$

E_e is the electron energy, P_{ij} is the ionization potential for the outer shell, P_{i2} is the binding energy for the first inner subshell, P_{i3} is the binding energy for the second inner subshell, etc. q_{ij} is the number of electrons in the jth shell. a_{ij}, b_{ij}, and c_{ij} are constants. For charge states greater than ~3, Lotz uses $a_{ij} = 4.5\times10^{-14}$ cm^2 eV^2 and $b_{ij} = c_{ij} = 0$. We use $a_{ij} = 9\times10^{-14}$ cm^2 eV^2 to take into account any effects caused by collisional ionization of excited state species. In the experimental situation the target ions are bathed in a light field as strong as 10^{16} W/cm^2 and are thus likely to be in excited electronic states. In an excited state the P_{ij} values become smaller, and consequently the ionization cross section increases. In fact Lotz[21] noted that the same formula could be used to calculate the enhanced cross section for a particular excited state. Nevertheless, the excited-state distribution is difficult to estimate. Furthermore, measurements of the impact-ionization cross section of electronically excited species are lacking. Consequently, for the remainder of this paper, we simply use twice the standard ground-state cross sections. This approximation should yield the correct qualitative contribution of collisional ionization to blue shifting but may underestimate its magnitude to some extent.

The source laser is represented by plane-polarized-wave solutions to Maxwell's equations in vacuum. A Gaussian temporal pulse profile is assumed for the laser intensity. The form of the fields incident upon the gas is thus

$$E_y = B_z = E_0\sin(\omega_0 t)\exp(-t^2/t_{pulse}^2),$$

where ω_0 is the frequency of the source laser and t_{pulse} is 0.6 times the pulse length (to give a Gaussian FWHM equal to the specified pulse length).

The numerical solution of Maxwell's equation is implemented in one dimension along the propagation direction x by using the advective differencing algorithm.[23] The transverse fields are separated into left- and right-going components. By choosing $\Delta x = c\Delta t$, these components are advanced in time simply by shifting the values over one cell. Let E_y^R and E_y^L be the right- and left-going components, respectively. These field components are advanced in time as

$$E_y^R(t+\Delta t, x+\Delta x) = E_y^R(t,x) - \pi\Delta t\left[J_y\left(t+\frac{\Delta t}{2}, x\right) + J_y\left(t+\frac{\Delta t}{2}, x+\Delta x\right)\right],$$

$$E_y^L(t+\Delta t, x) = E_y^L(t, x+\Delta x) - \pi\Delta t\left[J_y\left(t+\frac{\Delta t}{2}, x\right) + J_y\left(t+\frac{\Delta t}{2}, x+\Delta x\right)\right].$$

The transmitted and reflected fields are defined as $E_y^R(t, x=L)$ and $E_y^L(t, x=0)$, respectively. The boundary values for the incident fields are

$$E_y^R(t, x=0) = E_0\sin(\omega_0 t)\exp(-t^2/t_{pulse}^2),$$
$$E_y^L(t, x=L) = 0.$$

The electron current is advanced in time as

$$J_y\left(t+\frac{\Delta t}{2}, x\right) - J_y\left(t\frac{\Delta t}{2}, x\right) = \Delta t\frac{e^2}{m}n_e(t,x)[E_y^R(t,x) + E_y^L(t,x)].$$

The electron density is obtained from the ion charge state distribution, which evolves according to a simple Euler scheme. Tests were made for the stability of the ion charge state evolution by using more robust, but more time consuming, ordinary differential equation solvers.

Several checks of the Maxwell solution were performed by using an incident laser pulse with a Gaussian temporal profile. When the temporal derivative of the refractive index is constant for the duration of the pulse, the resulting profile will be a Gaussian of the same width as the original but centered at $\omega = \omega_0 + \Delta\omega$. The entire spectrum will have shifted by the amount $\Delta\omega$. As a test case we assume that the ionization is produced externally, independently of the laser pulse, is uniform over the interaction length, and has a constant temporal derivative. According to the Drude model, for an incident wavelength λ_0 and an interaction length L the shift is given by

$$\Delta\lambda = -\frac{e^2L\lambda_0^3}{2\pi m_e c^3}\frac{dn_e}{dt}.$$

For an incident laser wavelength of 620 nm and an interaction length of 10 μm an ionization rate of 2.805×10^{33} cm^{-3} s^{-1} will result in a spectral blue shift of 10 nm.

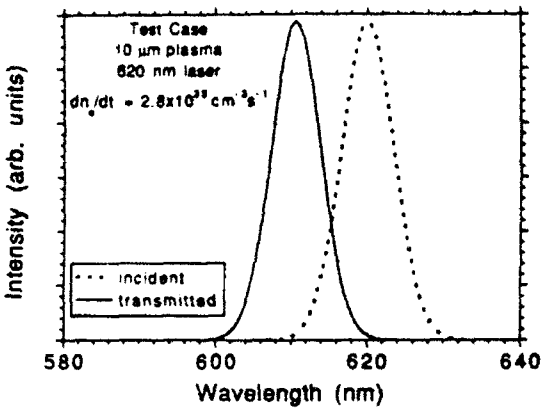


Fig. 1. Calculated spectra for a 100-fs laser pulse with an original wavelength of 620 nm. The laser is propagated through a spatially uniform plasma with an interaction length of 10 μm and an ionization rate of 2.805×10^{23} cm^{-3} s^{-1}. The dotted curve is the incident spectrum; the solid curve is the transmitted spectrum.

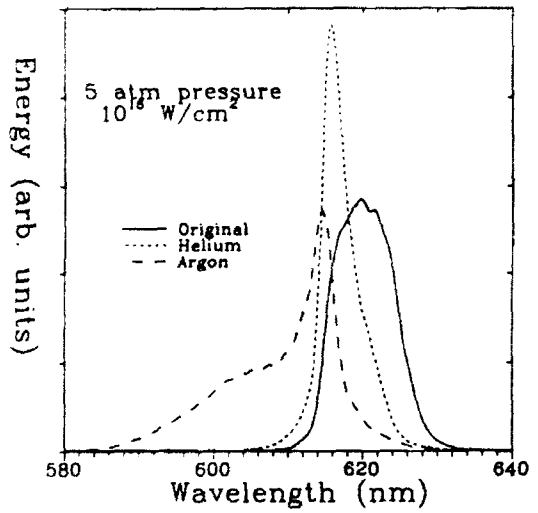


Fig. 2. Measured spectra after shifting through 5 atm (3.8 kTorr) of He and Ar gases with a peak laser intensity of 10^{16} W/cm^2.

Figure 1 shows the transmitted pulse spectrum as obtained from the Maxwell solver for this test case. Since there is no spatial variation within the plasma, the solution agrees with the Drude model, as expected.

3. EXPERIMENTAL RESULTS

In this section we present some of the experimental results and discuss the reason for invoking the significance of the collisional ionization process in explaining the observed frequency shifts.

In the experiments of Wood and Downer[1-3,12] 100-fs laser pulses with center wavelengths of 620 nm and energies less than or equal to 0.4 mJ are focused at *f*/5 into a glass cell containing a pressure of between 1 and 5 atm (0.76 and 3.8 kTorr) of He, Ne, Ar, Kr, or Xe gases. The minimal focal spot size is measured as 7 μm in diameter in order for the intensity to reach 1/*e* of the central intensity maximum. Light transmitted through the focal region is collected and analyzed by a spectrometer. The spectra of the laser pulses are recorded for varying gases, pressures, and pulse energies. Typically less than 1% of the pulse energy is lost in ionizing the gas.

Figure 2 shows the original pulse spectrum and the measured spectra after shifting through 5 atm of He and Ar gases with a nominal peak intensity of 10^{16} W/cm^2. In He a significant amount of the pulse energy remains unshifted, indicating that the derivative of the index of refraction or, equivalently, the ionization rate, is zero during a significant portion of the laser pulse. In Ar almost the entire pulse spectrum is shifted away from the original spectrum, indicating a nonzero value for the derivative of the index or, equivalently, for the ionization rate, during the entire pulse. The same qualitative feature is observed in Kr and Xe gases.

Figure 3 shows the effect of varying the laser pulse energy on the measured spectrum after breakdown in Kr gas at 5 atm of pressure. The center of the unshifted pulse spectrum is indicated by a vertical line. As the pulse energy is increased, the spectrum peak moves toward the blue until the pulse energy reaches 0.01 mJ [−log(Energy) = 5.0], after which the position of the peak appears to change little. As the energy increases further, a shoulder appears on the blue side of the spectrum, which becomes broader and shifts farther toward the blue.

Figure 4 shows the measured spectra after interaction with each of the noble gases at 5 atm of pressure for a pulse energy of 0.25 mJ. A blue shift occurs in all cases. However, the shape of the blue-shifted spectrum depends strongly on the gas species. In Ar, Kr, and Xe the narrow, less-shifted peak and the broad, blue shoulder are clearly discernible, and the shoulder appears as a separate peak. The blue shoulder shows greater shifting as the

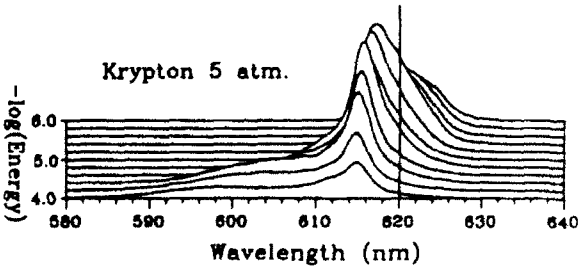


Fig. 3. Measured spectra after interaction with 5 atm of Kr as a function of pulse energy. Pulse energy is increasing toward the bottom of the figure in steps of ($\times 10^{0.2}$). The center of the original spectrum is indicated by a vertical line at 620 nm.

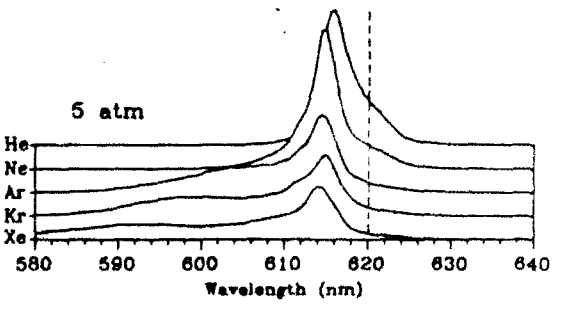


Fig. 4. Measured spectra after interaction of 0.25-mJ pulses in 5 atm of each of the noble gases studied. The center of original pulse spectrum is indicated by a vertical dashed line at 620 nm

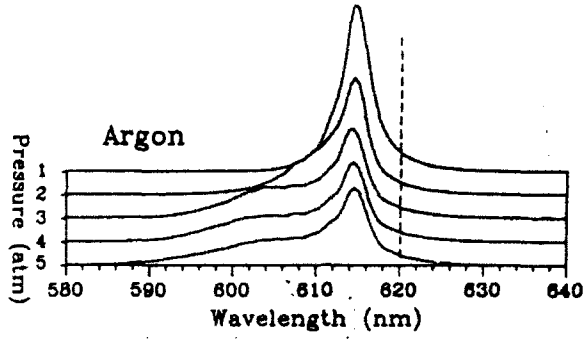


Fig. 5. Measured spectra after interaction of 0.25-mJ pulses with Ar gas as a function of gas pressure. Pressure is increasing toward the bottom of the figure. The center of original pulse spectrum is indicated by a vertical dashed line at 620 nm.

atomic weight of the gas is increased, corresponding to the plasma densities that are attained in the heavier gases. At the same time the height of the narrow, less-shifted peak decreases. In He and Ne, in contrast, the blue shoulder is absent. Instead a residual shoulder appears, corresponding approximately to the unshifted spectrum.

The effect of gas pressure on the measured spectra in Ar for a pulse energy of 0.25 mJ is shown in Fig. 5. At all pressures shown there is the narrow peak that is blue shifted by 6 nm. Only for pressures above 1 atm does the broad, more-blue-shifted shoulder appear to be discernible. The blue shoulder shows greater shifting, and the height of the narrow, less-shifted peak decreases as the gas pressure is increased. Also, the breadth of the broad blue shoulder increases as the gas pressure is increased. Note the absence of a residual shoulder even at the lower pressures of Ar.

Although the amplitude and the width of these spectral features depend on the focal profile, chirp, and other details of the ionizing pulse, numerous measurements confirm that the qualitative trends shown in Figs. 2–5 are reproducible features of femtosecond ionization of noble gases. Time-resolved pump–probe experiments were also performed.[3,12] These experiments show that (in Ar, Kr, and Xe) the broad blue shoulder evolves well in the leading edge of the ionizing pulse, while the less-shifted peak arises at or slightly after the peak of the ionizing pulse. Furthermore, the single blue-shifted peak in He and Ne (like the broad blue shoulder in Ar, Kr, and Xe) evolves in the leading edge.

In calculating the plasma density, using either Keldysh or Ammosov *et al.* strong-field ionization but without collisional ionization, it is observed that at all intensities, in each gas, the ionization has saturated by the peak of the laser pulse, or within 10–20 fs after the peak of the pulse (see the dashed curves, Fig. 6). This result implies that in all gases a significant portion of the ionizing pulse energy would remain unshifted, as is observed in He. Because the data for Ar, Kr, and Xe do not exhibit this type of behavior, there is the implication that significant ionization occurs during and for some time after the peak of the laser pulse. This additional ionization may be caused by electron–impact collisions with the gas. The initial interpretation that comes from the Drude model calculations[2,3] is the following: Strong-field ionization before the peak of the laser pulse gives rise to the broad blue shoulder; collisional ionization after the peak of the laser pulse gives rise to the narrow, less-blue-shifted peak. In Section 4 we present numerical simulations to investigate whether this preliminary interpretation of the experimental data is correct.

4. CALCULATIONS OF IONIZATION-INDUCED EFFECTS

Figure 6(a) shows the calculated evolution of the degree of ionization in 5 atm of the noble gases illuminated by a 100-fs laser pulse with a peak intensity of 10^{16} W/cm^2, with the Keldysh rates used for strong-field ionization. The solid curves are calculations with collisional ionization according to the Lotz formula, while the dashed curves are calculations without collisional ionization. Figure 6(b) shows the corresponding calculations made using the Ammosov *et al.* rates for strong-field ionization. Note that, except for He, during the peak of the laser pulse there is a distinct difference in the degrees of ionization calculated with and without collisional ionization. This difference is quite dramatic when the Keldysh rates are used for the strong-field ionization. Although these differences are less significant when the Ammosov *et al.* rates are used, they are still apparent. Also, note that the degrees of ionization obtained with the Ammosov

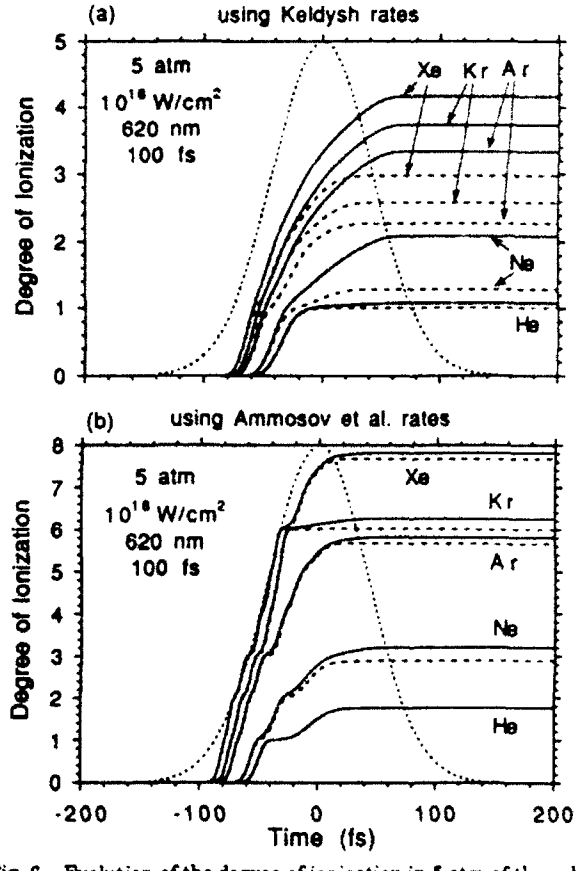


Fig. 6. Evolution of the degree of ionization in 5 atm of the noble gases with a peak laser intensity of 10^{16} W/cm^2, calculated using (a) Keldysh and (b) Ammosov *et al.* rates for strong-field ionization. The solid curves are calculations that include collisional ionization; the dashed curves are without collisional ionization. The time profile of the laser pulse is shown by the dotted curves

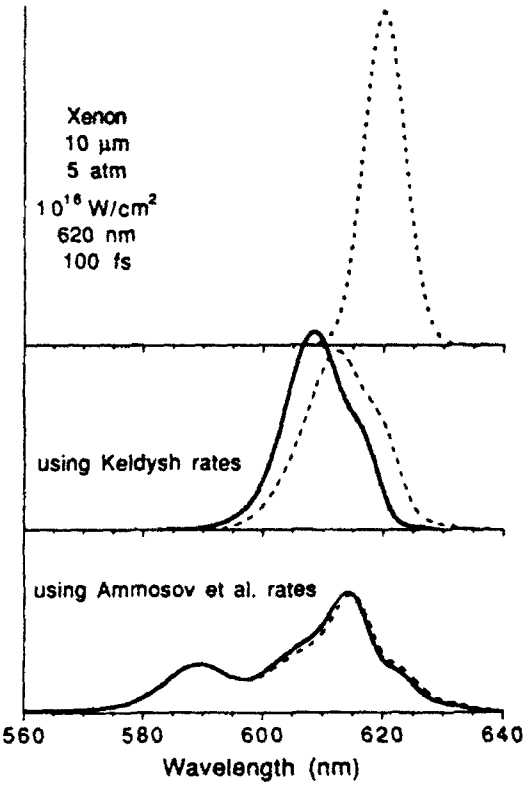


Fig. 7. Spectra for ionization of 5 atm of Xe by 100-fs pulse of peak intensity 10^{16} W/cm^2, calculated using Keldysh and Ammosov *et al.* rates for strong-field ionization. The laser-plasma interaction length is 10 μm. The dotted curves shows the original spectra; the solid curves are the shifted spectra including collisional ionization; the dashed curves are without collisional ionization.

et al. rates are much larger than those obtained with the Keldysh rates.

The magnitude of the differences in the ionization degree is reflected in the calculated spectral shifts. Figure 7 shows the spectra in 5 atm of Xe calculated by using the Keldysh rates, compared with that for the Ammosov *et al.* rates. The laser-plasma interaction length is assumed to be 10 μm in these calculations. With the Keldysh rates there is a dramatic difference in the spectral shifts with and without collisional ionization, whereas this difference is quite small when the Ammosov *et al.* rates are used to calculate the field ionization. The steeper rate of increase in the degrees of ionization obtained with the Ammosov *et al.* rates corresponds to much larger frequency blue shifts in the transmitted spectrum, as one would expect from the simple Drude model. Furthermore, only the Ammosov *et al.* formulation correctly yields the double-peaked spectral shape observed experimentally (see Fig. 4) with 5 atm of Xe, confirming that it provides the more accurate description of ionization under the experimental conditions.

The results shown in Figs. 6 and 7 suggest that the significance of collisional ionization depends strongly on the values of the field ionization rate. Recent experiments[17-18] show field ionization rates that are in better agreement with the tunneling formula of Ammosov *et al.*[16] For the remainder of this paper we use the Ammosov *et al.* rates to calculate the frequency shifts.

Figure 8 shows the calculated spectrum in 5 atm of the noble gases. The effect of collisional ionization is negligible in the case of He, as is evident also from Fig. 6(b), in agreement with earlier analyses.[2,3] For Ne the effect of collisional ionization is small but distinct. There is little contribution from collisional ionization during the second half of the laser pulse. Collisional ionization does not significantly alter the shape of the spectrum but merely shifts the entire spectrum a little bit toward the blue. The spectrum in Ne shows a prominent blue-shifted peak and a small residual peak, as is observed experimentally (see Fig. 4). For the heavier gases Ar, Kr, and Xe the frequency shifting manifests itself in two ways. First, there is the appearance of a broad, blue spectral shoulder that is shifted from the original by as much as 35 nm in Kr for the parameters used in Fig. 8. Temporally this feature arises mainly from the lower ionization stages, occuring well within the leading edge of the ionizing pulse, in agreement with the time-resolved experiments.[3] This feature is unaffected by including collisions in the model and thus originates purely from strong-field ionization. Second, the extra blue shifting resulting from the higher ionization degrees attained in the heavier gases sometimes causes the residual peak to be blue shifted even in the absence of collisions. Compare, for example, the

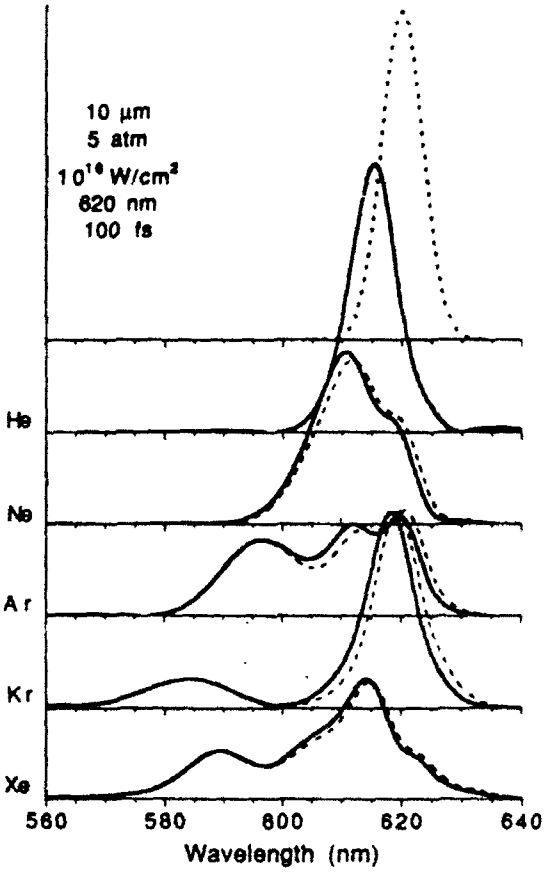


Fig. 8. Spectra for ionization of 5 atm of He, Ne, Ar, and Kr by a 100-fs pulse of peak intensity 10^{16} W/cm^2, calculated with the Ammosov *et al.* rates used for field ionization. The laser-plasma interaction length is 10 μm. The dotted curves shows the original spectra; the solid curves are the shifted spectra including collisional ionization; the dashed curves are without collisional ionization.

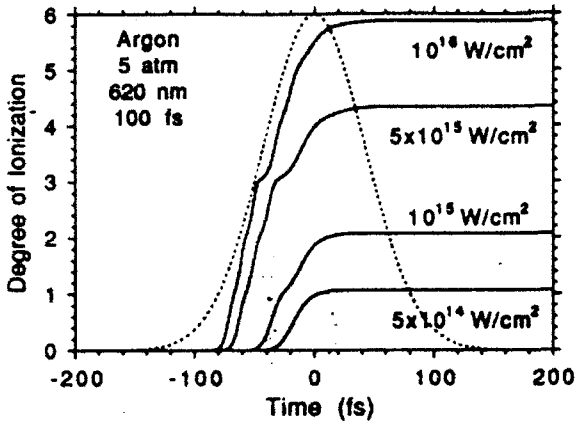


Fig. 9. Calculated evolution of the degree of ionization in 5 atm of Ar for various peak laser intensities. The time profile of the laser pulse is shown by the dotted curve.

spectra of Xe and Kr (dashed curves) in Fig. 8. Since the highest ionization stage in Xe occurs at a slower rate than the lower stages [Fig. 6(b)], it causes the blue shift of the tall, narrow, less-shifted peak rather than a further shift in the broad, blue shoulder. Furthermore, Fig. 6(b) shows that this final stage (and thus the blue shifting of the less-shifted peak) should occur near the peak of the ionizing pulse, in qualitative agreement with time-resolved experiments.[3] Thus femtosecond ionization of 5-atm Xe at 10^{16} W/cm^2 light intensity provides a specific example in which the main qualitative features of the data[3] can be explained without invoking collisional ionization at all. Elucidation of such an example is a major new point brought out by the current simulation, which was not evident in earlier analyses[2,3] of the blue-shift data.

Nevertheless, in other cases collisional ionization may still be necessary to account for the main features of the data. For example, without collisions in the model the Ar and Kr spectra (Fig. 8, dashed curves) each show a prominent unshifted spectral peak, in contradiction to the data (Fig. 4). With collisional ionization included, on the other hand, this residual peak is blue shifted. Furthermore, its extra shift occurs temporally near the peak of the ionizing pulse [see the solid curves in Fig. 6(b)], in qualitative agreement with time-resolved experiments.[3] Even in the case of Xe, collisional ionization supplements the effect described above by producing an extra shift of the less-blue-shifted feature near the temporal peak of the ionizing pulse. The calculated collision-induced shift, however, is smaller (1–2 nm) in magnitude than the shift (5 nm) observed experimentally. The assumption of twice the ground-state impact ionization cross section may be responsible for this discrepancy by underestimating the impact ionization rate of electronically excited species. Conceivably a high degree of electronic excitation could result in an even higher cross section.[21] A detailed discussion of this interpretation is in preparation.[10]

The calculated evolution of the degree of ionization in Ar is shown in Fig. 9 for various peak laser intensities. The corresponding spectra are shown in Fig. 10. At a laser intensity of 5×10^{14} W/cm^2, Ar is only singly ionized. At a pressure of 5 atm and an interaction length of 10 μm, this calculation produces a blue-shifted peak that is 4 nm from the original. A small residual shoulder corresponding to the far-red wing of the unshifted spectrum is evident. At a laser intensity of 10^{15} W/cm^2, Ar is doubly ionized. This produces a main blue shift of 7 nm. Furthermore, the residual shoulder is now centered closer to the center of the unshifted spectrum. At a laser intensity of 5×10^{15} W/cm^2, Ar is more than four times ionized. Almost the entire spectrum is shifted away from the original spectrum. The spectrum becomes broader and is shifted farther toward the blue. Concurrently relatively little unshifted light remains. This produces a new main peak, which is blue shifted by ~4 nm, and a broad, blue shoulder that is blue shifted by ~14 nm. At a laser intensity of 10^{16} W/cm^2, approximately half of this main peak is blue shifted farther, causing a subdivision into two peaks. The other half is once again centered on the original pulse spectrum. The broad, blue shoulder is also shifted farther, giving the appearance of a third peak.

The effect of gas pressure on the calculated spectra in Ar is shown in Fig. 11 for a peak laser intensity of 10^{15} W/cm^2. As the pressure is increased to 2 atm, only one peak is prominent, and the blue shift is simply proportional to the pressure. As the pressure is increased to 3 atm this peak begins to subdivide, giving the appearance of a plateau. At a pressure of 4 atm the broad, blue shoulder becomes evident, as well as the narrow, less-blue-shifted peak.

The above calculations show additional examples of conditions under which the simulation without collisions can reproduce the main qualitative features of the data. Most noteworthy is the case of 5-atm Ar at 5×10^{15} W/cm^2 (Fig. 10). In this case the simulation yields a broad blue shoulder and a less-shifted peak, with little residual un-

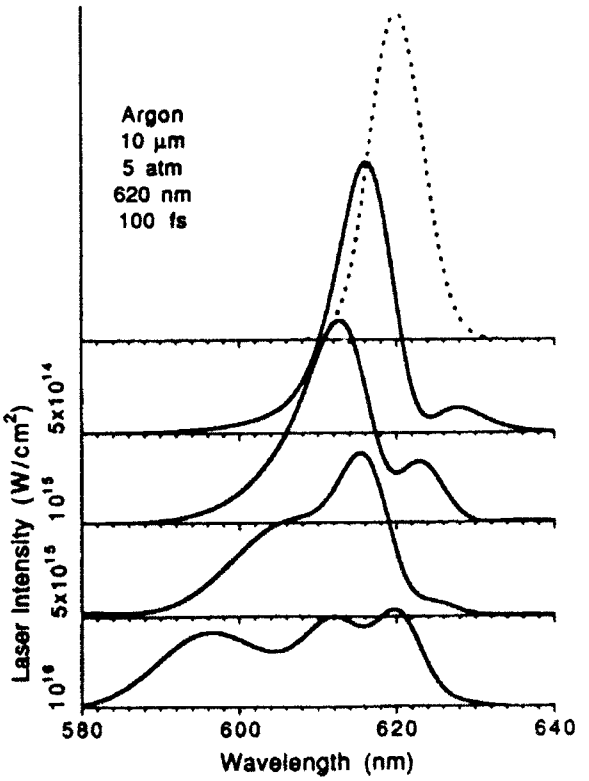


Fig. 10. Calculated spectra in 5 atm of Ar for various peak laser intensities. The dotted curve shows the original spectrum.

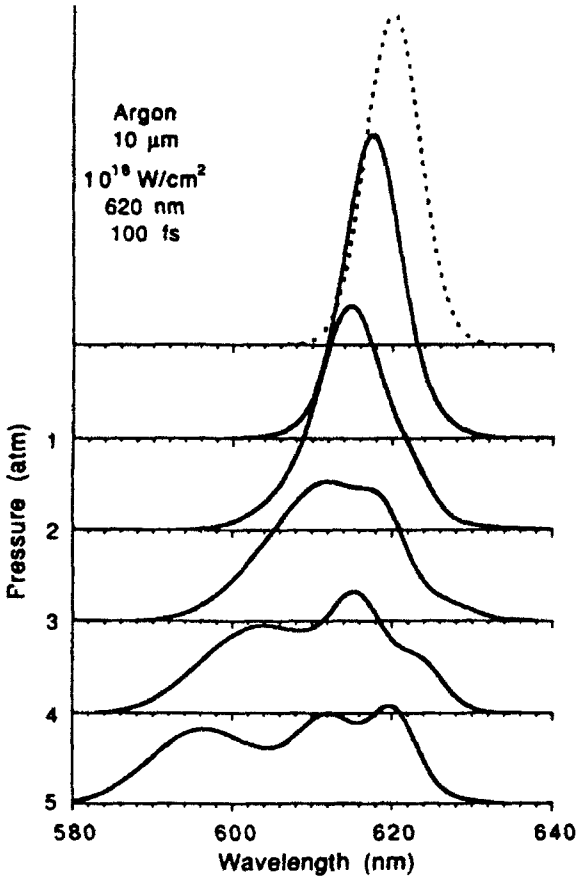


Fig. 11. Calculated spectra for ionization of Ar at various pressures with a peak laser intensity of 10^{16} W/cm^2. The dotted curve shows the original spectrum.

shifted energy, quite like the corresponding measured self-blue-shifted spectra (Figs. 4 and 5). Furthermore, the calculated temporal behavior qualitatively agrees with time-resolved experiments.[3] In other cases, however, the simulated blue shift disagrees significantly with the experiment. For example, as intensity increases to 10^{16} W/cm^2 in 5-atm Ar, the reappearance of a large fraction of unshifted energy (caused by completion of sixfold ionization in the leading edge of the pulse) is a trend not observed experimentally. On the other hand, collisional ionization at a sufficient rate might maintain the blue shifting of the entire pulse spectrum as intensity increases.

A more definitive comparison between experiment and simulation cannot be drawn at this time for two reasons. First, the experimental peak intensity at the focus is difficult to measure. Our estimates are only accurate to within a factor of 2 or 3. Second, the simulation is 1-D and thus does not take into account either the spatial averaging along the radius of the laser-plasma interaction region or the defocusing of the laser pulse at higher gas pressures. To include such effects would require a two-dimensional calculation. Nevertheless, even the 1-D simulation successfully reproduces most of the features of the data. Uncertainty remains in the interpretation of some details, notably the importance of collisional ionization in producing a secondary blue shift near the temporal peak of the ionizing pulse in several atmospheres of the heavier rare gases.

5. CONCLUSIONS

The technique of spectral blue shifting of intense femtosecond laser pulses provides a powerful diagnostic of plasma evolution at atmospheric densities. Both the intensity and the short rise times of these laser pulses lead to high rates of plasma formation, which in turn lead to measurable blue shifts in the transmitted spectra. Such experiments significantly complement standard diagnostics of strong-field ionization, such as ion yield and photoelectron spectroscopy, by providing femtosecond time resolution and by permitting experiments in much denser plasmas. Using a more rigorous procedure than previous analyses,[2,3] the 1-D electromagnetic plasma simulations presented here provide qualitative explanations of most of the features observed experimentally in the blue-shifted spectra.

The magnitude of the spectral shifts and the relative importance of collisional ionization depend strongly on both the field-ionization and collisional-ionization rates used in the simulations. Our simulations lead to two main conclusions that were not evident in previous analyses.[2,3] First, the contribution of collisional ionization to the evolution of the plasma density is almost negligible if we use the Ammosov *et al.* rates for field ionization and the Lotz rates for electron-impact ionization of electronic ground-state species. If, on the other hand, the impact ionization cross sections are enhanced (by >2 times) because of electronic excitation of the target ions, collisional ionization would then contribute substantially in a manner consistent with the data. Second, at some excitation levels in many of the gases the 1-D simulation explains the main qualitative features of the data without invoking a collisional ionization mechanism. However, some cases remain (high intensity and pressure in some of the heavy rare gases) for which explaining the data may still require a collisional mechanism. Further studies will be needed before the role of collisional ionization can be established or ruled out definitely.

For more detailed comparison with experiment, a two-dimensional calculation is necessary if effects such as beam defocusing are to be included.

ACKNOWLEDGMENTS

The research of B. M. Penetrante and J. N. Bardsley was performed under the auspices of the U. S. Department of Energy by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48. The research of W. M. Wood, C. W. Siders, and M. C. Downer was supported by the National Science Foundation (grant DMR8858388), and Robert A. Welch Foundation (grant F-1038), and the U.S. Air Force Office of Scientific Research (contract F49620-89-C-0044).

REFERENCES

1. W. M. Wood, G. Focht, and M. C. Downer, Opt. Lett. 13, 984 (1988).
2. M. C. Downer, G. Focht, D. H. Reitze, W. M. Wood, and T. R. Zhang, in *Ultrafast Phenomena VI*, T. Yajima, K. Yoshihara,

C. B. Harris, and S. Shionoya, eds. (Springer-Verlag, Berlin, 1988), pp. 128–131.
3. W. M. Wood, C. W. Siders, and M. C. Downer, Phys. Rev. Lett. **67**, 3523 (1991).
4. E. Yablonovitch and N. Bloembergen, Phys. Rev. Lett. **29**, 907 (1972)
5. N. Bloembergen, Opt. Commun. **8**, 285 (1973).
6. E. Yablonovitch, Phys. Rev. Lett. **31**, 877 (1973).
7. E. Yablonovitch, Phys. Rev. Lett. **32**, 1101 (1974).
8. E. Yablonovitch, Phys. Rev. A **10**, 1888 (1974).
9. S. C. Wilks, J. M. Dawson, and M. B. Mori, Phys. Rev. Lett. **61**, 337 (1988).
10. V. B. Gildenburg, A. V. Kim, and A. M. Sergeev, Pis'ma Zh. Eksp. Teor. Fiz. **51**, 91 (1990) [JETP Lett. **51**, 104 (1990)].
11. E. Yablonovitch, Phys. Rev. Lett. **60**, 795 (1988).
12. M. C. Downer, W. M. Wood, and J. I. Trisnadi, Phys. Rev. Lett. **65**, 2832 (1990).
13. R. R. Freeman, P. H. Bucksbaum, H. Milberg, S. Darak, D. Schumacher, and M. E. Geusic, Phys. Rev. Lett. **59**, 1092 (1987).
14. C. J. Joshi, C. E. Clayton, K. Marsh, D. B. Hopkins, A. Sessler, and D. Whittum, IEEE Trans. Plasma Sci. **18**, 814 (1990).
15. M. V. Ammosov, N. B. Delone, and V. P. Krainov, Zh. Eksp. Teor. Fiz. **91**, 2008 (1986) [Sov. Phys. JETP **64**, 1191 (1986)].
16. L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
17. S. Augst, D. Strickland, D. D. Meyerhofer, S. L Chin, and J. H. Eberly, Phys. Rev. Lett. **63**, 2212 (1989).
18. S. Augst, D. D. Meyerhofer, D. Strickland, and S. L. Chin, J. Opt. Soc. Am. B **8**, 858 (1991).
19. W. M. Wood, "Femtosecond time-resolved study of plasma generation and dynamics by blue shifting of high-intensity laser pulses," Ph.D. dissertation (University of Texas at Austin, Austin, Tex., 1991).
20. B. M. Penetrante and J. N. Bardsley, Phys. Rev. A **43**, 4100 (1991).
21. W. Lotz, Z. Phys. **206**, 205 (1967); **216**, 241 (1968).
22. H. Tawara and T. Kato, At. Data Nucl. Data Tables **36**, 167 (1987).
23. C. K. Birdsall and A. B. Langdon, *Plasma Physics via Computer Simulation* (McGraw-Hill, New York, 1985).

Quantum Electronics and Laser Science Conference

1992 Technical Digest Series
Volume 13

Conference Edition

Summaries of papers presented at the
Quantum Electronics and
Laser Science Conference
May 10–15, 1992
Anaheim, California

Sponsored by
American Physical Society
IEEE/Lasers and Electro-Optics Society
Optical Society of America

in cooperation with
Quantum Electronics Division of the European Physical Society
Japanese Quantum Electronics Joint Group

Optical Society of America
2010 Massachusetts Avenue, NW
Washington, DC 20036

Articles in this publication may be cited in other publications. In order to facilitate access to the original publication source, the following form for the citation is suggested:

Name of Author(s), "Title of Paper," in *Quantum Electronics and Laser Science Conference*, 1992 OSA Technical Digest Series, Vol. 13 (Optical Society of America, Washington, DC, 1992), pp. xx-xx.

ISBN Number

Conference Edition	1-55752-239-1
Postconference Edition	1-55752-240-5
(Note: Postconference Edition includes postdeadline papers.)	
1992 Technical Digest Series	1-55752-261-8

Library of Congress Catalog Card Number

Conference Edition	91-80623
Postconference Edition	91-80624

IEEE catalog number 92CH3153-4

Copyright © 1992, Optical Society of America

Individual readers of this digest and libraries acting for them are permitted to make fair use of the material in it, such as to copy an article for use in teaching or research, without payment of fee, provided that such copies are not sold. Copying for sale is subject to payment of copying fees. The code 1-55752-261-8/92/$2.00 gives the per-article copying fee for each copy of the article made beyond the free copying permitted under Sections 107 and 108 of the U.S. Copyright Law. The fee should be paid through the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970.

Permission is granted to quote excerpts from articles in this digest in scientific works with the customary acknowledgment of the source, including the author's name and the name of the digest, page, year, and name of the Society. Reproduction of figures and tables is likewise permitted in other articles and books provided that the same information is printed with them and notification is given to the Optical Society of America. Republication or systematic or multiple reproduction of any material in this digest is permitted only under license from the Optical Society of America; in addition, the Optical Society may require that permission also be obtained from one of the authors. Address inquiries and notices to Director of Publications, Optical Society of America, 2010 Massachusetts Avenue, NW, Washington, DC 20036. In the case of articles whose authors are employees of the United States Government or its contractors or grantees, the Optical Society of America recognizes the right of the United States Government to retain a nonexclusive, royalty-free license to use the author's copyrighted article for United States Government purposes.

Printed in the U.S.A.

Thursday AFTERNOON
May 14, 1992 QThD
Convention Center Hall B
1:00 pm **QELS Poster Session 3**

OPTICAL INTERACTIONS WITH CONDENSED MATTER AND ULTRAFAST PHENOMENA

QThD1 Temperature dependence of the refractive indices and optical band gap of thin amorphous silicon films

N. Do, P. T. Leung, L. Klees, A. C. Tam, F. Tong,* W. P. Leung†
IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose California 95120

The variation of the optical constants with temperature for amorphous silicon (a-Si), as well as other solids, is important in photothermal processing, such as laser annealing. Although crystal silicon and other semiconductor materials have been studied extensively, a-Si is less studied in this aspect. Previous studies have reported that there is insignificant temperature dependence of the absorption coefficient at the Nd:YAG laser wavelength (1.06 μm)[1,2] for a-Si samples of submicron thickness. This has then been sometimes assumed to be a general property of a-Si in studies involving laser pulse interaction with the material.[3]

In this work, we present a thorough study of the dependence of the real and imaginary refractive index $n(T)$ and $k(T)$ as functions of temperature T for a-Si. Such dependence is sensitive to sample thickness and optical wavelength.

To determine the optical constants for the samples, we have followed the standard reflectance (R) and transmittance (I) measurements in which R and I are measured at almost normal incidence and $n(T)$ and $k(T)$ at a fixed temperature are obtained from an iterative computer program. The samples we study have the thicknesses of 0.2 μm and 1.0 μm. The two different incident wavelengths are 0.752 μm and 1.15 μm. The results are shown in Fig. 1 and 2 from which we observe the almost linear rise of $n(T)$ with temperature which is comparable to that with crystal silicon (c-Si). However, $k(T)$ varies quite differently and, in general, increases at a slower rate. For the 0.2-μm thickness at 1.15-μm wavelength, dk/dT is indeed negligible; but not for a case with more absorptive wavelength or greater sample thickness. Thus, we conclude that, contrary to previous expectations,[1-3] the absorption coefficient of a-Si can have a relatively strong temperature dependence in the highly absorptive regime, and must be considered in laser processing of a-Si.

In addition, by applying the Mott-Davis formula to the data obtained from $n(T)$ and $k(T)$, results are obtained for the temperature dependence of the optical gap energy E_g for a-Si. Here, while E_g for a-Si is, in general, larger than that for c-Si, they have similar temperature variation and can all be fitted using the well-known empirical formula.[5] Results for $E_g(T)$ are shown in Fig. 3.

IBM Watson Research Center, Yorktown Heights, New York 10598; †IBM Storage System Product Division, 5600 Cottle Road, San Jose, California 95193

1. M. R. T. Siregar, M. von Allmen, W. Luthy, Helv. Phys. Acta 52, 45 (1979).
2. M. von Allmen *et al.*, Appl. Phys. Lett. 34, 82 (1979).
3. C. K. Ong, H. S. Tan, E. H. Sin, Mater. Sci. Eng. 79, 79 (1986).
4. See, e.g., the review article by G. E. Jellison, Jr., in *Semiconductors and Semimetals*, Vol. 23 (Academic Press, New York, 1984), R. F. Wood, C. W. White, R. T. Young, Eds., p. 95, and references therein.
5. Y. P. Varshni, Physics (Utrecht) 34, 149 (1967).

QThD2 Rapid-scan femtosecond ellipsometry of $Si_{1-x}Ge_x$ and C_{60} crystalline films

H. R. Choo, M. C. Downer, V. P. Kesan*
Physics Department, University of Texas, Austin, Texas 78712

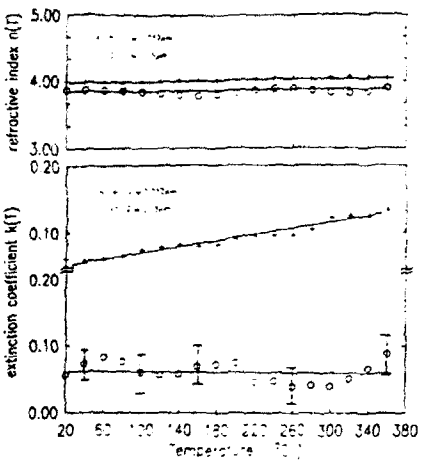
Recent studies have shown that the femtosecond optical response of semiconductors depends sensitively on, and can provide a quantitative indicator of, key sample imperfections such as implantation damage,[1] defects induced by low growth temperature,[2] nonuniform doping profiles,[1] and interface roughness.[1] In this paper, we use an unamplified colliding-pulse mode-locked (CPM) laser (λ = 620 nm) to obtain the femtosecond ellipsometric response in air of $Si_{1-x}Ge_x$ and C_{60} crystalline films, for which, to our knowledge, no previous femtosecond studies exist. We show that femtosecond response of $Si_{1-x}Ge_x$ grown epitaxially on Si depends sensitively on the depth of the strained SiGe/Si interface beneath the surface, while the response of C_{60} films depends on the degree of crystallinity as determined independently by x-ray diffraction. In each case, the femtosecond response of the samples of highest crystalline quality reveals the basic nonequilibrium bulk carrier dynamics, whereas the response of less perfect samples reveals alterations in carrier dynamics attributable to strained interfaces, alloy separation, or amorphization. A "rapid scan"[3] of pump–probe time delay Δt provides rapid data acquisition. The samples are pumped at normal incidence and probed either at near-normal incidence or at oblique incidence ($\Theta \sim 70°$) through an ellipsometric optical sequence (polarizer, compensator, sample, analyzer, differential detector) to provide complete optical characterization of film and substrate at each Δt.

Figure 1(a) shows the femtosecond reflectivity response of a complete compositional family of optically thick $Si_{1-x}Ge_x$ films immediately following growth by molecular beam epitaxy (MBE) on Si substrates. For these samples, the strained SiGe/Si interface lies well below the optical probe depth. Thus, the carrier dynamics probed are those of relaxed $Si_{1-x}Ge_x$, and are understandable in terms of the bulk band structure. Specifically, with increasing x, the amplitude of the initial response increases with increasing pump absorption coefficient, and thus increasing initial carrier density. In addition, the one-component temporal relaxation for Si-like films ($x \leq 0.6$) changes to a two-component response for Ge-like ($x \geq 0.6$) films, because of a change in carrier dynamics resulting from different initial k-space distributions. In Ge-like samples, direct $L \rightarrow L$ valley pump absorption creates sufficient excess hole energy to allow delayed secondary carrier generation by impact ionization which, together with intervalley $L \rightarrow \Gamma$ hole scattering, produces the observed delayed secondary reflectivity decrease. As the epitaxial film becomes thinner than an optical probe depth, however, carrier dynamics within the strained interface region increasingly dominate the observed probe response. In these cases, we observe sharply reduced carrier lifetime, as illustrated by the reflectivity responses in Fig. 1(b) for strained 800 Å SiGe films grown by RPCVD (bottom) and MBE (middle). The top trace in Fig. 1(b) was obtained from the same $x = 0.5$ sample as the corresponding Fig. 1(a) trace after several months of room-temperature storage. A completely different, small-amplitude, two-component response is now observed, suggesting alloy separation into local regions of pure Si and Ge. However, no such changes occurred in the other samples.

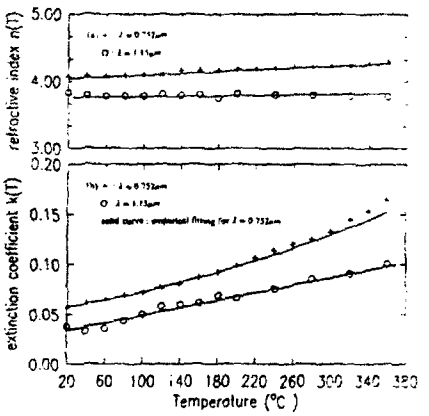
Figure 2 (top trace) shows the femtosecond oblique incidence response of an optically thick film grown by sublimating unpurified C_{60} powder onto glass. X-ray diffraction showed oriented epitaxial fcc crystalline structure.[4] An initial 1-ps recovery is followed by a steady state, which we tentatively interpret as ultrafast relaxation of initially hot electron-hole pairs to long-lived, 1.5-eV direct-gap band-edge states at the X point.[5] The x-ray diffraction of other grown films indicated amorphized C_{60}. In these cases the femtosecond response shown in the bottom trace of Fig. 2 was observed. Although the first 1 ps is similar, continuing relaxation is observed instead of a steady state, suggesting that amorphization induces rapid nonradiative recombination channels.

IBM Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598

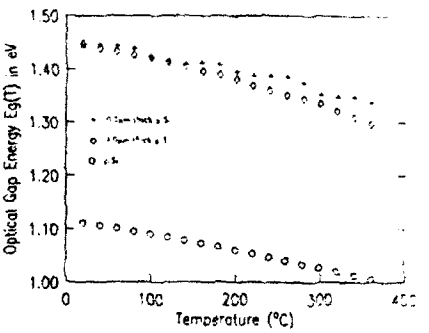
1. W. Kutt *et al.*, "Nonlinear optical probing of surface defects in semiconductors with femtosecond pulses," in *Quantum Electronics and Laser Science Conference*, Vol. 11, 1991 Technical Digest Series (Optical Society of America, Washington DC, 1991), paper QWD17, p. 142; A. Esser *et al.*, Appl. Surf. Sci. 46, 446 (1990); G. C. Cho *et al.*, Phys.



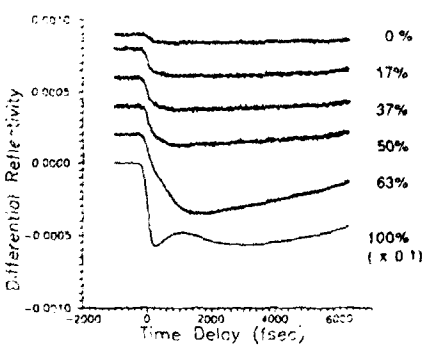
QThD1 Fig. 1. 0.2-μm a-Si sample; (a) Refractive index as a function of temperature, (+) λ = 0.752 μm; (O) λ = 1.15 μm. (b) Extinction coefficient as a function of temperature; (+) λ = 0.752 μm; (O) λ = 1.15 μm.

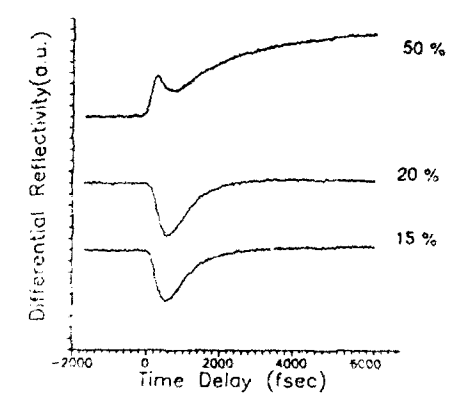


QTHD1 Fig. 2. 1.0-μm a-Si sample; (a) Refractive index as a function of temperature, (+) λ = 0.752 μm; (O) λ = 1.15 μm. (b) Extinction coefficient as a function of temperature; (+) λ = 0.752 μm; (O) λ = 1.15 μm solid curve is a quadratic polynomial fit for λ = 0.752 μm.

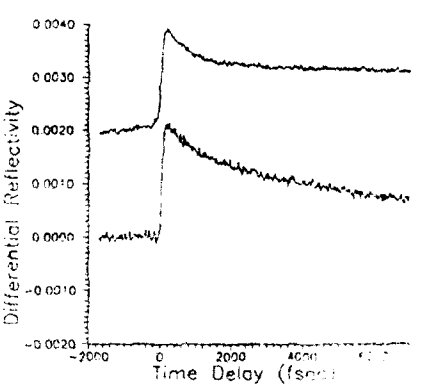


QThD1 Fig. 3. Optical gap energy E_g as a function of temperature for Si at λ = 0.752 μm. (+) 0.2-μm a-Si sample (0) 1.0-μm a-Si sample (O) c-Si.





QThD2 Fig. 1. Femtosecond reflectivity response of the SiGe alloys. Numbers represent the Ge content in each alloy. The reflectivities at time delay zero are displaced for clarity. (a) Relaxed optically thick SiGe alloys freshly grown by MBE. (b) Femtosecond reflectivity response of the strained, optically thin alloys. The 50% alloy is the same sample as in (a) after 6 months.



QThD2 Fig. 2. Femtosecond reflectivity response of the C_{60} films using oblique incident angle probe; (a) crystalline C_{60} film on glass; (b) amorphousized C_{60} film on glass.

Rev. Lett. 65, 764 (1990); F. E. Doany, D. Grischkowsky, Appl. Phys. Lett. 52, 36 (1987).
2. S. Gupta *et al.*, Appl. Phys. Lett. 57, 1343 (1990).
3. M. Strahnen, W. Kutt, H. Kurz, in *Proceedings of the International Conference on VME-bus in Research*, C. Eck, Ed. (North-Holland, Amsterdam, 1988).
4. W. M. Tong *et al.*, J. Phys. Chem. 95, 4709 (1991).
5. S. Saito, A. Oshiyama, Phys. Rev. Lett. 66, 2637 (1991).

QThD3 Diffuse transmission spectroscopy of structure in dense colloids

P. D. Kaplan, A. G. Yodh, D. J. Pine*
Physics Department, University of Pennsylvania, 209 S. 33rd Street, Philadelphia, Pennsylvania 19104

Multiple light scattering spectroscopies are of growing importance to the laser community because of their potential to provide information on dense random media, and because the behavior of diffuse light is interesting in its own right. In this contribution, we show that it is possible to derive quantitative information about the microscopic structure of a dense colloid by measuring its transmission as a function of wavelength.

Photon transport in isotropic random systems, such as dense colloids, is characterized by two quantities: the photon random-walk step length l^* and the photon absorption length. It has been previously recognized that l^* is related to an integral over the material interparticle structure factor, $S(q)$.[1] This fact has been used to correct for the effects of particle structure in diffusion measurements, but generally, the effect is viewed as a nuisance. We demonstrate by experiment and theory that measurements of the *wavelength dependence* of l^* can provide information with high sensitivity on the structure factor of the dense colloid.

To visualize the connection between interparticle structure and photon transport, consider a photon impinging on a small group of particles [Fig. 1(a)]. The probability of scattering through wave vector q is given by the product of the individual particle form factor $F_{k_0a}(q)$ and the interparticle structure factor $S(q)$. The form factor depends on wave vector k_0 as well as particle diameter a. In multiple scattering experiments, this type of scattering occurs many times, and the formal expression for $1/l^*$ is proportional to an integral over $S(q)$ with a weighting function $q^3F_{k_0a}(q)$, i.e.,

$$\frac{1}{l^*} = \frac{\pi\rho}{(k_0a)^6}\int_0^{2k_0a} q^3a^3F_{k_0a}(qa)S(qa)d(qa), \quad (1)$$

where ρ is the particle number density. Diffuse transmission spectroscopy works because the upper cutoff of this integral ($2k_0a$) depends on the photon wavelength. Thus, the value of l^* at every wavelength is an average over *different parts* of the structure factor.

In this experiment, we measure the wavelength-dependent transmission, between 450 and 900 nm, of an optically dense slab of polystyrene spheres (essentially hard spheres) suspended in water. The light is collected [Fig. 1(b)] using an integrating sphere and a photodiode. The data in Fig. 2 were taken from a cell with volume fraction ϕ = 0.33, a = 0.460 μm, and cell thickness L = 500 μm. The transmission coefficient T derived within the diffusion approximation is proportional to l^*,[2] $T = 5l^*/3L = O(l^*/L)$. The solid line in Fig. 2 shows a theoretical prediction for T using the Percus-Yevick approximation to calculate the structure factor and Mie theory for the form factors. The dashed line is calculated with $S(q) = 1$. Clearly, diffuse transmission spectroscopy can test the validity of different models for the material structure. The weak oscillations of l^* with respect to wavelength reflect the periodicity of $S(q)$ and F_{k_0a}. The detailed role of wavelength is discussed further in Fig. 3.

In the future, diffuse transmission spectroscopy will enable us to study a wide range of systems. Although the system studied here is essentially hard sphere, it may be possible to observe the fluid/glass/crystal hard-sphere transitions in similar colloids.

Exxon Research and Engineering, Annandale, New Jersey 08801

1. S. Fraden, G. Maret, Phys. Rev. Lett. 65, 512 (1990); X. Qiu, X. L. Wu, J. Z. Xue, D. J. Pine, D. A. Weitz, P. M. Chaikin, Phys. Rev. Lett. 65, 516 (1990).
2. –. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic Press, New York, 1978). Vol. 1.

QThD4 High-resolution nonlinear magneto-optical spectroscopy in GaAs grown by molecular beam epitaxy

Min Jiang, Hailin Wang, R. Merlin, D. G. Steel
Harrison M. Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan 48109-1120

The presence of strong magnetic fields significantly modifies electronic states of a semiconductor, such as quantization of energy levels (Landau levels) and shrinkage of exciton wave functions.[1] Consequently, significant changes in optical as well as transport properties of the material are expected. In this paper, we report observations of an enhanced nonlinear optical response and a dramatic reduction of mobilities for magnetoexcitons in GaAs. We also discuss nonlinear optical measurements of excitons associated with higher Landau levels.

Our measurements are based on frequency-domain four-wave mixing (FWM),[2] and are carried out as 2.5 K on 0.2-μm thick GaAs film grown by molecular beam epitaxy. At zero field, the sample is homogeneously broadened, and has a 1-*s* exciton absorption line width of 0.2 meV as shown in Fig. 1(a). The light heavy-hole degeneracy is lifted by strain introduced during the etching process. The linear absorption displays a rich structure at 6 T [see Fig. 1(b)]. The features are likely the combined result of strain and impurity-bound excitons. However, the nonlinear optical response [Fig. 1(c)] obtained through cw degenerate FWM exhibits a quite different structure. The energy separation and the magnetic field dependence of the dominant peaks in the nonlinear spectrum suggest that these peaks are due to magnetoexcitons associated with Landau levels, indicating a significant enhancement of the excitonic nonlinear magneto-optical response above the heavy-hole exciton compared to the nonlinear response at 0 T.

The exciton diffusion coefficient is obtained using standard nearly degenerate FWM methods in which grating decay rates are measured as a function of the grating spacing.[3] Fig. 2(a) and (b) display nearly degenerate FWM line shapes obtained at two different field strengths. The width of the response is given by $\Gamma = \gamma = 4\pi^2D/\Lambda^2$, where γ is the excitation decay rate and Λ is the grating spacing. Dramatic field-induced changes in the grating decay rate are evident in Fig. 2(b). Figure 2(c) is a plot of the magnetic dependence of the exciton diffusion coefficient which, at 1 T, is reduced by a factor of 5 compared to the value at zero field. The reduction of the exciton mobility is likely the result of the smaller Bohr radius in the presence of a magnetic field. In this case, exciton scattering by weak or small-scale disorder may become important. Note that the smaller grating decay rates also lead to enhanced nonlinear optical responses for magnetoexcitons.

Nearly degenerate FWM measurements on excitons associated with the second and third Landau levels also reveal interesting phenomena. First, decay of the cw nonlinear response is of the order of the exciton recombination time. Second, measurements using selective optical excitation show that the nonlinear signal arises from excitons associated with the σ^- transition. If the nonlinear optical response is primarily due to phase-space filling and exchange effects, the first observation would imply that the inter-Landau-level transfer rate is slower than or comparable with the exciton recombination rate. If the inter-Landau-level interaction is responsible for the nonlinear optical response of excitons associated with high Landau levels, the second observation would suggest a spin-dependent interexciton interaction. Further studies including magnetoexciton photoluminescence and cw pump–probe measurements, which are helpful in understanding current nonlinear measurements, are also presented.

1. R. J. Elliot, R. Loudon, J. Phys. Chem. Solids 15, 196 (1991).
2. H. Wang, M. Jiang, D. G. Steel, Phys. Rev. Lett. 65, 1225 (1990).
3. J. T. Remillard *et al.*, Opt. Lett. 14, 1131 (1989).

BEST AVAILABLE COPY

Friday MORNING
May 15, 1992 QFB
Convention Center Room C4

8:00 am Ultrafast Electron Dynamics—Surfaces

Tony F. Heinz, *IBM Watson Research Center, Presider*

8:00 am (Invited)

QFB1 Ultrafast electron dynamics and recombination on the Ge(111) 2 × 1 π-bonded surface

Richard Haight, Martina Baeumler
IBM Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598

Elucidation of the microscopic scattering and energy loss processes involved in ultrafast electron dynamics is crucial to the understanding of macroscopic transport in semiconductor systems. In a semiconductor heterostructure system, for example, an electron that crosses the interface between the two materials will scatter from the band minimum of the substrate to a new band in the overlayer, a process that may be complicated by interfacial disorder and interdiffusion. The surface of a semiconductor, in this case the cleaved Ge(111) 2 × 1 surface, provides a particularly interesting model system for the study of such processes. Upon cleavage, one-dimensional chains of π-bonded atoms are formed whose electronic states disperse into the bulk band gap of Ge. The processes that couple bulk excited electrons into states at the surface are operative in a wide class of systems.

We used angle-resolved laser photoemission techniques to examine the ultrafast electron scattering and recombination processes on the Ge(111) π-bonded surface with subpicosecond time resolution. Electrons photoexcited into the bulk Ge conduction band scatter into the unoccupied surface antibonding π* band whose minimum is at the $\bar{J}$ point in the surface Brillouin zone. Rapid relaxation to the surface band minimum is followed by a unique phonon-assisted process in which electrons recombine with bulk holes at the valence band maximum, which we find to be the primary mechanism responsible for the decay of the transient π* population. Time-dependent measurements at 300 K and 120 K have been performed to determine the role of energetic phonons in the scattering processes. These processes are modeled with a set of rate equations, whose fits to the data yield scattering times used to directly determine a surface recombination velocity. Ultrafast surface-state hole dynamics are observed and a renormalization of the surface band gap is studied as a function of electron density. Since the π-bonded states are fundamentally one-dimensional in nature, these results represent the first such studies of bandgap renormalization in a one-dimensional system.

Fraunhofer Institute fur Angewandte Festkörper Physik, Tullastr. 72, D7500, Freiburg, Germany

8:30 am

QFB2 Space-charge-limited thermionic emission from femtosecond-excited metal surfaces

X. Wang, D. M. Riffe, M. C. Downer, J. E. Erskine, R. M. More
Physics Department, University of Texas, Austin, Texas 78712

With the advent of high-intensity femtosecond sources has come the ability to excite, on a subpicosecond time scale, the electrons in a solid to temperature T_e greatly in excess of the lattice vibrational temperature T_i.[1] In a previous experiment on Ag,[2] it was shown that non-equilibrium heating of the electrons leads to subpicosecond thermionic emission, but that space–charge fields distort the initial thermal distribution. Here, we achieve quantitative understanding of the non-equilibrium thermionic emission process by simultaneously measuring the reflectivity of the incident light, the total electron yield, and electron–energy distribution curves (EDCs) of the emitted electrons. The results are modeled with an analytic Richardson-Dushman type equation and with particle simulation codes which include space–charge.

The non-equilibrium excitation is achieved with 10-Hz YAG-pumped, multistage amplification of 2.0-eV colliding-pulse mode-locked (CPM) pulses.[3] Average intensities in each 80-fs pulse are varied up to 4.3 × 10^{12} W/cm^2 (35 mJ/cm^2 fluence) on the sample surface. Total yield is determined by measuring the average current through the sample to ground, whereas the electron-energy distributions are obtained from time-of-flight (TOF) spectra measured with a multichannel plate coupled to a transient digitizer.[4] All data reported here were taken with the laser intensity kept below the cumulative damage threshold.

We find that the temperature-dependence (deduced calorimetrically from the absorption of the metal) of the total yield is in good agreement with a space-charge-limited extension of the Richardson-Dushman equation for thermionic emission. For a square-wave temperature pulse of duration τ_0, the total number N of emitted electrons is given by:[5]

$$N \approx 409\, R_0\, (k_B T_e)\, \ln\{1 + 7.76 \times 10^6\, R_0 \tau_0\, (k_B T_e)\, e^{-\phi_0/k_B T_e}\}$$

Here R_0 is the radius (in microns) of the spot being heated, $k_B T_e$ and ϕ_0 are the temperature and work function, respectively, in electron volts. For reasonably high temperatures (>0.2 eV for a typical metal), the yield is very nearly linear with temperature. Figure 1 has plotted the measured and theoretical Eq. (1) yields for Al (110) using a measured value of R_0 = 158 μm and assumed values of τ_0 = 0.5 ps and ϕ_0 = 4 eV. Note that the yield is not very sensitive to the last two values since they appear inside the log term, and hence it is a simple and direct measure of the non-equilibrium rise T_e. The good agreement between the two curves and the linear behavior of the experimental results for the highest temperatures establish the space–charge dominating behavior to the yield as described by Eq. (1).

Comparison of the total yield with the TOF spectra indicates that, for a given temperature increase, the electron distributions are skewed by the space–charge fields to kinetic energies (KEs) much higher than the corresponding thermal distribution. This is seen in TOF data from Al (110), shown in curves (a) and (c) of Fig. 2. The curve below each data set is a Boltzman distribution at the same temperature T_e as the data, where T_e has been deduced from the simultaneously measured total yield and Eq. (1). In addition to the electrons at much higher KEs, a sharp quantum-excitation peak, most likely due to a surface-related state, at a KE of 1.5 eV is also evident. At present, particle-simulation codes are being used to further understand the measured thermionic distributions.

Physics Department, Lawrence Livermore National Laboratory, Livermore, California 94550

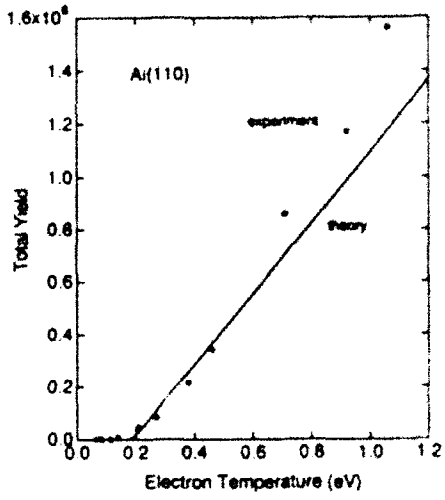
1. H. M. Milchberg, R. R. Freeman, S. C. Davey, R. M. More, Phys. Rev. Lett. 61, 2364 (1988).
2. W. C. Banyai *et al.*, in *Ultrafast Phenomena VII* (Springer-Verlag, Berlin, 1990), p. 116.
3. W. M. Wood, G. Focht, M. C. Downer, Opt. Lett. 13, 984 (1988).
4. D. C. Anacker, J. L. Erskine, Rev. Sci. Instrum. 62, 1246 (1991).
5. D. M. Riffe, X. Y. Wang, M. C. Downer, R. M. More (unpublished).

8:45 am

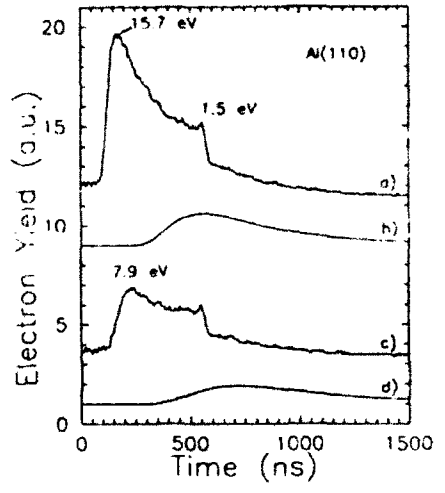
QFB3 Observation of the thermalization of electrons in a metal excited by femtosecond optical pulses

H. W. K. Tom, W. S. Fann, J. Bokor, R. H. Storz
AT&T Bell Laboratories, 4B-429 Crawfords Corner Road, Holmdel, New Jersey 07733

We have measured the evolution of the electron energy distribution excited in a 300-Å Au film by a 180-fs optical pulse. The distribution is measured directly by uv photoemission with a time resolution of 270 fs and an electron energy resolution of ~0.05 eV. Earlier, we reported the first time-resolved photoemission measurement of laser-heated metals with a time resolution of ~700fs. We found that the distribution function, especially around 0.3 eV above the Fermi level, deviated from Fermi-Dirac until ~800 fs after the heating pulse.[1] Here, with better time resolution we are now able to observe (a) evidence of the nascent (as photoexcited) electron distribution, (b) the time evolution from the nascent to a Fermi Dirac distribution, and (c) that thermalization is more rapid for higher exci-



QFB2 Fig. 1. Total yield vs electron temperature for Al(110): • measured yield vs temperature rise T_e deduced from the absorptivity of the metal. For all data shown, the absorptivity is negligibly different from the low-intensity absorptivity of Al. The solid line is calculated from Eq. (1).



QFB2 Fig. 2. Time-of-flight spectra from Al(110) for two different electron temperatures T_e. Curves (a) and (c) are measured TOF spectra corresponding to temperature rises of 0.45 and 0.73 eV, respectively, deduced from simultaneous total yield data and Eq. (1). Curves (b) and (d) are Boltzman TOF distributions, also corresponding to temperature rises of 0.45 and 0.73 eV respectively. Also labeled are kinetic energies corresponding to the various peaks in the TOF spectra.

tattice levels. These results are in marked contrast to experiments in which femtosecond visible reflection spectroscopy[2] was used to deduce the electron-phonon constant coupling[3] g with a model[4] that assumes that the electrons are thermalized much faster than the duration of the optical pulse. These results show that electron thermalization occurs on the same time scale as electron-phonon coupling and that electron-lattice energy transfer is much more complicated than previously thought.

Typical time-resolved photoemission spectra are shown in Fig. 1 for three time delays relative to the arrival of the laser-heating pulse. The energy is referred to E_F, the Fermi energy. The laser heating pulse was generated by amplifying the output of a two-jet prism-compensated synchronously pumped dye laser operating at 1.84 eV. The autocorrelation FWHM was 1.4 × 180 fs. The uv pulse (5.52 eV) was generated by frequency tripling the 1.84 eV pulse in two BBO crystals. The uv pulse FWHM was 270 fs, deduced by observing a 450-fs FWHM cross-correlation with the fundamental. This duration was consistent with the BBO crystal length. The heating and uv pulse were overlapped on a 300-Å Au film kept in a uhv chamber. The photoemitted electrons were detected with a time-of-flight detector with resolution of ~0.05 eV at E_F in the direction normal to the Au surface in a 2° cone angle.

The deviation from Fermi-Dirac is best observed on a logarithmic scale. In Fig. 2 we show the time-evolution of the film for an absorbed fluence of 120 μJ/cm². The best-fit Fermi-Dirac function is shown in dashed line. The data are normalized with respect to the density of states for which we have solved self-consistently. At t = 133 fs, a plateau from 0.7 to 1.8 eV reaches its maximum value. If there were no energy relaxation, the distribution would be the integrated nascent distribution, a flat distribution extending from 0 to 1.84 eV (the pump photon energy), with a value of 0.0028. We can easily see the one-photon cutoff that is a signature of the nascent distribution. The level is within a factor of two of what we'd expect with no energy relaxation. This observation suggests that the energy relaxation time for the 1.8-eV electrons is about one-half the pump pulse duration, e.g., ~100 fs. This value is consistent with energy range experiments[4].

At t = 400 fs, the Fermi-Dirac temperature that fits the data best reaches its maximum, 775 K. The maximum temperature calculated from the heat capacity, assuming no energy transfer to the lattice, is 770 K. There is still approximately 30% of the energy in the hot tail. Because the energy relaxation time should scale as E^{-2}, we expect the 0.9- and 0.7-eV relaxation times to be ~400 and ~670 fs, which are consistent with the data. By 500 fs, the competition between electron thermalization and electron energy transfer to the lattice becomes important. The electron system loses the energy before it can thermalize: even at 1.3 ps there is a nonthermal tail.

In Fig. 3, we show data similar to that in Fig. 2, except that the absorbed fluence was 2.5 × higher, 300 μJ/cm². The plateau from 0.8 to 1.84 eV has the same shape as in the lower fluence case and approximately twice the amplitude consistent with the fluence. The energy relaxation might be faster in this case because there are more electrons above E_F with which to scatter. In this case, only electrons of 0.5 eV have to relax for the distribution to thermalize. By 667 fs the thermalization seems complete. The electron temperature maximum calculated from the heat capacity is 1200 K. The peak temperature reached at 400 fs is 1700 K. Despite what appears to be good agreement with a Fermi-Dirac distribution by 667 fs, the electron temperatures that fit the data do not obey the standard electron-phonon model, e.g., $C_e \partial T_e/\partial t = -\chi(T_e - T_L) = -C_L \partial T_L/\partial t$ until t = 900 fs. This discrepancy is currently being investigated. It may be related to the observation that the best Fermi-Dirac fit is at a temperature much higher than that expected from the heat capacity.

This study shows that thermalization of electrons in a metal occurs relatively slowly. It challenges the standard electron-phonon coupling model. It opens the possibility that ultrashort pulsed excitation on metals produces large numbers of electrons with nascent energy distribution for ~100 fs. It may be crucial in understanding the recent observation of enhanced chemical reaction rates of absorbed molecules on metal surfaces irradiated with ultrashort pulses.

1. W. S. Fann, R. H. Storz, H. W. K. Tom, J. Bokor, "Direct measurement of nonequilibrium electron-energy distributions in sub-picosecond laser-heated gold films," submitted to *Phys. Rev. Lett.*
2. R. W. Schoenlein, W. Z. Lin, J. G. Fujimoto, G. L. Eesley, Phys. Rev. Lett. **58**, 1680 (1987).
3. H. E. Elsayed-Ali, T. B. Norris, M. A. Pessot, G. A. Mourou, Phys. Rev. Lett. **58**, 1212 (1987); S. D. Brorson *et al.*, Phys. Rev. Lett. **64**, 2172 (1990).
4. S. I. Anisimov, B. L. Kapeliovich, T. L. Perelman, Sov. Phys. JETP **39**, 375 (1975).
5. S. M. Sze, J. L. Moll, T. Sugano, Solid-State Electron. **7**, 509 (1964).

9:00 am

QFB4 Heating and hydrodynamic expansion of intensely irradiated metal targets measured by femtosecond time-resolved reflectivity

X. Y. Wang, H. Y Ahn, M. C. Downer
Physics Department, University of Texas, Austin, Texas 78712

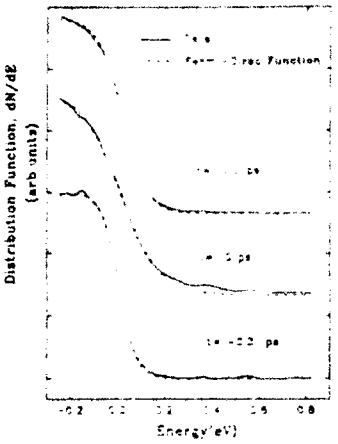
Intense femtosecond excitation of metal targets has attracted fundamental interest in the properties of solid-density plasmas[1-3] as well as practical interest in the generation of ultrafast soft x-ray pulses.[4] In several recent studies,[1,2] the ultrafast response of reflectivity from metal targets was observed with a single intense pulse serving both as target excitation and as optical probe. Consequently, it was difficult to distinguish the competing processes of electron heating and hydrodynamic surface expansion which influence the optical reflectivity to different degrees and at different times.

We report femtosecond time-resolved experiments in which a 90 fs, 620-nm *p*-polarized pump pulse 45° incidence excites Ag, Al, or W surfaces at intensities up to 5 × 10^{13} W/cm². The reflectivity of a weak *s*- or *p*-polarized probe pulse (0° incidence, 90 fs, 620 nm) focused to a spot centered on the pump is then monitored during the pump-probe time delay Δt from −2 ps to +15 ps. The sample is rastered during data collection.

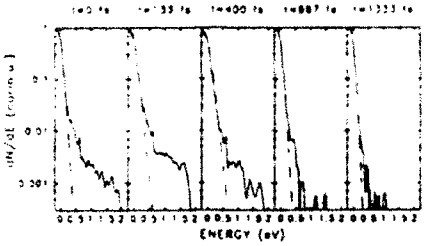
Time-resolved *p*-polarized reflectivity data for the nearly free electron metals Ag and Al excited at 10^{13} W/cm² are shown in Fig. 1(a) and (b), respectively. The reflectivity decreases with pulse-width-limited fall time from an initial value of 0.88 (Ag) or 0.80 (Al), reaching a minimum of about 0.52 in Δt ~ 1 ps, followed by a much slower partial recovery during 1 < Δt < 10 ps. Beyond 10 ps, reflectivity levels off, or decreases slightly as the ablation cloud becomes optically thick.

The reflectivity data are analyzed at each time delay by numerically solving the Helmholtz wave equations.[5] The model includes instantaneous heating and subsequent cooling of the electrons and development of a plasma gradient that expands at constant velocity normal to the metal surface. The evolving material parameters are electron collision frequency ν, plasma frequency ω_p (related by the Drude dielectric function) and expanding plasma gradient scale length L. The dotted curves in Fig. 1(a) and (b) were derived only on the basis of electron heating (L = 0), yielding a pulse-width-limited fall time, as observed. However, an unreasonably long electron relaxation time (~10 ps) is required to fit the recovery. Alternatively, the dashed curves in Fig. 1(a) and (b) show the reflectivity response of Ag and Al, considering plasma expansion alone at a velocity of approximately the speed of sound. But this model alone cannot account for the rapid fall time without invoking an unphysical plasma expanding speed >10^7 cm/s. Nevertheless, when both effects are taken into account, a very satisfying fit for Ag and Al is obtained, as shown by the solid curves in Fig. 1(a) and (b).

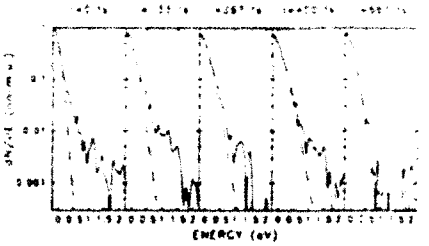
Time-resolved data for the nonfree-electron metal W are shown in Fig. 2. In contrast to Ag and Al, under the same intensity excitation 10^{13} W/cm² [Fig. 2(a)], the reflectivity increases sharply from an initial value of 0.20 to a peak value of ~0.25, indicating conversion to a more reflective metal (higher free-electron density) upon melting. The subsequent reflectivity drop and recovery, however, closely resembles that observed in Ag and Al, and is similarly explained. At lower intensity [5 × 10^{12} W/cm², see Fig. 2(b)], the reflectivity only increases and stays high, indicating melting with minimal subsequent hydrodynamic expansion. In all cases, *s*-polarized reflectivity changes very little, in approximate agreement with the model presented. The calculated curves in Fig. 3 show a different temporal response expected with λ = 310 nm probe for Al. Experiment with a uv probe are being performed to verify the proposed interpretation.



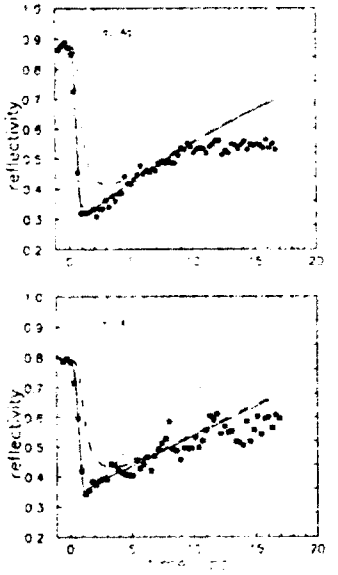
QFB3 Fig. 1. Electron energy distribution function vs energy for laser-excited 300-Å Au film for three time delays. The film was evaporated on a sapphire substrate and was cleaned by annealing in uhv. The absorbed fluence from the laser heating pulse was 120 $\mu J/cm^2$. The spectra were obtained using uv photoemission. The uv fluences were kept low enough to avoid space-charge effects. Each spectrum is the accumulation of 25 000 electrons.



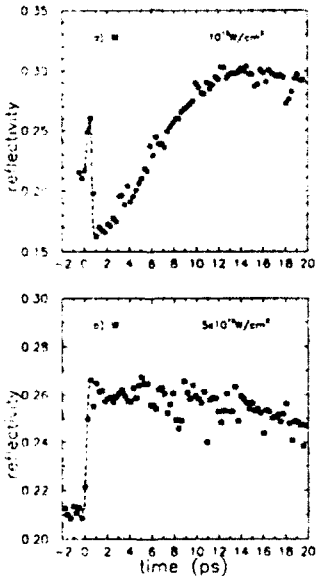
QFB3 Fig. 2. Electron energy distribution function vs energy for Au film with 120-$\mu J/cm^2$ absorbed laser fluence at five time delays. The vertical scale is in units of the density of states. The artifact coming from two-photon ($\omega + 3\omega$) photoemission from levels below E_F can be no larger than one-half the 1.8-eV signal at $t = 0$ fs, e.g., $\lesssim 0.0005$ at $t = 0$ fs and $\lesssim 0.00035$ by $t = 133$ fs. The Fermi-Dirac fits shown in dashed line are fit with $T_e = 516$ K, 562 K, 706 K, 550 K, and 470 K, respectively.



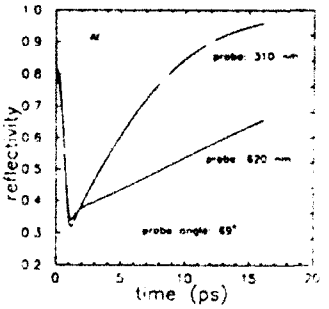
QFB3 Fig. 3 Electron energy distribution function vs energy for Au films with 300-$\mu J/cm^2$ absorbed laser fluence at five time delays. The Fermi-Dirac fits shown in dashed line are fit with $T_e = 753$ K, 1238 K, 1536 K, 1684 K, and 1274 K, respectively.



QFB4 Fig. 1. Time-resolved reflectivity of (a) Ag and (b) Al surfaces pumped by p-polarized, 90-fs, 620-nm pulses incident at 45° from the normal at peak intensity 10^{13} W/cm^2, and probed by s-polarized pulses (data in squares) of same duration and wavelength incident at 69°. Each data point is averaged over 20 shots. Curves represent theoretical models based on numerical solutions of Helmholtz equations, as follows: electron heating included (maximum reflectivity drop shown corresponds to υ/ω from 0.06 to 1.12 for Ag and from 0.35 to 2.41 for Al), plasma expansion neglected (dotted curves); plasma expansion included (with sound speed of 1.17×10^6 cm/s for Ag and 1.5×10^6 cm/s for Al), electron heating neglected (dashed curves); electron heating and plasma expansion both included (solid curves).



QFB4 Fig. 2. Time-resolved reflectivity of W under identical conditions as for Ag and Al at peak intensity (a) 10^{13} W/cm^2 and (b) 5×10^{12} W/cm^2. A numerical fit of reflectivity evolution for W is difficult because the Drude model may not be easily applied with complicated initial transition metal features.



QFB4 Fig. 3. Theoretical curves by extrapolating the parameters $\upsilon(t)$, $\omega_p(t)$, $L(t)$ used to fit the data as in Fig. 1(b) for Al, indicating significantly different temporal response of reflectivity for probes of 620 nm and 310 nm.

1. H. M. Milchberg *et al.*, Phys. Rev. Lett. 61, 2364 (1988); R. Fedosejevs *et al.*, Appl. Phys. B 50, 79 (1990).
2. J. C. Kieffer *et al.*, Phys. Rev. Lett. 62, 760 (1989).
3. H. M. Milchberg, R. R. Freeman, J. Opt. Soc. Am. B 6, 1351 (1989).
4. M. M. Murnane *et al.*, Phys. Rev. Lett. 62, 155 (1989).

9:15 am

QFB5 Generation of high-current-density, subpicosecond electron bunches from Ag film and Cu mirror using amplified CPM laser pulse

Triveni Srinivasan-Rao, T. Tsang, J. Fischer
Instrumentation Division, Brookhaven National Laboratory, 535-B, Upton, New York 11973

Electrons generated by ultrashort laser pulses have proven to be excellent sources for high-brightness electron guns,[1] much sought after for short-wavelength FEL, and high-energy accelerators. The brightness of the e-beam can be increased by increasing the current and decreasing the emittance. We explore the feasibility of increasing the current and the current density, using photoemission with subpicosecond laser pulses on metal cathodes.

The laser system used in these studies is a colliding pulse mode-locked (CPM) dye laser followed by a nitrogen-laser-pumped dye amplifier generating up to 3 μJ energy with a pulse duration of 280 fs at 2-eV photon energy. This beam is focused onto the cathode to a spot size of 60 × 90 μm² FWHM. The cathode is either a copper mirror or a thin Ag film, held parallel to the anode, at a pressure of $\sim 10^{-5}$ Torr. The electrons leaving the cathode are measured using a charge-sensitive preamplifier and a shaping amplifier, and are displayed on an oscilloscope.

The intensity dependence of the current density from copper upon irradiation by the amplified CPM is shown in Fig. 1 (+). Since the work function of copper is ~4.3 eV, multiphoton process involving at least two photons is expected. However, contrary to this expectation, the slope of the line through the data is 1, implying a single-photon process. The maximum current density delivered by the cathode exceeds 10 kA/cm², with a quantum efficiency of 2×10^{-4}. No surface damage has been observed on the sample up to laser intensities of 2×10^{11} W/cm².

Intensity dependence of photoemission from Ag film for both the unamplified and the amplified CPM are also shown in Fig. 1. In keeping with the energy required to overcome the work function of silver, two-photon photoemission is observed in both the intensity regimes, and the consistency between the data in the two regimes is very good. The maximum current density observed with this sample is 3 kA/cm². No surface damage to the Ag film was observed after irradiating with laser intensities of $\sim 10^{12}$ W/cm². The autocorrelation signal of the amplified laser beam and the e-beam are shown in Fig. 2. The electron pulse appears to be 25% longer than the laser pulse. This discrepancy could be attributed either to the nonuniform heating of the electrons or the low signal-to-noise ratio and fluctuation in the electron signal, and is currently under investigation. In Fig. 3, the charge, Q is plotted against the extraction field, $\sqrt{E}$, to illustrate the magnitude of the Schottky effect on photoemission from a smooth film for various input energies. Theoretically, for a two-photon process, the slope of the line should be proportional to $A \cdot I$ where A contains material constants and I is the intensity of the laser. The data indicate that the slope changes as $\sqrt{I}$. This discrepancy could be due to the dependence of the material constants on the electron and lattice temperature and hence the laser intensity.

In conclusion, subpicosecond electron bunches of current densities exceeding 10 kA/cm² has been generated. Since no optical damage was observed so far, extracting higher current densities are feasible, if larger extraction fields are available. Scaling to larger currents is also possible by using more energetic laser beams of comparable intensities.

1. H. G. Kirk *et al.*, "Experimental results from the BNL ATF photocathode gun," to be published in *Proc. Thirteenth International Free-Electron Laser Conference*, Santa Fe, New Mexico, Aug. 25–30, 1991.

9:30 am Invited

QFB6 Electronic disordering during femtosecond laser melting of GaAs

P. N. Saeta,* J.-K. Wang, Y. Siegal, N. Bloembergen, E. Mazur
Gordon McKay Laboratory, Harvard University, Cambridge, Massachusetts 02138

An intense femtosecond pulse of visible light incident on the surface of a crystalline semiconductor can deposit energy in the electronic system in a time that is short compared to the electron–phonon coupling time. If the density of deposited energy is high enough, a layer of material at the surface may be able to disorder before appreciable energy is transferred from the electrons to the ions of the lattice.[1] Pump–probe experiments on silicon using 90-fs pulses showed a rise in reflectivity and a decay in second-harmonic generation (SHG) consistent with the melting of the surface layer after the excitation pulse.[2–4] Because the intensity of the reflected second-harmonic signal depends on crystallographic orientation, its decay serves as a sensitive probe of structural changes. GaAs offers the advantage offers SI that SHG is dipole-allowed in the bulk, so that the surface contribution is negligible compared to the strong bulk signal.[5]

Pump–probe experiments were performed with 160-fs pulses (FWHM) on a semi-insulating GaAs(110) wafer in air. The probe was s-polarized and incident at 45° from the normally incident pump beam. The origin of time was determined from the sum-frequency signal between the pump and probe beams. The sample was translated in plane between laser shots, since the excitation fluence exceeds the damage threshold.

Figure 1 shows the fluence dependence of the detected reflectivity and second-harmonic signals at a probe delay of 120 fs, normalized to their values without excitation. At a threshold of 0.1 J/cm² the second-harmonic intensity begins to drop, by 0.2 J/cm² it vanishes even at this short delay. Above the threshold of 0.1 J/cm² the reflectivity rises, tending toward a saturated value 44% above the unexcited crystalline value. In Fig. 2, the same signals are shown as a function of probe delay for an incident fluence of 0.33 J/cm². The curves through the data points are exponential fits giving response times of 90 and 170 fs for the SHG decay and reflectivity rise, respectively.

The high reflectivity that obtains 500 fs after excitation was compared to a Drude model of molten GaAs. The observed 44% rise is consistent with 50% ionization of valence electrons in the excited phase. This is more than an order of magnitude greater than the density of photoexcited electrons, 8×10^{20} cm⁻³, based on linear absorption. This strongly suggest an electronic phase transformation. The sharp decay of SHG with a 90-fs time constant indicates that the excited electronic state is a centrosymmetric one, in which SHG is dipole forbidden. Since the time scale for energy relaxation from electrons to phonons in the crystal is ≈ 2 ps,[6] the electronic system—and likely the atomic lattice—disorders before appreciable energy is transferred to lattice vibrations.

*Current address: *AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974*

1. J. A. Van Vechten, R. Tsu, F. W. Saris, Phys. Lett. A **74**, 422 (1979).
2. C. V. Shank, R. Yen, C. Hirlimann, Phys. Rev. Lett. **50**, 454 (1983).
3. C. V. Shank, R. Yen, C. Hirlimann, Phys. Rev. Lett. **51**, 900 (1983).
4. H. W. K. Tom, G. D. Aumiller, C. H. Brito-Cruz, Phys. Rev. Lett. **60**, 1438 (1988).
5. P. Saeta, J.-K. Wang, Y. Siegal, N. Bloembergen, E. Mazur, Phys. Rev. Lett. **67**, 1023 (1991).
6. J. A. Kash, J. C. Tsang, J. M. Hvam, Phys. Rev. Lett. **54**, 2151 (1985).

Effect of barrier thickness asymmetries on the electrical characteristics of AlAs/GaAs double barrier resonant tunneling diodes

A. J. Tsao, V. K. Reddy, D. R. Miller, K. K. Gullapalli, and D. P. Neikirk
Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712

(Received 16 September 1991; accepted 30 October 1991)

We report the impact of small barrier thickness asymmetry on the dc I–V characteristics of AlAs/GaAs double barrier resonant tunneling diodes. With a bottom AlAs barrier 6 ML thick and GaAs well 18 ML thick, the effects of varying the top AlAs barrier thickness from 5 to 8 ML produced significant changes in peak current density, peak voltage, and peak-to-valley current ratio (PVCR). PVCRs of 5.6 were obtained on a 7/18/6 monolayer structure, the highest reported to date for an AlAs/GaAs DBRTD structure.

I. INTRODUCTION

For high performance double barrier resonant tunneling diodes (DBRTDs), precise control of the quantum well thickness and the barrier and spacer layer thicknesses is critical. Minor variations in any of these layer thicknesses can have tremendous impact on the resulting dc I–V parameters, such as the peak-to-valley current ratio (PVCR), the peak current density (J_p), the difference between the peak voltage and the valley voltage (ΔV), and the difference between the peak current density and the valley current density (ΔJ).

Previous studies of barrier thickness asymmetries in DBRTDs have dealt with thick AlGaAs barriers which varied by many monolayers.[1-3] Asymmetries in dc current–voltage characteristics have also been observed in nominally symmetric structures, where the cause has been attributed to interface roughness from the inverted interface, GaAs on AlGaAs or GaAs on AlAs.[4,5] Here we present the effects of intentionally grown barrier thickness asymmetries on thin barrier, high current density AlAs/GaAs DBRTDs. Experimental measurements are also compared to simulation results using a self-consistent calculation that couples the Schrödinger equation in the quantum well region to the drift-diffusion equation in the rest of the device.[6]

II. EXPERIMENTAL PROCEDURE

The DBRTDs were grown in a Varian Gen II molecular beam epitaxy (MBE) system using n^+ (2×10^{18} cm^{-3}) silicon-doped (100) liquid encapsulated Czochralski (LEC) GaAs substrates. On all structures, the substrate temperature was kept at 600 °C with a 0.5–1.0 μm n^+ (4×10^{18} cm^{-3}) GaAs buffer layer grown at a rate of 1 ML/s. Three spacer layers surround the quantum well, consisting of 50 Å of nominally undoped GaAs, 100 Å of 5×10^{16} cm^{-3} n-type GaAs, and 100 Å of 6×10^{17} cm^{-3} n-type GaAs. A n^+ (4×10^{18} cm^{-3}) GaAs cap layer was grown on all samples to facilitate ohmic contact formation. All growth rates were calibrated using reflection high-energy electron diffraction (RHEED) intensity oscillations, immediately before and after actual growth of the DBRTD devices. The AlAs growth rates used were 0.25 or 0.30 ML/s, while the growth rate for the GaAs spacer layers and GaAs well was 0.4 ML/s. Growth was interrupted for 4 s at each heterointerface. For all devices, the bottom AlAs barrier and GaAs well growth times were set to produce nominal layer thicknesses of 6 and 18 ML, respectively. The AlAs top barrier growth time was altered for the different DBRTDs to produce nominal layer thicknesses of 5, 5.5, 6, 6.5, 7, and 8 ML.

The device structures were mesa isolated using an 8:1:1 H_2SO_4:H_2O_2:H_2O etch. The top Ni/AuGe/Ni contact metallization was annealed at 450 °C for 30 s. The backside ohmic contacts were formed by the indium metallization used to mount the GaAs substrates to the molybdenum blocks during MBE growth.

The devices were tested using a Keithley 230 programmable voltage source and a Keithley 195A digital multimeter controlled by an IBM PC-AT. The devices ranged from 3 to 7 μm in diameter.

III. RESULTS AND DISCUSSION

Experimental and simulation results are summarized in Figs. 1 and 2. In our simulations, an AlAs effective mass of $0.11m_e$ provided the best fit to the experiments. The change from the band edge value of $0.15m_e$ can be attributed to strong renormalization of the Γ electron effective mass in the AlAs.[7] Figure 1 shows that for biases ("reverse" bias) such that electrons are incident on the top AlAs barrier first, the peak current density drops dramatically as this width is varied from 5 to 8 ML. For electron injection through the fixed-thickness bottom barrier first ("forward" bias), the variation is significantly less. Figure 2 shows similar behavior for the peak voltage in forward and reverse bias. Both our simulation and experimental results show that thickness variations in the entry barrier have a much larger impact on dc I–V characteristics than thickness variations in the exit barrier.

Figure 3 shows that structural asymmetries also have a significant impact on peak-to-valley current ratio. Signifi-

 0734-211X/92/021042-03$01.00 ©1992 American Vacuum Society

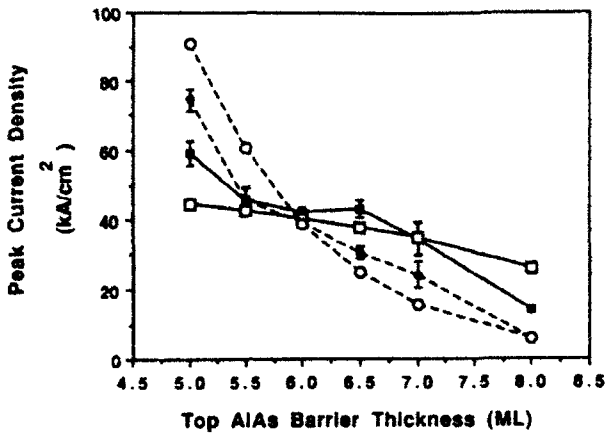


FIG. 1. Peak current density vs top AlAs barrier thickness for both simulation and experimental results. (- -●- -): Electron injection through variable thickness top AlAs barrier first; (- -○- -): Simulation results for electron injection through variable thickness top AlAs barrier first; (—■—): Electron injection through 6 ML bottom AlAs barrier first; (—□—): Simulation results for electron injection through 6 ML bottom AlAs barrier first. One-sigma error bars are included for experimental data.

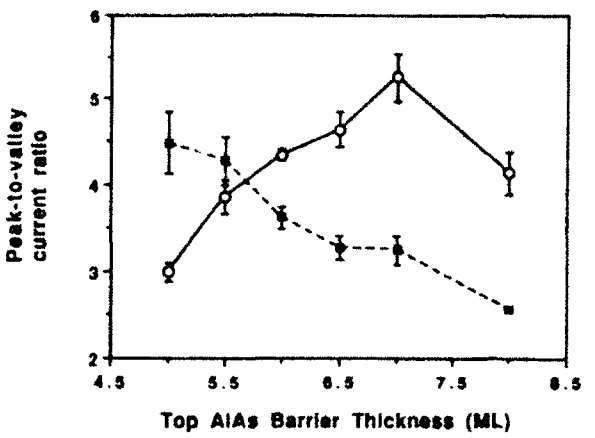


FIG. 3. Peak to valley current ratio vs top AlAs barrier thickness for experimental results. (—○—): Electron injection through variable thickness top AlAs barrier first; (- -■- -): Electron injection through 6 ML bottom AlAs barrier first. One-sigma error bars are included for this experimental data.

cantly higher PVCRs were observed when the electrons passed through the thicker barrier first, regardless of whether the thicker barrier is on the top or bottom of the GaAs well. As the top AlAs barrier thickness was varied from 5 to 7 ML, an increase in the PVCR in the reverse bias case (electron injection through the top AlAs first) was seen. This trend did not continue, however, as the top Alas barrier thickness was increased to 8 ML. It is believed that as the thickness increased from 7 to 8 ML Γ_{GaAs}–X_{AlAs} tunneling effects may become significant, which tends to degrade the PVCR. Other efforts to improve PVCR for AlGaAs/AlAs/GaAs DBRTDs have also used asymmetric barriers, producing PVCRs of 6.3.[4,8] Our 7/18/6 monolayer DBRTD demonstrated a PVCR of 5.6 in reverse bias, which is the highest reported to date for a simple AlAs/GaAs DBRTD.

The amount of dc I–V asymmetry in each of the devices is summarized by taking the ratios of the forward and reverse bias data, as shown in Fig.4. Interestingly, the device that exhibited the most symmetric characteristics was not the nominally 6/18/6 monolayer DBRTD, but rather the 5.5/18/6 monolayer DBRTD. This indicates that there may be a MBE growth asymmetry of unknown origin which causes the AlAs barrier grown after the GaAs quantum well to be approximately half a monolayer thicker than the barrier grown before the well.

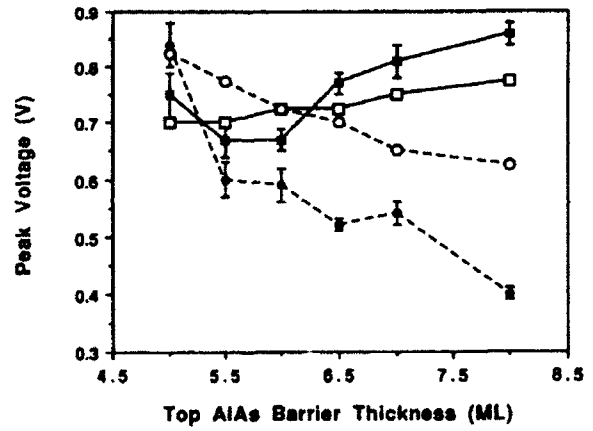


FIG. 2. Peak voltage vs top AlAs barrier thickness for both simulation and experimental results. (- -●- -): Electron injection through variable thickness top AlAs barrier first; (- -○- -): Simulation results for electron injection through variable thickness top AlAs barrier first; (—■—): Electron injection through 6 ML bottom AlAs barrier first; (—□—): Simulation results for electron injection through 6 ML bottom AlAs barrier first. One-sigma error bars are included for experimental data.

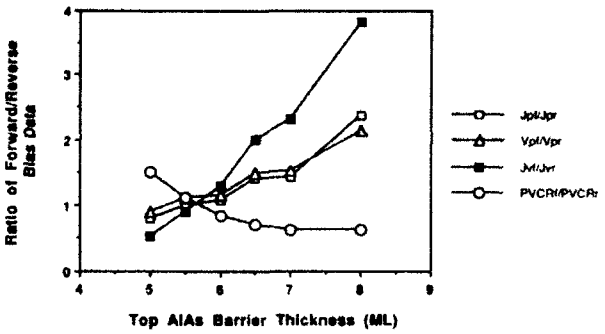


FIG. 4. Ratio of the forward and reverse bias dc I–V parameters vs the top AlAs barrier thickness. This graph displays the asymmetry in dc I–V parameters as top AlAs barrier thickness is varied. The device with a 5.5 ML top AlAs barrier is shown to be most symmetric. (—□—): Ratio of the forward and reverse bias peak current densities. (—■—): Ratio of the forward and reverse bias valley current densities. (—○—): Ratio of the forward and reverse peak-to-valley ratios. (- -△- -): Ratio of the forward and reverse bias peak voltages.

IV. CONCLUSIONS

We have reported the impact of small barrier thickness asymmetries on the dc I–V characteristics of AlAs/GaAs DBRTDs. By carefully monitoring growth rates before and after each device structure growth run, DBRTD structures were grown with nominal top AlAs barrier thicknesses of 5, 5.5, 6, 6.5, 7, and 8 ML. It was seen that the PVCR is significantly affected by the amount of asymmetry in barrier thickness, with higher PVCRs occurring when the electron moves through the thicker barrier first, regardless of whether the thicker barrier was on the top or the bottom of the structure. For the 7/18/6 DBRTD, a PVCR of 5.6 was measured, which is the highest reported to the best of our knowledge. For peak current densities and voltages, simulation results are in qualitative agreement with our experimental results. Simulation and experimental results show that thickness variations in the entry barrier have a much larger impact on dc I–V characteristics than thickness variations in the exit barrier.

ACKNOWLEDGMENTS

This work was sponsored by the Joint Services Electronics Program under Grant No. AFOSR 49620-89-C-0044, the Texas Advanced Technology Program, and the Texas Advanced Research Program.

[1] E. S. Alves, L. Eaves, M. Henini, O. H. Hughes, M. L. Leadbeater, F. W. Sheard, and G. A. Toombs, Electron. Lett. **24**, 1190, (1988).
[2] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Solid State Electron. **31**, 731, (1988).
[3] C. Rosel, P. Gueret, and H. P. Meier, J. Appl. Phys. **67**, 900, (1990).
[4] T. J. Shewchuk, P. C. Chapin, P. D. Coleman, W. Kopp, R. Fischer, and H. Morkoc, Appl. Phys. Lett. **46**, 508, (1985).
[5] P. Cheng and J. S. Harris, Jr., Appl. Phys. Lett. **56**, 1676 (1990).
[6] D. R. Miller, PhD dissertation, The University of Texas at Austin, 1992.
[7] G. Brozak, E. A. de Andrada e Silva, L. J. Sham, F. DeRosa, P. Miceli, S. A. Schwarz, J. P. Harbison, L. T. Florez, and S. J. Allen, Jr., Phys. Rev. Lett. **64**, 471, (1990).
[8] V. K. Reddy, A. J. Tsao, and D. P. Neikirk, Electron. Lett. **26**, 1742, (1990).

Influence of growth interruption on *I–V* characteristics of AlAs/GaAs double barrier resonant tunneling diodes

V. K. Reddy and D. P. Neikirk
Microelectronics Research Center, Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, Texas 78712

(Received 16 September 1991; accepted 29 October 1991)

We have examined the influence of growth interruption on the electrical characteristics of thin barrier, high-current density AlAs/GaAs resonant tunneling diodes (RTDs) grown by molecular-beam epitaxy. Interrupt schedules were determined by independent reflection high-energy electron diffraction measurements of specular spot intensity oscillations during prototypical device growth sequences. Our data suggests that interface roughness at inverted and normal interfaces does not play a significant role in determining the perpendicular transport characteristics of high current density AlAs/GaAs RTDs.

I. INTRODUCTION

The double barrier resonant tunneling diode (DBRTD) has been investigated extensively for device applications such as microwave oscillators.[1,2] To improve the performance of these molecular-beam epitaxy (MBE) grown devices both structural and growth parameters need to be optimized. One such growth parameter is the interrupt time used at each heterointerface in the device.

Investigations of the effect of growth interruption during MBE have focused mainly on the correlation between interrupt time and quantum well photoluminescence linewidths[3] and reflection high-energy electron diffraction (RHEED) specular spot intensity oscillations.[4] Interrupts of an appropriate duration during quantum well growth have been shown to result in luminescence spectra indicative of atomically smooth and abrupt interfaces. The effect of growth interruptions on the low temperature electrical characteristics of thick barrier, very low current density AlGaAs/GaAs DBRTDs has been previously investigated by Gueret *et al.*[5] The optimization of DBRTDs for oscillator applications, however, requires the study of high current density devices. Here, we examine the influence of growth interruption on the current–voltage (*I–V*) characteristics of thin barrier, high current density AlAs/GaAs DBRTDs.

II. EXPERIMENT

The interrupt durations used for the DBRTDs were based on the results of independent RHEED experiments. These consisted of measurements of the specular spot intensity oscillations during prototypical device growth sequences consisting of 6 monolayers (ML) AlAs, interrupt, 18 ML GaAs, interrupt, and finally 6 ML of AlAs. The interrupt time was varied from 0 to 60 s. The GaAs and AlAs growth rates were 0.4 and 0.3 ML/s, respectively. The As/Ga beam equivalent pressure ratio was 30 and As/Ga incorporation ratio was 3.1 at a substrate temperature of 600 °C.

For electrical tests, four structurally symmetric DBRTDs were grown at 600 °C with different growth interruption schedules. The growth conditions were nominally the same as those employed in the RHEED measurements. The layer schematic of the structure along with the interrupt schedule are given in Fig. 1. Sample A is grown without any interruption. For samples B and C, growth interruptions of 20 and 60 s, respectively, were employed only after the GaAs layers, while no interruptions followed the growth of the AlAs layers. For sample D, growth interruptions of 60 s only on the AlAs layers were used, with no interruptions on the GaAs layers. All samples, grown on (100) n^+ GaAs substrates, consisted of a nominally undoped (n type, 1×10^{15} cm^{-3}) 18 ML GaAs quantum well sandwiched between 6 ML AlAs barriers, followed by a three-step dopant transition region consisting of 50 Å of nominally undoped GaAs adjacent to the AlAs barriers, 100 Å of n type, 6×10^{16} cm^{-3} GaAs, and finally 100 Å of n type, 4.3×10^{17} cm^{-3} GaAs. Mesa isolated diodes were formed with Ni/AuGe/Ni/Au and backside indium ohmic contacts that were annealed at 450 °C for 30 s. The *I–V* characteristics were obtained for 10–15 devices per sample with a computerized data acquisition system.

III. RESULTS AND DISCUSSION

The results of the independent RHEED measurements are shown in Fig. 2. This figure illustrates the evolution of intensity oscillations as the interrupt time between layers was varied from 0 to 60 s. In this figure, the interrupt time scales have been compressed to allow easy visual comparison of the relative oscillation amplitudes of the three layers. In all cases, RHEED oscillations were maintained throughout the entire growth sequence. Also, very strong oscillations were observed during the growth of the first 6 ML AlAs layer. However, interruptions on AlAs did not result in specular spot intensity recovery with the intensity remaining essentially constant until the GaAs quantum well growth was initiated. Furthermore, as the growth interruption was increased on AlAs, the GaAs grown on top of it became rougher as evidenced by a reduced oscillation amplitude during the GaAs quantum well growth. This may indicate that the very thin 6 ML AlAs layer was fairly smooth, and so growth interruption was not beneficial.

In contrast to the behavior of the AlAs layer, specular

 0734-211X/92/021045-03$01.00 ©1992 American Vacuum Society

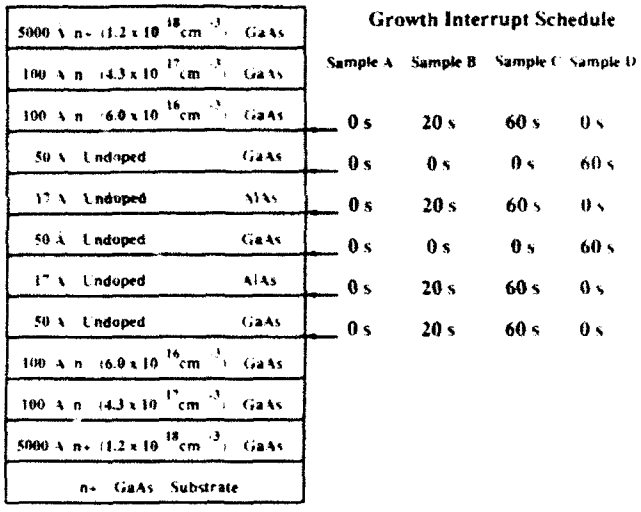


FIG. 1. Layer schematic of the DBRTD structure along with the interrupt schedule for the four DBRTDs studied.

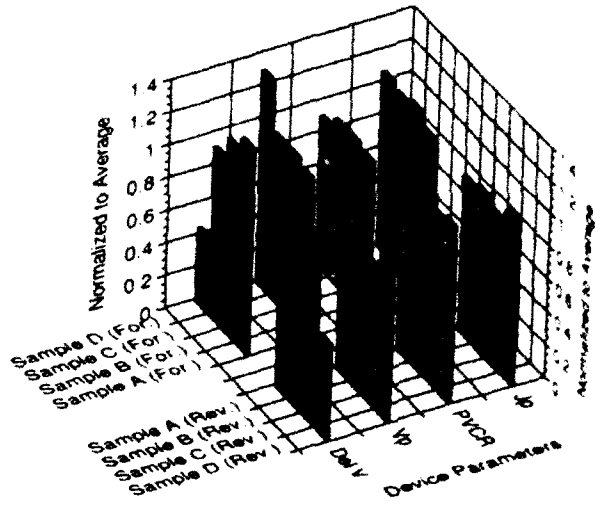


FIG. 3. Summary of measured I–V data for samples A, B, C, and D. The values for J_p, PVCR, V_p, and ΔV are normalized to their respective averages based on all samples and both bias directions. The average values for J_p, PVCR, V_p, and ΔV were 43 kA/cm^2, 4.4, 0.83 V, and 0.22 V, respectively. The standard deviations for all quantities presented were less than 5%.

spot intensity recovery was observed for interrupts on the 18 ML thick GaAs layer. If no interruption was used after the GaAs growth, the RHEED oscillations seen during the growth of the final AlAs layer were much worse than those seen during the growth of the first AlAs layer. However, a 60 s interruption on the GaAs layer was sufficient to produce oscillations from the final AlAs layer with amplitude comparable to that of the first AlAs layer. Since the interrupt schedule affects the interface quality, it is possible it may affect the device I–V characteristics. From the data shown in Fig. 2, it might be expected that a DBRTD device grown using interruptions only after the growth of GaAs layers, and not after AlAs barrier layers, should improve the I–V characteristics over a device grown with no interruptions. Further, a device grown with long interruptions (60 s) after AlAs layers, but not after the GaAs layers, might result in poorer characteristics compared to one grown with no interruptions.

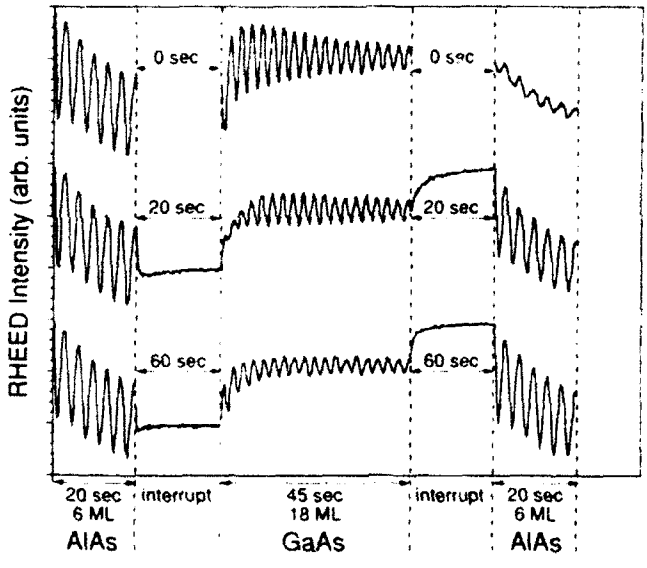


FIG. 2. Evolution of RHEED specular spot intensity oscillations during prototypical growth sequences consisting of 6 ML AlAs/interrupt/18 ML GaAs/interrupt/6 ML AlAs. The interrupt time was varied from 0 to 60 s. The interrupt time scales have been compressed to allow easy comparison of the relative oscillation amplitudes of the three layers.

Measured I–V data for samples A, B, C, and D are summarized in Fig. 3, which shows normalized values for peak current density (J_p), peak-to-valley current ratio (PVCR), peak voltage (V_p), and difference between peak and valley voltage (ΔV) for both electron injection through the bottom barrier first (forward bias) and for injection through the top barrier first (reverse bias). The quantities are scaled to their respective averages based on all samples and both bias directions. The average values for J_p, PVCR, V_p, and ΔV are 43 kA/cm^2, 4.4, 0.83 V, and 0.22 V, respectively.

For both bias directions, J_p and PVCR are independent of both interrupt time and schedule. The V_p and ΔV appear to be sample dependent in that a higher V_p and smaller ΔV is obtained in sample D (60 s growth interruption on AlAs only). Although this may be due to the interrupt schedule, it could also be due to parasitic series resistance. In fact, a similar structure grown during a previous system cycle exhibited characteristics similar to samples A, B, and C. Surprisingly, the 60 s growth interruption on GaAs did not improve the PVCR and J_p, even though the RHEED measurements indicated that this interrupt did cause significant interface smoothing. Growth interruption on the AlAs seems to have had little impact on I–V characteristics. These observations indicate that any interface roughness present in our thin barrier, high current density AlAs/GaAs DBRTDs does not play a significant role in determining perpendicular transport characteristics.

All the DBRTD samples studied here, regardless of in-

terrupt schedule, display a slight asymmetry in both PVCR and J_p between forward and reverse bias. Electron injection from top to bottom (i.e., reverse bias) results in higher PVCR and lower J_p, whereas for electron injection from bottom to top (forward bias), the PVCR is lower and the J_p is higher. If the PVCR is taken to be a measure of the interface roughness[6,7] then our data suggests that the role of the inverted (GaAs on AlAs) and normal (AlAs on GaAs) interfaces in our context is not very significant. In fact, all of the samples studied here exhibit a higher PVCR for electron transport from top to bottom, where electrons encounter a nominally rougher AlAs interface first, rather than bottom to top, where electrons encounter a nominally smoother AlAs interface first. For our devices, we believe the observed asymmetry is due to a growth-induced thickness asymmetry, which yields a top AlAs barrier that is slightly thicker than the bottom AlAs barrier.[8]

IV. CONCLUSION

We have examined the influence of growth interruption on the electrical characteristics of thin barrier, high-current density AlAs/GaAs DBRTDs grown by MBE. Interrupt schedules were determined by independent RHEED measurements of the intensity oscillations during prototypical device growth sequences. Four nominally symmetric DBRTDs with different interrupt sequences were fabricated and electrically characterized. Our data suggests that interface roughness at inverted and normal interfaces does not play a significant role in determining the perpendicular transport characteristics of high current density AlAs/GaAs DBRTDs.

ACKNOWLEDGMENTS

The authors thank K. K. Gullapalli, D. R. Miller, and A. J. Tsao for useful discussions. This work was sponsored by the Joint Services Electronics Program under Grant No. AFOSR 49620-89-C-0044 and the Texas Advanced Technology Program.

[1] V. P. Kesan, A. Mortazawi, D. R. Miller, V. K. Reddy, D. P. Neikirk, and T. Itoh, IEEE Trans. Micro. Theory Tech. **37**, 1933 (1989).
[2] V. K. Reddy, A. J. Tsao, and D. P. Neikirk, Electron. Lett. **26**, 1742 (1990).
[3] M. A. Herman, D. Bimberg, and J. Christen, J. Appl. Phys. **70**, R1 (1991).
[4] A. Madhukar, T. C. Lee, M. Y. Yen, P. Chen, J. Y. Kim, S. V. Ghaisas, and P. G. Newman, Appl. Phys. Lett. **46**, 1148 (1985).
[5] P. Gueret and C. Rossel, in *Proceedings of NATO Advanced Study Workshop on Resonant Tunneling in Semiconductors. Physics and Applications* (Plenum, New York, 1990).
[6] H. Riechert, D. Bernklau, J. P. Reithmaier, and R. D. Schnell, J. Cryst. Growth **111**, 1100 (1991).
[7] H. C. Liu and D. D. Coon, J. Appl. Phys. **64**, 6785 (1988).
[8] A. J. Tsao, V. K. Reddy, D. R. Miller, K. K. Gullapalli, and D. P. Neikirk, Proceedings of the 11th Molecular-Beam Epitaxy Workshop, Austin, Texas, 1991 (unpublished).

Photoluminescence study of the effects of growth interruption on integer and fractional monolayer AlGaAs/GaAs quantum wells

T. R. Block,[a] D. P. Neikirk, and B. G. Streetman
Microelectronics Research Center, Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712

(Received 16 September 1991; accepted 9 October 1991)

A set of quantum wells containing wells of integer n and integer plus half-monolayer $n + 1/2$ thickness were grown. The photoluminescence of these samples contained distinct peaks for both the n and $n + 1/2$ wells. The structure was regrown with growth interruption between layers. With sufficiently long interrupt times (~ 60 s) the peaks due to the $n + 1/2$ wells disappeared leaving only peaks from wells of integer monolayer thickness. This result can be interpreted in terms of structural changes at the interfaces such as island coalescence. The intensity of the peaks and shifts in their energy also provide a measure of contamination and microroughness.

I. INTRODUCTION

Photoluminescence (PL) is an extremely powerful tool for the analysis of GaAs/AlGaAs quantum wells (QWs). Due to the confinement of carriers the PL transition energy is shifted to higher energies as the well thickness is decreased.[1] The linewidth of the PL transition from the well is sensitive to the structure of the interfaces between the layers[2-4] and the intensity of the transition is related to the material quality.[5,6] This sensitivity to layer structure and quality can be used to study the growth processes which occur during molecular beam epitaxy (MBE).

A common growth technique is to interrupt growth between layers during the growth of QWs. Migration of the adatoms on the surface changes the size and distribution of islands present and is reflected in the PL transition energy and linewidth. Often the growth of a single QW with interruption will give rise to more than one peak due to the formation of wells of different thickness distributed laterally across the sample. These multiple peaks can be interpreted in terms of a simple model in which the exciton samples the potential in its local environment.[4,7] This behavior has been related to the recovery of the reflection high-energy electron diffraction (RHEED) intensity upon growth interruption.[8,9]

II. EXPERIMENT

In this article we propose an additional technique involving growth interruption and PL of QWs. Shown in Fig. 1 is the PL spectrum of a set of GaAs QWs grown with wells of integer n monolayer thickness. In this sample wells of 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 19, 25, and 35 ML were grown separated by 300 Å $Al_{0.3}Ga_{0.7}As$ barriers. The transitions from this sample give a series of sharp and well separated peaks, one for each well thickness. This suggested the idea of growing wells of nominal thickness equal to $n + 1/2$ ML. A sample was grown in which the shutters were timed to give nominal well thicknesses of 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, and 11 ML separated by 300 Å barriers. The PL of this sample is shown in Fig. 2 where 17 distinct peaks can be seen, again one for each of the well thicknesses grown. This same structure was also grown with growth interruptions of 10, 30, 60, and 120 s between all layers. The PL of the samples grown with 10 and 30 s interrupts are shown in Fig. 3. In this figure the peaks are beginning to blur together such that the number of distinct peaks is decreased. This trend is continued for the case of 60 s interrupts shown as the upper spectrum of Fig. 4. In this spectrum the half-monolayer transitions have disappeared leaving nine relatively distinct transition which corresponds to the number of integer monolayer wells grown. The spectrum for interrupts of 120 s is shown as the bottom spectrum in Fig. 4 where again distinct transitions are seen, but surprisingly there are even fewer peaks present.

All samples were grown in a Varian Gen II at 620 °C with an As cracking source.[10] The Ga growth rate was set to 0.7 ML/s and the Al growth rate was set to 0.3 ML/s. The As/Ga incorporation ratio was kept at approximately 1.4. All PL spectra were taken at 4 K with the samples immersed in liquid He. The samples were excited with the 514.5 nm line of an Ar ion laser at a power density of 100 mW/cm^2. The luminescence was passed through a SPEX 1 m, single-grating spectrometer and detected with an RCA C31034 GaAs photocathode detector.

III. DISCUSSION

The most prominent trend to be seen in Figs. 2–4 is the decrease in the number of well transitions as the interrupt time increases. Without interruption 17 peaks are distinguishable, and this reduces to nine peaks after 60 s interrupts. This is apparently due to a coalescence of the nominal half-monolayer wells into wells of monolayer thickness. Thus, for example, one can imagine the nominal 3.5 ML well as possessing closely spaced regions of 3 and 4 ML thicknesses. This results in a transition energy intermediate between that of a 3 and a 4 ML well, which disappears as the regions of 3 and 4 ML become larger than the exciton. A potential problem for this explanation is the apparent absence of some peaks from the sample grown with 120 s interrupts. However, we believe this is due to the degradation of the sample from the long interrupt time.

Degradation with interrupt time is indicated by the in-

 0734-211X/92/020832-03$01.00 ©1992 American Vacuum Society

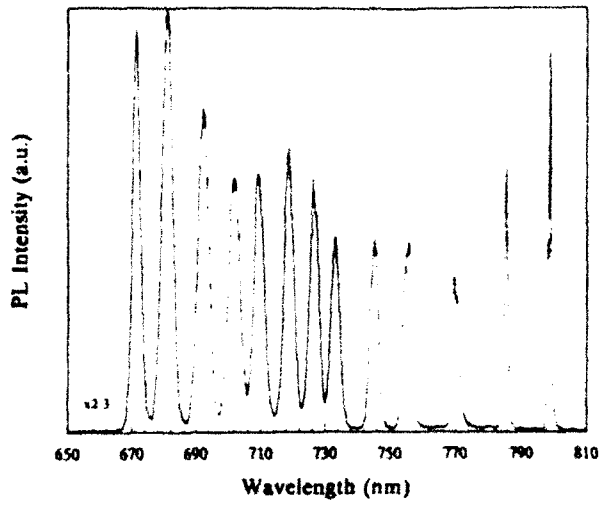


FIG. 1. PL spectrum of QW sample containing 13 wells, 4, 5, 6, 7, 8, 9, 10, 11, 13, 15, 19, 25, and 35 ML thick, separated by 300 Å Al_0 ,Ga_0 ,As barriers.

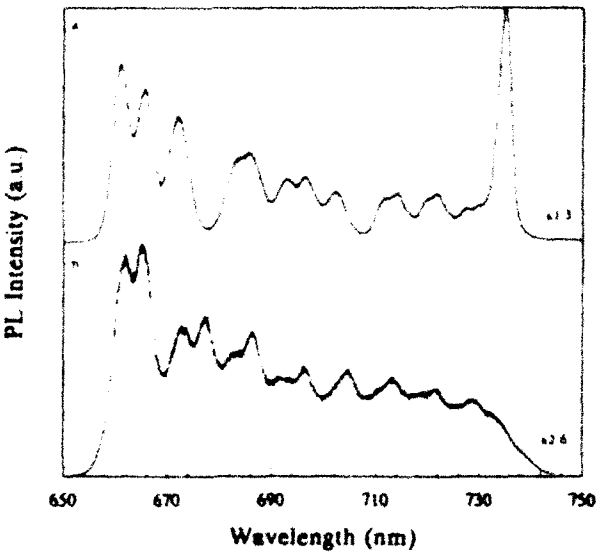


FIG. 3. PL spectrum of QW sample containing integer and half-monolayer wells grown with interrupt times of (a) 10 s and (b) 30 s.

tegrated intensities from this series of samples which was found to monotonically decrease with increasing interrupt time (the integrated intensity was down a factor of 2 for the 120 s interrupt sample). We have previously found degradation in the PL intensity and linewidth of QW samples grown at different As cracking temperatures, which we attributed to the generation of small amounts of AsO.[6] Even when the AsO was reduced below the detection limits of the residual gas analyzer after an extensive bake, we still found some degradation in the PL indicating the contaminant had not been completely removed. Since contaminants may be expected to build up on the surface during growth interruption, it is not surprising that PL degradation is seen in the present work. Our earlier work also indicated that the larger wells grown near the substrate seemed to suffer a proportionally higher loss in peak intensity than the thinner wells near the surface, similar to the effect seen in Fig. 4.

There is clearly not a distinct change of half-monolayer wells into regions of atomically flat, integer-monolayer wells. Instead, we observe a shift in the transition energies of the nominally integer-monolayer wells. This is shown in Fig. 5 where the transition energy for the 3 and 4 ML wells are plotted as a function of interrupt time. In Fig. 5 one can see a relatively large shift in the transition wavelengths to longer wavelengths as the growth interruption time increases, although it does seem to approach a constant value for longer interrupt times. Shown for comparison in Fig. 5 is the calculated transition wavelengths for well thicknesses

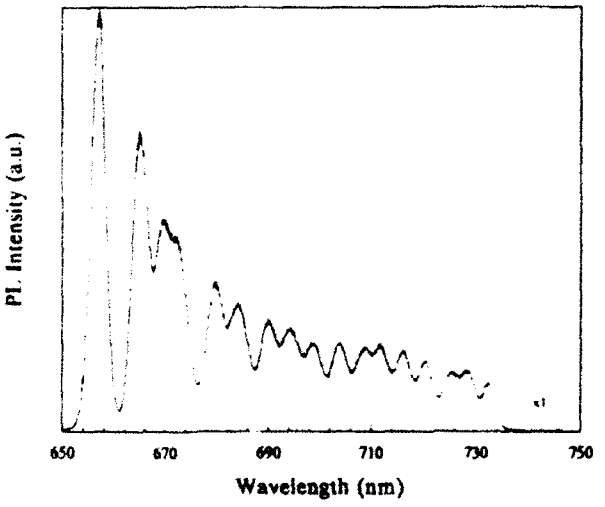


FIG. 2. PL spectrum of QW sample containing integer and half-monolayer wells, of nominal thicknesses 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 10, 10.5, and 11 ML, separated by 300 Å Al_0 ,Ga_0 ,As barriers.

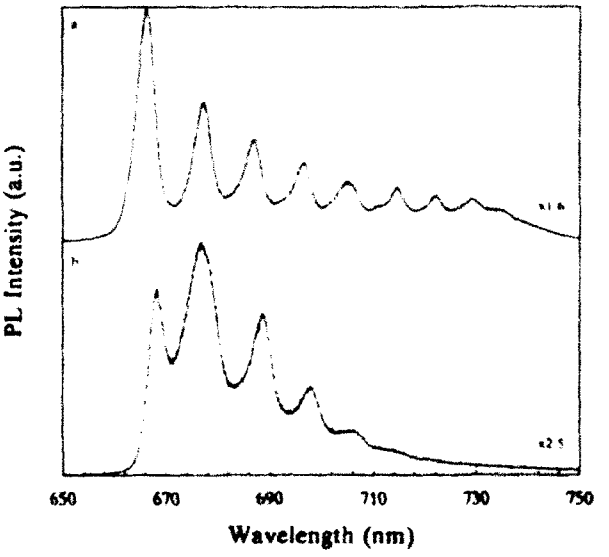


FIG. 4. PL spectrum of QW sample containing integer and half-monolayer wells grown with interrupt times of (a) 60 s and (b) 120 s.

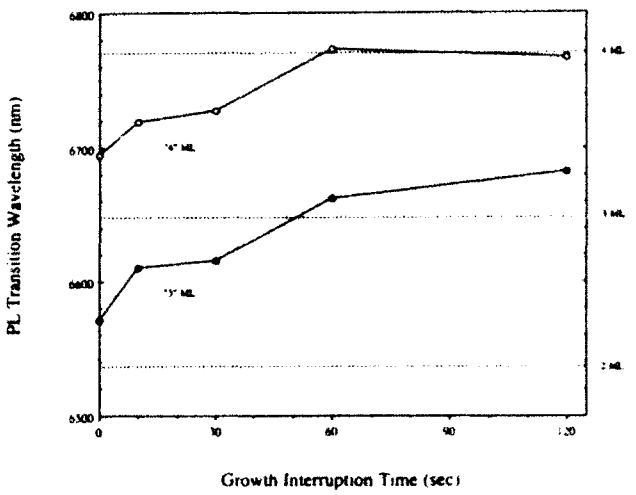


FIG. 5. Measured PL transition wavelengths for wells of nominal 3 and 4 ML thickness with increasing interrupt time. Also shown for comparison are the calculated transition wavelengths (dashed lines).

of 2, 3, and 4 ML using an effective-mass model.[11] These results can be interpreted in terms of the "microroughness" model.[4] In this model the interface between the layers is nearly abrupt on an atomic scale with small regions (much less than an exciton diameter in size) deviating in thickness by one or more monolayers. The exciton sees an average potential and gives rise to a sharp transition but shifted in energy from one with perfectly abrupt interfaces. The shift in the PL transition wavelength would mirror the relaxation of the microroughness. If this interpretation is correct it would also imply that the samples grown without interruption have significant microroughness. It may be fruitful to examine the magnitude of the peak shift in conjunction with a simulation of kinetic processes occurring on the surface such as that of Ogale *et al.*[3]

IV. CONCLUSION

A novel method involving the growth of quantum wells having nominal integer plus one-half monolayer thickness, in conjunction with growth interruption has been used to examine MBE growth processes. Our results indicate that approximately 60 s is sufficient to smooth the interfaces under these growth conditions. The continuous shift in the transition energies with interrupt time indicates the presence of microroughness. However, degradation of the samples with increasing interrupt time illustrates the importance of the vacuum ambient and reducing contaminants in the system.

ACKNOWLEDGMENTS

This work was supported by the Joint Services Electronics Program under Contract No. AFOSR F491620-89-C-0044, the Army Research Office under Contract No. DAAL-03-88-K-0060, and the Science and Technology Center program of the National Science Foundation, Grant No. CHE 8920120.

[a]Present address: TRW Electronics & Technology Division, Electronic Systems Group, Redondo Beach, CA 90278

[1]K. Ploog and G. H. Dohler, Adv. Phys. 32, 285 (1983).
[2]J. Singh and K. K. Bajaj, J. Appl. Phys. 57, 5433 (1985).
[3]S. B. Ogale, A. Madhukar, F. Voillot, M. Thomsen, W. C. Tang, T. C. Lee, J. Y. Kim, and P. Chen, Phys. Rev. B 36, 1662 (1987).
[4]R. F. Kopf, E. F. Schubert, T. D. Harris, and R. S. Becker, Appl. Phys. Lett. 58, 631 (1991).
[5]C. T. Foxon, J. B. Clegg, K. Woodbridge, D. Hilton, P. Dawson, and P. Blood, J. Vac. Sci. Technol. B 3, 703 (1985).
[6]T. R. Block and B. G. Streetman, J. Cryst. Growth 111, 98 (1991).
[7]D. Bimberg, D. Mars, J. N. Miller, R. Bauer, and D. Oertel, J. Vac. Sci. Technol. B 4, 1014 (1986).
[8]T. Fukunaga, K. L. I. Kobayashi, and H. Nakashima, Surf. Sci. 174, 71 (1986).
[9]B. A. Joyce, J. Zhang, J. H. Neave, and P. J. Dobson, Appl. Phys. A 45, 255 (1988).
[10]T. J. Mattord, V. P. Kesan, G. E. Crook, T. R. Block, A. C. Campbell, D. P. Neikirk, and B. G. Streetman, J. Vac. Sci. Technol. B 6, 1667 (1988).
[11]G. Bastard, Phys. Rev. B 24, 5693 (1981).

operation on asymmetrical Y lasers. We obtain up to 32 dB sidemode suppression. The tuning response of a device with $\Delta L = 56\,\mu m$ is shown in Fig. 4 together with the theoretical tuning scheme curve. With one single current I_4 the MZ filter is tuned across closely spaced FP cavity modes within a 5·5 nm wavelength range which corresponds to the MZ filter width. The correlation to our simple theory is fine at low injection currents. At higher injection we find deviation due to our assumptions, taking no gain saturation, carrier heating and interaction of the four carrier reservoirs into account.

In a further operation mode (not shown in the Figure), again with one single current I_4, the gain curve is shifted across the MZ modes of the asymmetrical device. As an example, for a Y laser with $\Delta L = 81\,\mu m$ (corresponding to an MZ spacing of 4 nm) an 8 nm shift across three MZ modes is accomplished in a first experimental demonstration.

Conclusions: To overcome the unfavourable tuning control of the symmetrical Y laser, we have designed and realised the first asymmetrical Y-laser structures. For these devices the expected much simpler tuning scheme is confirmed. Further extensibility of the actually measured single current tuning spans is expected for optimised devices and appropriately designed MZ filter functions. The asymmetrical Y-laser structures offer easily controllable wavelength addressing schemes. Therefore, the successful implementation of the enormous potential of Y lasers into photonic switching and WDM transmission systems is expected.

Acknowledgments: We thank especially K. Daub and A. Nowitzki for excellent technical assistance. This work is partially financially supported by the German Minister for Research and Technology (project TK0581), and in part by the European RACE II project 2039 (ATMOS).

3rd July 1992

M. Schilling, K. Dütting, W. Idler, D. Baums, G. Laube, K. Wünstel and O. Hildebrand (*Alcatel-SEL Research Centre, Optoelectronic Components Division ZFZ/WO, Lorenzstr. 10, D-7000 Stuttgart 40, Germany*)

References

1 SCHILLING, M., SCHWEIZER, H., DÜTTING, K., IDLER, W., KÜHN, E., NOWITZKI, A., and WÜNSTEL, K.: 'Widely tunable Y-coupled cavity integrated interferometric injection laser', *Electron. Lett.*, 1990, 26, (4), pp. 243–244

2 SCHILLING, M., IDLER, W., KÜHN, E., LAUBE, G., SCHWEIZER, H., WÜNSTEL, K., and HILDEBRAND, O.: 'Integrated interferometric injection laser: novel fast and broad-band tunable monolithic light source', *IEEE J. Quantum Electron.*, 1991, QE-27, (6), pp. 1616–1624

3 SCHILLING, M., IDLER, W., BAUMS, D., LAUBE, G., WÜNSTEL, K., and HILDEBRAND, O.: 'Multifunctional photonic switching operation of 1500 nm Y-coupled cavity laser (YCCL) with 28 nm tuning capability', *IEEE Photonics Technol. Lett.*, 1991, 3, (12), pp. 1054–1057

4 IDLER, W., SCHILLING, M., BAUMS, D., LAUBE, G., WÜNSTEL, K., and HILDEBRAND, O.: 'Y laser with 38 nm tuning range', *Electron. Lett.*, 1991, 27, (24), pp. 2268–2269

5 HILDEBRAND, O., SCHILLING, M., IDLER, W., BAUMS, D., LAUBE, G., and WÜNSTEL, K.: 'The integrated interferometric injection laser (Y-laser): one device concept for various system applications'. Proc. ECOC/IOOC'91 Paris, 1991, vol. 2, pp. 39–46

6 WÜNSTEL, K., IDLER, W., SCHILLING, M., LAUBE, G., BAUMS, D., and HILDEBRAND, O.: 'Y-shaped semiconductor device as a basis for various photonic switching applications'. Proc. OFC'92 San Jose, 1992, p.125

7 IDLER, W., SCHILLING, M., BAUMS, D., DÜTTING, K., LAUBE, G., WÜNSTEL, K., and HILDEBRAND, O.: 'Wide range 2·5 Gb/s wavelength conversion with a tunable Y-laser'. Proc. 18th ECOC Berlin 1992

8 ZIELINSKI, E., KEPPLER, F., HAUßER, S., PILKUHN, M. H., SAUER, R., and TSANG, W. T.: 'Optical gain and loss processes in GaInAs/InP MQW laser structures', *IEEE J. Quantum Electron.*, 1989, QE-25, (6), pp. 1407–1416

9 WIEDEMANN, P., KLENK, M., KOERNER, U., WEINMANN, R., ZIELINSKI, E., and SPEIER, P.: 'MOVPE of In(GaAs)P/InGaAs MQW-structures', *J. Cryst. Growth*, 1991, 107, pp. 561–566

10 SPEIER, P., BOUAYAD-AMINE, J., CEBULLA, U., DÜTTING, K., KLENK, M., LAUBE, G., MAYER, H. P., WEINMANN, R., WÜNSTEL, K., ZIELINSKI, E., and HILDEBRAND, O.: '10 Gbit/s MQW-DFB-SIBH lasers entirely grown by LPMOVPE' *Electron. Lett.*, 1991, 27, (10), pp. 863–864

HIGH EFFICIENCY MICROWAVE DIODE OSCILLATORS

S. Javalagi, V. Reddy, K. Gullapalli and D. Neikirk

Indexing terms: Oscillators, Microwave oscillators, Microwave devices, Diodes

Data on the operation of very high efficiency microwave oscillators using AlAs/InGaAs quantum well injection transit (QWITT) diodes are presented. A DC-to-RF power conversion efficiency as high as 50% was achieved, which, to the authors' knowledge, is the highest efficiency reported for continuous wave (CW) operation of a two terminal semiconductor device.

Introduction: Because the high frequency output power of an individual semiconductor device is limited, there has been increasing interest in new power combining techniques for microwave to submillimetre wave systems. For instance, large output powers have been generated using two dimensional grid array sources [1]. Although total RF output power may be increased, DC-to-RF power conversion efficiency remains an intrinsic property of the semiconductor device. For two-terminal diodes, the highest DC-to-RF conversion efficiencies have been obtained from IMPATTs, which in pulsed operation have achieved 60% [2]. In CW operation, only 43% has been achieved [3]. We present DC and RF data for two AlAs/InGaAs quantum well injection transit (QWITT) diodes with 1200 and 1800 Å depletion region lengths. DC-to-RF power conversion efficiencies in the range 44–50% were obtained under optimum RF circuit/device matching conditions. In addition, the 1800 Å device produced a single device output power of 20 mW without requiring any device heatsinking or substrate thinning.

Experiment: The two AlAs/InGaAs QWITT diodes with drift region lengths of 1800 and 1200 Å, referred to as QWITT 1 and QWITT 2, respectively, were grown on $n+$ InP substrates by molecular beam epitaxy. The motivation for the QWITT diode is that the voltage and current differences between the peak and the valley, ΔV_{pv} and ΔI_{pv}, respectively, must be made as large as possible to increase output power [4, 5]. The quantum well is formed by sandwiching a 47 Å $In_{0.53}Ga_{0.47}As$ layer between 25 Å strained AlAs layers. The quantum well is between a symmetric layer profile consisting of 15 Å of undoped $In_{0.53}Ga_{0.47}As$ adjacent to the well followed by ~100 Å n ($2·3 \times 10^{18}\,cm^{-3}$) $In_{0.53}Ga_{0.47}As$ and 100 Å n ($1·3 \times 10^{17}\,cm^{-3}$) $In_{0.53}Ga_{0.47}As$. The drift region is placed below the quantum well. Standard processing techniques were used to fabricate mesa isolated diodes with Au topside contacts. The room temperature peak current density was ~8 kA/cm² and the peak to valley current ratio was ~20. The stabilised (non-oscillating) current density against voltage characteristics of typical diodes are shown in Fig. 1.

Two open circuit single stub resonator circuits were designed using 50 Ω microstrip lines for oscillation at 1 and 2·7 GHz. The design is based on the Kurokawa oscillation criterion (i.e. $Z_{device} = -Z_{circuit}$) [6]. Both small signal and large signal calculations on the QWITT diodes show a very nearly real impedance for low microwave frequencies (< 5 GHz). The open circuit stub in the circuit produces a real impedance at the device only over a narrow band of frequencies; thus the Kurokawa condition can be satisfied only in this frequency band. For the stub lengths chosen the circuit impedance should be purely real at ~1 and 2·7 GHz for circuits 1 and 2, respectively. Because the diode impedance is nearly real, the frequency of oscillation is expected to be independent of the diode area or structure (i.e. QWITT 1 or QWITT 2). The diodes were biased with an HP11612A bias tee and the DC port was monitored for bias line oscillations.

Results and discussion: The diode mesas were contacted by a tungsten whisker at the end of the circuit and different areas on each sample were probed. The RF output power was measured by placing each sample at the end of the circuit and varying the DC bias voltage until maximum output power

was obtained. The large signal specific negative resistance of the device is a strong function of the RF amplitude, hence for a fixed RF circuit the power density should vary with device

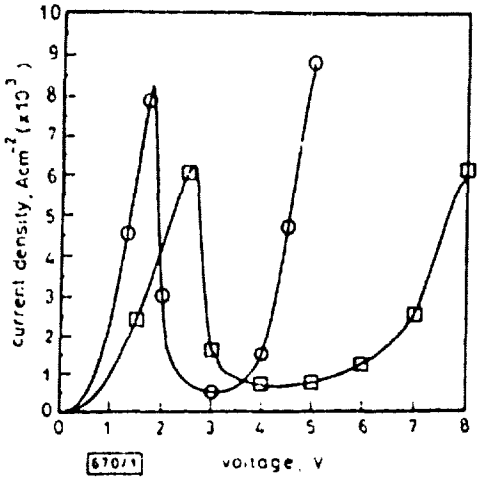


Fig. 1 *Typical stabilised (non-oscillating) DC J–V curves for QWITT 1 and QWITT 2*

□ QWITT 1
○ QWITT 2

area. The frequency spectrum of the QWITT oscillators was fairly narrow, producing a typical spectrum which was 15 dB down 5 kHz off the centre frequency for a resolution bandwidth of 1 kHz. The linewidth was limited by the very low Q circuit used here. Particular care was taken to ensure the accuracy of the RF power measurements through the use of a calibrated RF power meter. The values reported here should be conservative, because the only measurement error corrected for was a 1 dB cabling loss; no compensation for the loss due to the microstrip circuit, connectors, and bias tee was made.

The experimental results for both QWITT 1 and QWITT 2 operating in the two circuits are summarised in Table 1. Values quoted are averages for four to five nominally identical diodes; scatter was ~5% of the average value. For some devices, a DC-to-RF power conversion efficiency as high as 50% was achieved at a power density of 4 kW/cm^2 for the QWITT 1 structure; the average result for several diodes was 45%. To our knowledge, this is the highest efficiency ever reported for continuous wave operation with a two terminal device. Conventional IMPATT diodes have been shown to have efficiencies of 60%, but only in pulsed operation [2]. For CW operation, IMPATTs have demonstrated efficiencies only as high as 43%, but more typically in the range 16–36% [3, 7]. Gunn diodes have DC-to-RF conversion efficiencies of less than 15% [8, 9]. The highest efficiency reported for quantum well diode oscillators, hitherto, was 11% at a frequency of 1 GHz [10].

The maximum power obtained from a single device was 20 mW, which was achieved without any heatsinking. This is the highest power reported to date for quantum well diode oscillators, a factor of 20 greater than previously reported in the literature (980 μW at 1 GHz [10]). Such high power is possible because these QWITT diodes have very high specific negative resistance, allowing the use of very large area diodes without producing bias line oscillations.

Quasistatic power calculations were performed by numerically varying the RF voltage swing at different bias points in the NDR region, using the stabilised DC J–V curves of both diodes [5]. We find that both the maximum measured and calculated RF power obtained are greater than the frequently quoted '3/16 $\Delta V\ \Delta I$' rule [10]. This is because the DC I–V curve is not cubic, as used to obtain the '3/16 $\Delta V\ \Delta I$' rule [11], and that the optimum bias point is not the average of the peak and valley voltages, but rather closer in value to the valley voltage. At the optimum bias and RF voltage swing, simple large signal calculations on QWITT 1 (based on the J–V curve shown in Fig. 1) yield a power density of 4·1 kW/cm^2 with a DC-to-RF conversion efficiency of 42%, and QWITT 2 yields 3·0 kW/cm^2 with 40% efficiency.

In conclusion, we have presented data on AlAs/InGaAs QWITT microwave oscillators that have produced record DC-to-RF power conversion efficiencies as high as 50% and CW output powers of 20 mW from a single device.

Acknowledgments: This work was sponsored by the Texas Advanced Research Program and the Joint Services Electronics Program under grant number AFOSR 49620-89-C-0044.

10th July 1992

S. Javalagi, V. Reddy, K. Gullapalli and D. Neikirk (*Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712, USA*)

References

1 RUTLEDGE, D. B., POPOVIC, Z. B., WEIKLE II, R. M., KIM, M., POTTER, K. A., COMPTON, R. C., and YORK, R. A.: 'Quasi-optical power-combining arrays', 1990 IEEE MTT-S, Int. Microwave Symp. Dig., 1990, 3, pp. 1201–1204

2 CHANG, K. N.: 'Avalanche diodes as UHF and L-band sources', *RCA Rev.*, 1969, 30, pp. 3–14

3 IGLESIAS, D. E., and EVANS, W. J.: 'High-efficiency CW IMPATT operation', *Proc. IEEE.*, 1968, 56, p. 1610

4 KESAN, V. P., MORTAZAWI, A., MILLER, D. R., REDDY, V. K., NEIKIRK, D. P., and ITOH, T.: 'Microwave and millimeter-wave QWITT diode oscillators', *IEEE Trans.*, 1989, MTT-37, (12), pp. 1933–1941

5 REDDY, V. K., TSAO, A. J., JAVALAGI, S., KUMAR, G. K., MILLER, D. R., and NEIKIRK, D. P.: 'Quantum well injection transit time (QWITT) diode oscillators'. Proc. 15th Int. Conf. Infrared and Millimeter Waves, 10th–14th December 1990, pp. 88–90

6 KUROKAWA, K.: 'Some basic characteristics of broadband negative resistance oscillator circuits', *Bell Syst. Tech J.*, 1969, 48, pp. 1937–1955

7 GEWARTOWSKI, J. W.: 'Progress with CW IMPATT diode circuits at microwave frequencies', *IEEE Trans.*, 1979, MTT-27, (5), pp. 434–442

8 FRANK, F. B., and DAY, G. P.: 'High CW power K-band Gunn oscillators', *Proc. IEEE.*, 1969, 57, p. 339

9 CHANG, K. (Ed.): 'Handbook of microwave and optical components: Vol. 2' (Wiley, New York, 1990), pp. 298–299

Table 1 AVERAGE RESULTS FOR POWER DENSITY AND DC-TO-RF CONVERSION EFFICIENCY USING QWITT 1 AND QWITT 2 IN CIRCUITS 1 AND 2

Device		Circuit 1			Circuit 2		
Sample	Area	Frequency	Power density	Efficiency η	Frequency	Power density	Efficiency η
	cm^2 (x 10^{-8})	GHz	kW/cm^2	%	GHz	kW/cm^2	%
QWITT 1	300	0·941	3·90	42	2·380	3·6	39
W = 1800 Å	160	0·951	2·85	39	2·313	4·0	45
	68	0·953	2·40	27	2·448	2·8	36
QWITT 2	300	0·943	2·60	35	2·238	1·5	20
W = 1200 Å	144	0·930	3·00	37	2·320	2·2	26
	68	0·947	2·40	35	2·357	1·7	24

Results averaged over four to five devices; maximum efficiency as high as 50% was obtained for QWITT 1 in circuit 2 for a nominal 160 x 10^{-8} cm^2 area device

10 BROWN, E. R., PARKER, C. D., CALAWA, A. R., MANFRA, M. J., SOLLNER, T. C. L. G., CHEN, C. L., PANG, S. W., and MOLVAR, K. M.: 'High-speed resonant tunneling diodes made from the InGaAs/AlAs material system', *SPIE* (High Speed Electronics and Device Scaling), 1990, 1288, pp. 122–135
11 CHOW, W. F.: 'Principles of tunnel diode circuits' (New York, 1964)

VOLTAGE CONTROLLED OSCILLATORS USING COMPLEMENTARY CURRENT MIRROR INVERTER

R. R. Pal, H. N. Acharya, S. Kai and N. B Chakrabarti

Indexing terms: Inverters, Voltage controlled oscillators, Ring oscillators, Circuit design

Complementary current mirror inverters have large bandwidth, small time delay but small gain whereas complementary inverters have high gain but large time delay. Ring oscillators have been realised using a combination of both types of inverter to achieve stable high frequency oscillation. Voltage controlled oscillators were obtained by tuning the frequency of oscillation with supply voltage.

Introduction: Voltage controlled oscillators (VCOs) are widely used in such applications as phase locked loops, timing recovery, modulators and demodulators, and frequency synthesisers. Ring oscillator VCOs have the potential of very high frequency operation and low cost. The oscillation frequency of a ring oscillator is fixed and is determined by the propagation delay encountered by the number of inverter stages used to realise the oscillator. Silicon bipolar technology using merged *npn-pnp* complementary inverters is attractive for the realisation of high frequency VCOs because of high integrability and reduced process complexity. Merged *npn-pnp* complementary inverters with current mirrors appear to be particularly suited for VCO applications because the associated circuits for the phase frequency detector are realisable in current mirror logic using the same configuration [1]. Merged *npn-pnp* inverters can be used in three basic forms: the *npn* driver, *pnp* current souce [2, 3], the complementary inverter (CI), and the complementary inverter with current mirror (CCMI). This Letter describes the implementation of a VCO using a CCMI ring oscillator.

VCO using complementary current mirror inverters (CCMIs): The VCO, as shown in the inset of Fig. 1, is primarily a ring oscillator wherein the output of an odd number of serially connected CCMIs is fed back to the input stage. The frequency of oscillation depends on the number of inverter stages, the load capacitances and the current level.

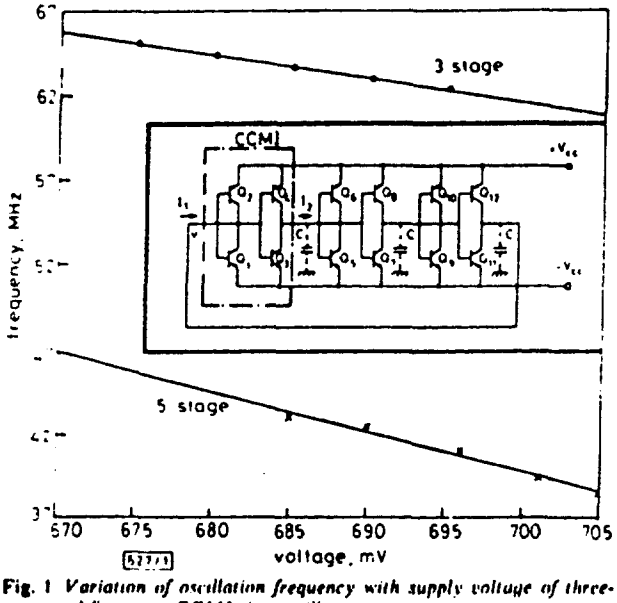


Fig. 1 *Variation of oscillation frequency with supply voltage of three-stage and five-stage CCMI ring oscillator*

Inset: circuit diagram of three-stage CCMI ring oscillator

To determine the DC behaviour of the inverter, we assume that the base input voltage of Q_3 is V and its collector voltage is V_0. If $V_{cc} \leq 0.5$ V, the quiescent current is very small. The input and output currents I_1 and I_2, as shown in the Figure, are given by

$$I_1 = (I_{sc1} + I_{sb3}) \exp [q(V + V_{cc})/kT] - (I_{sc2} + I_{sb4}) \exp [q(V_{cc} - V)/kT] \quad (1)$$

$$I_2 = I_{sc4} \exp [q(V_{cc} - V)/kT] - I_{sc3} \exp [q(V + V_{cc})/kT] \quad (2)$$

where I_{sc} is the saturation current with the collector shorted to the base and I_{sb} is the base current with active collector. I_{sc2}, I_{sc4} and I_{sc3} are the collector currents of transistors Q_2, Q_4 and Q_3, respectively, and I_{sc4} is the base current of Q_4.

Assuming $\sqrt{(I_{s1}I_{s2})} = I_{sm}$, $\sqrt{(I_{sc3}I_{sc4})} = I_{s1}$, $\sqrt{[(I_{s1}/I_{s2})]} = \exp(C_1)$ and $\sqrt{[(I_{sc3}/I_{sc4})]} = \exp(C_2)$ where $I_{s1} = (I_{sc1} + I_{sb3})$ and $I_{s2} = (I_{sc2} + I_{sb4})$, we obtain

$$I_2 = -I_1 \frac{I_{s1}}{I_{sm}} \frac{\sinh q(V + C_2)/kT}{\sinh q(V + C_1)/kT} \quad (3)$$

This implies that transistors Q_3 and Q_4 would need to have a larger area for current gain. The small signal AC gain of an inverter stage consisting of a complementary inverter (CI) followed by a mirror can be expressed in terms of the transconductance and admittances of the two stages. The small signal DC gain is given by

$$G_0 = (g_{mn} + g_{mp})/(g_{nm} + g_{pm} + g_{dn} + g_{dp}) \quad (4)$$

where g_{dn} and g_{dp} are the output conductances of the inverter, g_{nm} and g_{pm} are the conductances of the mirror transistors. The above expression shows that the DC gain can be controlled by varying the relative sizes of the inverters and mirror transistors.

Results and discussion: The VCO circuit (shown in Fig. 1) was wired in a PCB. The transistors used are BF 494B (*npn*) and BC 157B (*ppn*) whose f_t are 300 and 150 MHz, respectively. The experimental results of the VCO realised using three-stage and five-stage CCMI ring oscillators are shown in Fig. 1. The frequency of oscillation increases from 62 to 66 MHz for a supply voltage variation of 30 mV for the three stage oscillator whereas a variation of 10 MHz (35–45 MHz) is obtained over the same voltage variation for the five-stage CCMI VCO. Variation of frequency with load capacitance (shown by the dotted line in the inset of Fig. 1) for a three-stage CCMI VCO is shown in Fig. 2. The VCO performance is linear and suitable for PLL applications. The photograph of a typical VCO waveform obtained from a three-stage CCMI VCO without load capacitance is shown in the inset of Fig. 2.

It is interesting to study the variation of oscillator frequency when a combination of the complementary current mirror inverter (CCMI) and complementary inverter (CI) is used (Fig. 3 (inset)). Fig. 3 shows the experimental results of a three-stage ring oscillator where combination of a CCMI and CI is used. The frequency of oscillation reduces in proportion to the number of CCMI stages which are replaced by a CI. This is consistent with the fact that CCMIs have large bandwidth, small time delay but small gain, whereas CIs have high gain but large time delay. The maximum frequency of oscillation reduces from 65 to 4·00 MHz when all of the three CCMI stages are replaced by a CI (Figs. 1 and 3). Because CCMI gain and time delay depend on the relative areas of Q_3, Q_4 and Q_2, Q_1, the operating frequency can be changed by placing the inverters in parallel, either at the mask stage or subsequently. It will be noticed from Fig. 3 that the frequency of oscillation initially increases with supply voltage and thus with current. However when the current exceeds 5 mA per stage, the output frequency decreases. This may be attributed to the increase of diffusion capacitance.

Fig. 3 shows the waveforms in the NSIS as observed by a streak camera. The input control pulse was broadened from 9 to 12 ps as it traversed the loop due to the interplay of the normal dispersion and selfphase modulation, and walked through the signal by 10 ps (Fig. 3*a*). Fig. 3*b* shows transmitted (switched) and reflected (unswitched) signals for a CW probe signal input with a wavelength of 1552 nm. The peak level of the switched signal was about 60% of that with complete switching due to the low control pulse power (420 mW peak at Port *b*3). The switched signal had a 13 ps width. The switching contrast was about 17 dB mainly due to the poor extinction ratio of the polarisation controller in the loop. Total insertion loss of the NSIS for the signal was 12 dB.

Time-division demultiplexing was then demonstrated using a fixed pattern input data stream. The input signal with a wavelength of 1552 nm was generated from a gain-switched DFB-LD driven at 5·6 GHz. The resultant pulses with 30 ps pulse width were then compressed to 8 ps using a 1250 m dispersion-shifted fibre. The compressed pulses were injected into a fibre optic pulse pattern generator consisting of two highly-birefringent fibres and a following fibre polariser. The two birefringent fibres, whose polarisation dispersion was 2 ps/m, were about 10 and 20 m in length and connected with their polarisation axes offset by 45°. The two fibres delay the input pulses by about 20, 40 and 60 ps to produce a 50·4 Gbit/s '... 011110000011110 ...' pattern data stream as shown in Fig. 3*c*. The resultant data stream was then injected into the NSIS. Fig. 3*d–e* show the demultiplexed 5·6 Gbit/s '... 11 ...' signal (Fig. 3*d*) and the remaining '... 011010000011010 ...' signal (Fig. 3*e*). The PLC was seen to be stable even though no temperature control was employed.

We have reported the first NSIS that is practical. It is based on an integrated silica-based planar lightwave circuit which provides wavelength selecting functions with the low insertion loss (<4·3 dB), high isolation (>40 dB), and high environmental stability. The NSIS was tested and demonstrated diode-pumped time-division demultiplexing at 50·4 Gbit/s with a 13 ps switching time and a 5·6 GHz repetition rate.

Acknowledgment: The authors would like to express their appreciation to T. Sugie, M. Koga, A. Takada, and T. Morioka for their valuable technical advice and supplying the erbium doped fibre, the polarisation-independent optical circulator, and the DFB-LD. They also wish to thank H. Ishio, M. Kawachi, and T. Miya for their continual encouragement and guidance.

2nd December 1991

M. Jinno and T. Matsumoto (*NTT Transmission Systems Laboratories, 1-2356, Take, Yokosuka, Kanagawa, 238-03, Japan*)

T. Kominato and Y. Hibino (*NTT Opto-electronics Laboratories, 162, Shirakata-shirane, Tokai, Naka, Ibaraki, 319-11, Japan*)

References

1 JINNO, M., and MATSUMOTO, T.: 'Demonstration of laser-diode-pumped ultrafast all-optical switching in nonlinear Sagnac interferometer', *Electron. Lett.*, 1991, **27**, pp. 75–76

2 TAKADA, A., AIDA, K., and JINNO, M.: 'Demultiplexing of 40-Gb/s optical signal to 2.5-Gb/s using a nonlinear fibre loop mirror driven by amplified, gain-switched laser diode pulse'. Tech. Dig. OFC '91, San Diego, 1991, TuN3

3 NELSON, B. P., BLOW, K. J., CONSTANTINE, P. D., DORAN, N. J., LUCEK, J. K., MARSHALL, I. W., and SMITH, K.: 'All-optical Gbit/s switching using nonlinear optical loop mirror', *Electron. Lett.*, 1991, **27**, pp. 704–705

4 TAKATO, N., JINGUJI, K., YASU, M., TOBA, H., and KAWACHI, M.: 'Silica-based single-mode waveguides on silicon and their application to guided-wave optical interferometers', *J. Lightwave Technol.*, 1988, **LT-6**, pp. 1003–1010

5 SUGIE, T., OHKAWA, N., IMAI, T., and ITO, T.: 'A 2·5 Gb/s, 364 km CPFSK repeaterless transmission experiment employing an Er-doped fibre amplifier and SBS suppression optical link'. Tech. Dig. Optical Amplifiers and Their Applications Topical Meeting, 1990, PDP2-1

6 KOGA, M., and MATSUMOTO, T.: 'Polarization-insensitive high-isolation nonreciprocal device for optical circulator application', *Electron. Lett.*, 1991, **27**, pp. 903–904

WAVELET ANALYSIS OF ELECTROMAGNETIC BACKSCATTER DATA

H. Kim and H. Ling

Indexing terms: Electromagnetic waves, Wavelets, Signal processing, Scattering

An analysis of electromagnetic backscatter data using the wavelet transform is presented. The multiresolution property of wavelets enables scattering events with different scales in the frequency domain to be simultaneously resolved. The backscattered signal from a dispersive plasma cylinder is studied using the wavelet analysis. The different scattering mechanisms are clearly identifiable in the time-frequency representation.

Introduction: The electromagnetic energy backscattered from an unknown scatterer can provide useful information for identifying and characterising the scatterer. Recently, the running window Fourier transform (RWFT) (also known as the short time Fourier transform [2, 3]) was applied to the backscattered frequency response from an open-ended waveguide cavity to obtain a joint time-frequency representation of the original signal [1]. In the time-frequency plane, localised events in time (scattering centres), localised events in frequency (resonances), and dispersive scattering mechanisms can be simultaneously displayed. Consequently, a good insight was gained on the different scattering mechanisms which, when properly interpreted, can be correlated with the scatterer features for target identification applications. However, there is one intrinsic limitation associated with the RWFT. The scattered signal in the frequency domain generally comprises scattering mechanisms with different characteristic scales. For example, high-Q resonance phenomena are small-scale frequency events whereas contributions from nondispersive scattering centres extend over large frequency scales. Because the RWFT uses a fixed window width, it cannot resolve multiple events with different scales simultaneously. In this Letter, a wavelet analysis of backscatter data is presented as an alternative tool to the RWFT. The wavelet transform has attracted much attention in the signal processing community recently [2–5]. In contrast to the RWFT, the wavelet transform technique has a multiresolution property through the use of multiscale windows. Therefore, multiscale events of frequency can be more effectively resolved by the wavelet transform. As an example, a wavelet analysis of the backscattered frequency data from a dispersive plasma cylinder is presented.

Wavelet analysis: We define the wavelet transform of a frequency signal $F(\omega)$ as

$$W_f(\tau, \Omega) = \int F(\omega)\tau^{1/2} H(\tau(\omega \cdot \Omega))\, d\omega \qquad (1a)$$

$H(\omega)$ is usually referred to as the 'mother wavelet'. Note that the wavelet transform is defined in the frequency domain for our application, contrary to its usual definition in the time domain. Eqn. 1 can be interpreted as the decomposition of the frequency response $F(\omega)$ into a family of shifted and dilated wavelets $H(\tau(\omega \cdot \Omega))$. By shifting $H(\omega)$ with a fixed scale parameter τ, the τ-scale mechanisms in the frequency response $F(\omega)$ can be extracted and localised. Alternatively, by dilating $H(\omega)$ at a fixed Ω, all of the multiscale events of $F(\omega)$ at Ω can be analysed according to the scale parameter τ. This is the so called 'multiresolution' property of the wavelet transform and is an important advantage over the RWFT.

By properly manipulating eqn. 1*a*, the wavelet transform can also be carried out on the time domain signal $f(t)$ of the original frequency domain response $F(\omega)$.

$$W_f(\tau, \Omega) = \int f(t)\tau^{-1/2} h(t/\tau) e^{j\Omega t}\, dt \qquad (1b)$$

$h(t)$ is the Fourier transform of $H(\omega)$ and must satisfy the 'admissibility condition', i.e. $h(0) = 0$. Because eqn. 1*b* is essentially the Fourier transform of $[f(t)h(t/\tau)]$, it is the preferred

numerical implementation of the wavelet transform through the use of the FFT.

Backscatter from plasma cylinder: The backscatter from a plasma cylinder is considered. Our interest in this problem stems from the study of exhaust plume plasma from electrothermal arcjet thrusters [6]. The circular plasma cylinder has a radius of 11·94 cm. A cold plasma model is assumed to model the cylinder as a dispersive dielectric with $\varepsilon_r = [1 - (f_p/f)^2]$. f_p is the plasma frequency and is chosen to be 10 GHz. To generate the backscatter frequency response, the eigenfunction solution [7] for a circular dielectric cylinder is implemented numerically. A complex Bessel function routine [8] is used in generating the numerical data. The simulated backscattered response from the plasma cylinder from 3 to 17 GHz is shown in Fig. 1. We observe that the frequency response consists of two regions: the flat response below the plasma frequency and the oscillatory pattern above the plasma frequency. By examining the frequency response more closely around the plasma frequency, we also observe additional small-scale features which are due to resonance and dispersion phenomena. In this instance, the frequency response is composed of multiscale events and the wavelet analysis is ideally suited to analysing these multiscale frequency data.

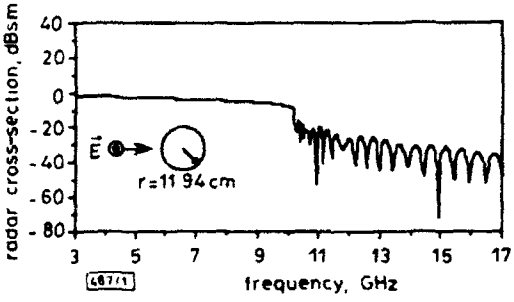


Fig. 1 *Backscattered frequency response of uniform plasma cylinder with f_p = 10 GHz and radius 11·94 cm*

$\varepsilon_r = 1 - (f_p/f)^2$

The wavelet transform was implemented based on eqn. 1*b* using the FFT routine. The function $h(t)$ was chosen to be a two-sided Kaiser-Bessel window with a constant Q factor of 0·3. (The Q factor is defined as the ratio between the window width and the window centre.) The resulting time-frequency plot obtained using the wavelet transform is shown in Fig. 2. Also plotted along the two axes are the time-domain and the frequency-domain responses. Two distinct scattering mechanisms can be identified in the wavelet domain: a vertical line due to the exterior specular reflection of the cylinder and a dispersion curve due to the once-internally reflected contribution from the energy penetrating into the cylinder. Below the plasma frequency, the plasma is cutoff to wave propagation and the backscattering comes solely from strong exterior reflection. Above the plasma frequency, the plasma cylinder is partially penetrable and the once-internally reflected contribution is present. The once-internally reflected wave traverses twice the diameter of the cylinder through the dispersive plasma. Because the phase velocity of the wave in the plasma is proportional to the travel time, the curve in the wavelet plane in fact represents the dispersion diagram of the plasma. At the plasma frequency, the phase velocity approaches infinity. Physically, this corresponds to the electron resonance around the heavy ions in the plasma. As the frequency increases beyond the plasma frequency, the phase velocity approaches the speed of light in vacuum. In between the two regions the plasma is highly dispersive. The wavelet transform successfully resolves the different events with widely varying scales in the frequency domain. It is worthwhile to point out that the RWFT would not be able to simultaneously resolve the large-scale (corresponding to the initial peaks in time domain) and small-scale events (resonance and dispersion).

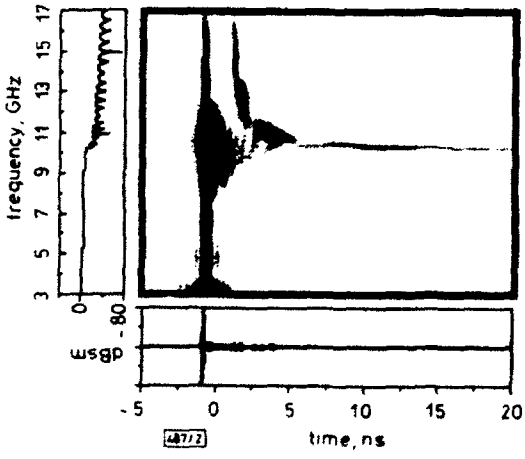


Fig. 2 *Time-frequency wavelet representation of backscatter data from uniform plasma cylinder*

Grey scale plots of intensity are in dB with dynamic range of 48 dB

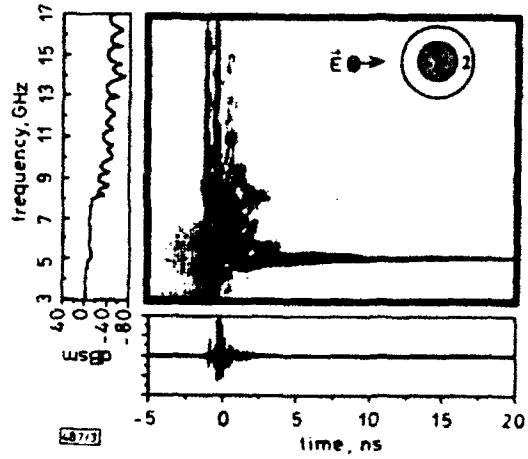


Fig. 3 *Time-frequency wavelet representation of backscatter data from two-layer plasma cylinder*

Inner layer has radius of 4·78 cm with f_p = 8 GHz and outer layer is of thickness 7·16 cm with f_p = 5 GHz
Grey scale plots of intensity are in dB with dynamic range of 52 dB
f_{p1} = 8 GHz, r_1 = 4·78 cm
f_{p2} = 5 GHz, r_2 = 11·94 cm

The result for a two-layer plasma cylinder is shown in Fig. 3. The inner layer has a radius of 4·78 cm with f_p = 8 GHz and the outer layer is of thickness 7·16 cm with f_p = 5 GHz. We observe three regions in the frequency response with additional small-scale features around the plasma frequencies. Again, the time-frequency representation using wavelets allows the detailed scattering mechanisms to be resolved.

Acknowledgment: This work is supported by the Joint Services Electronics Program and in part by National Science Foundation Grant ECS-8657524.

2nd December 1991

H. Kim and H. Ling (*Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084, USA*)

References

1 MOGHADDAR, A., and WALTON, E. K.: 'Time-frequency-distribution analysis of dispersive targets'. Workshop on High Frequency Electromagnetic Modeling of Jet Engine Cavities, Wright Laboratory, Wright-Patterson AFB, Ohio, USA, August 1st–2nd, 1991

2 MALLAT, S.: 'Multifrequency channel decompositions of images and wavelet models', *IEEE Trans.*, 1989, ASSP-37, pp. 2091–2110

3 DAUBECHIES, I.: 'The wavelet transform, time-frequency localization and signal analysis', *IEEE Trans.*, 1990, IT-36, pp. 961–1005

4 HEIL, C. E., and WALNUT, D. F.: 'Continuous and discrete wavelet transforms', *SIAM Review*, 1989, 31, pp. 628–666

5 COMBES, J. M., GROSSMANN, A., and TCHAMITCHIAN, PH. (Eds.): 'Wavelets, time-frequency methods and phase spaces' (Springer-Verlag, Berlin, 1989), pp. 2–20

6 LING, H., KIM, H., HALLOCK, G. A., BIRKNER, B. W., and ZAMAN, A. J. M.: 'Effect of an arcjet plume on satellite reflector performance', *IEEE Trans.*, 1991, AP-39, pp. 1412–1420

7 BUSSEY, H. E., and RICHMOND, J. H.: 'Scattering by a lossy dielectric circular cylindrical multi-layer-numerical values', *IEEE Trans.*, 1975, AP-23, pp. 723–725

8 AMOS, D. E.: 'A portable package for Bessel function of a complex argument and nonnegative order', *ACM Trans. Math. Software*, 1986, 12, pp. 265–273

CONDUCTOR-BACKED TELLEGEN SLAB AS TWIST POLARISER

I. V. Lindell, S. A. Tretyakov and M. I. Oksanen

Indexing terms: Polarisation, Electromagnetic waves

A conductor-backed slab of nonreciprocal bi-isotropic material with nonzero Tellegen parameter is shown to act as a twist polariser (crosspolarising reflector) when the medium parameters of the slab satisfy a certain condition, whence a linearly polarised incident plane wave is reflected as crosspolarised. This effect is seen to be broadbanded for large values and narrowbanded for low values of the Tellegen parameter and the operation is independent of the polarisation direction of the incident wave. Analysis of the Tellegen slab is outlined and the two operating modes for the twist polariser are discussed.

Introduction: Novel effects and devices have been recently suggested based on applications of chiral media, which are reciprocal special cases of the most general isotropic, or bi-isotropic (BI), media. The chiral parameter κ and the Tellegen parameter χ, which is responsible for nonreciprocity of the BI medium, appear in similar positions in the medium equations [1]

$$D = \varepsilon E + (\chi - j\kappa)\sqrt{(\mu_0 \varepsilon_0)}H \tag{1}$$

$$B = \mu H + (\chi + j\kappa)\sqrt{(\mu_0 \varepsilon_0)}E \tag{2}$$

For lossless BI media, the four parameters ε, μ, χ, κ are real quantities. In this Letter, it is demonstrated that a slab of BI medium with nonzero Tellegen parameter χ and backed by a perfectly conducting plane can be designed to rotate the incident linearly polarised field in reflection by 90°. This type of operation is basic in its action and its use in antenna and microwave devices has been described earlier [2–5] with structures that require the polarisation of the incident wave to be at a fixed angle to an axis of the structure.

An isotropic medium with a nonreciprocity parameter, denoted here by χ, was first suggested by Tellegen in 1948 [6] to produce a nonreciprocal device called gyrator. He also explained that to make such a medium in practice would require a microscopic structure of similar parallel permanent electric and magnetic dipoles. This possibility has also been considered by others [7, 8] and a method has been suggested which applies parallel electric and magnetic fields to certain substances such as CrO_2 while they are cooled to a temperature at which the electric and magnetic polarisations become permanent [9]. The problem of fabricating such a medium is not considered here in any more detail.

Reflection dyadic: Applying vector transmission-line analysis*, where any layered medium with homogeneous BI media and boundaries parallel to the xy plane can be replaced by a transmission line with dyadic characteristic impedances and dyadic propagation factors, the reflection dyadic, corresponding to normal incidence of plane waves, of a BI slab of thickness d and with a perfectly conducting backplane can be written in the dyadic form

$$\bar{R} = e^{j\psi}(\cos\phi \bar{I}_t + \sin\phi \bar{J}) = R_{co}\bar{I}_t + R_{cr}\bar{J} \tag{3}$$

where $\bar{I}_t = a_x a_x + a_y a_y$ denotes the two-dimensional unit dyadic and $\bar{J} = a_z \times \bar{I} = a_y a_x - a_x a_y$ the 90° rotator. R_{co} and R_{cr} are the, respective, reflection coefficients corresponding to the copolarised and crosspolarised components of the reflected field when the incident field is linearly polarised. The angles ψ and ϕ are real for a lossless slab, ψ gives the phase shift and ϕ the polarisation rotation, of the reflected wave with respect to the incoming wave. The rotation angle ϕ is positive in the right-hand direction when looking at the slab interface.

The dependence of ψ and ϕ on the parameters of the slab can be expressed as

$$\tan\psi = \frac{\eta\eta_0 \cos\nu \sin(2kd\cos\nu)}{(\eta^2 + \eta_0^2)\sin^2(kd\cos\nu) - \eta_0^2\cos^2\nu} \tag{4}$$

$$\tan\phi = \frac{2\eta\eta_0 \sin\nu \sin^2(kd\cos\nu)}{(\eta^2 - \eta_0^2)\sin^2(kd\cos\nu) + \eta_0^2\cos^2\nu} \tag{5}$$

We denote $\eta_0 = \sqrt{(\mu_0/\varepsilon_0)}$ and quantities of the BI medium by $\eta = \sqrt{(\mu/\varepsilon)}$, $k = nk_0 = \omega\sqrt{(\mu\varepsilon)}$, $n = \sqrt{(\mu\varepsilon/\mu_0\varepsilon_0)}$ and $\sin\nu = \chi/n = \chi_r$. χ_r is called the relative Tellegen parameter. It is seen that the chirality parameter κ does not affect the reflection dyadic at all in the present notation; with other formalisms this is not always true [1]. On the hand, the Tellegen parameter χ has an essential rôle in producing polarisation rotation. For a lossless slab with real ψ, eqn. 3 implies the relation $|R_{co}|^2 + |R_{cr}|^2 = 1$, or the sum of co- and crosspolarised reflected powers equals the incident power.

The nonreciprocal character of a medium with nonzero Tellegen parameter χ is seen most easily from the property of having nonzero R_{cr}. In fact, in trying to reverse the reflection, it is seen that an incident wave with the polarisation of the previously reflected wave would not in general reflect with the polarisation of the previously incident wave.

Twist polariser: Let us find the condition under which the copolarised reflection coefficient vanishes, in which case all the power incident with linear polarisation is reflected in crosspolarisation. From eqn. 3 we see that the condition $R_{co} = 0$ is satisfied when $\cos\phi = 0$, i.e. when the parameters kd, χ and η/η_0 are such that the denominator of $\tan\phi$ in eqn. 5 vanishes.

$$\sin(kd\cos\nu) = \pm\frac{\eta_0\cos\nu}{\sqrt{(\eta_0^2 - \eta^2)}} = \pm\sqrt{\left(\frac{1-\chi_r^2}{1-(\eta/\eta_0)^2}\right)} \tag{6}$$

This has real solutions for lossless BI media when the inequality $\eta/\eta_0 \leq |\sin\nu| = |\chi_r|$, or $\mu_r \leq |\chi|$ is satisfied with $\mu_r = \mu/\mu_0$. To have $R_{co} = 0$ and $R_{cr} \neq 0$, the Tellegen parameter obviously must not be zero. The crosspolarised reflection coefficient is from eqn. 3 simply the phase factor $R_{cr} = e^{j\psi}$, with $\cos\psi = \eta/\eta_0 \sin\nu = \mu_r/\chi$. Thus, the conductor-backed BI slab with parameters satisfying eqn. 6 acts as a crosspolarising reflector (twist polariser) creating a rotation of 90° and a phase shift ψ for the reflected field. The total field in front of the slab will in this case be elliptically polarised with the axial ratio $AR = \tan(\mu/2)$.

The twist polariser effect is dependent on the Tellegen parameter, which, in addition to the previous inequality must satisfy another one, $|\chi_r| \leq 1$, for the energy function to be positive semidefinite†.

There are two interesting modes of operation for the crosspolarising reflector: wide-band operation which damps the copolarised reflection in a wide frequency band and narrowband operation which suggests use in filtering narrow bands of signals.

In wideband operation, the value of the relative Tellegen parameter χ_r should be close to unity. In theory, for $\chi_r = 1$ there is an infinite band for which $|R_{co}|$ is smaller than a given number Δ, but, if Δ is small, the thickness d becomes large. Thus, for practical use, somewhat lower values are more suitable. As seen in Fig. 1, for $\eta/\eta_0 = 0.8$ and $\chi_r = 0.85$ we have

* LINDELL, I. V., TRETYAKOV, S. A., and OKSANEN, M. I.: 'Vector transmission-line and circuit theory for bi-isotropic layered structures', submitted to *J. Electro Magnetic Wave Appl.*

† LINDELL, I. V.: 'Methods for electromagnetic field analysis (Oxford: Clarendon Press). To be published

Wavelet Analysis of Backscattering Data from an Open-Ended Waveguide Cavity

Hao Ling, *Member, IEEE*, and Hyeongdong Kim, *Student Member, IEEE*

***Abstract*— The wavelet analysis technique is applied to analyze the electromagnetic backscattering data from an open-ended waveguide cavity. Compared to the conventional short-time Fourier transform, the wavelet transform results in a better representation of the scattering features in the time-frequency plane, due to its multiresolution property.**

I. Introduction

THE electromagnetic energy backscattered from an unknown target can provide information useful for classifying and identifying the target. This is commonly accomplished by interpreting the radar echo in either the time or the frequency domain. For target characteristics which are not immediately apparent in either the time or the frequency domain, the joint time-frequency representation of the radar echo can often provide more insight into the scattering mechanisms and, when properly interpreted, can lead to successful identification of the target. In a recent paper by Moghaddar and Walton [1], the joint time-frequency analysis of an open-ended waveguide cavity was carried out. In that work, both the short-time Fourier transform (STFT) and the Wigner–Ville distribution were applied to the broad-band backscattering data from a dispersive cavity structure to arrive at the time-frequency representation. Good insights on the scattering mechanisms were gained from the results. However, the STFT is limited by its fixed resolution in both the time and the frequency domain. The Wigner–Ville distribution, although providing good localization of scattering mechanisms, introduces additional cross terms which leads to "ghosts" in the time-frequency plane.

The theory of wavelets is currently attracting a great deal of attention in many disciplines of applied science [2]–[5]. In this letter, the wavelet transform is applied to the backscattering data from an open-ended waveguide cavity in order to derive the time-frequency representation of the signal. Contrary to the conventional STFT that has fixed resolution in both time and frequency, the wavelet transform, when properly defined, can provide variable resolution in time and multiresolution in frequency. Since the early-time radar echo from finite objects usually consists of sharp peaks [6], very fine time resolution is needed to resolve the various scattering centers. On the other hand, since the late-time arrivals are characterized by resonant ringing, good frequency resolution (or coarse time resolution) is needed for isolating the target resonances. In the intermediate region, dispersive phenomena require good resolution in both time and frequency. The multiple resolution property of the wavelet transform is ideally suited for this task. Consequently, the wavelet representation can provide a better time-frequency characterization of the backscattering data.

Manuscript received November 22, 1991. This work was supported by the Joint Services Electronics Program and in part by National Science Foundation Grant ECS-8657524.

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084.

IEEE Log Number 9107677.

II. Short-Time Fourier Transform and Wavelet Transform

The conventional STFT of a time signal $f(t)$ is defined as [3], [4]:

$$S(\tau,\Omega)=\int f(t)g(t-\tau)e^{-j\Omega t}dt. \tag{1a}$$

It is essentially the Fourier transform operation with the addition of a time window function $g(t)$. The translation of the window as a function of τ results in a two-dimensional time-frequency representation, $S(\tau,\Omega)$, of the original time function. By manipulating (1a), the STFT can also be expressed in the frequency domain:

$$S(\tau,\Omega)=e^{-j\Omega t}\int F(\omega)G(\Omega-\omega)e^{j\tau\omega}d\omega. \tag{1b}$$

Here $F(\omega)$ and $G(\omega)$ are the Fourier transforms of $f(t)$ and $g(t)$, respectively. We observe from (1a) and (1b) that the STFT representation can be obtained through either a moving window in time $g(t)$ or a corresponding moving window in frequency $G(\omega)$.

We will now introduce the (continuous) wavelet transform of a time signal $f(t)$ for our application:

$$W_f(\tau,\Omega)=\int f(t)\tau^{-1/2}h(t/\tau)e^{j\Omega\tau}dt. \tag{2a}$$

By comparing (1a) and (2a), we see that $h(t)$ is similar to the window function $g(t)$ in the STFT. However, $h(t)$ must satisfy an additional "admissibility condition" in wavelet theory [5], viz., $h(t=0)=0$. To satisfy this condition, $h(t)$ is usually chosen to be a translated window function with its center at t_0. By changing τ, the center of the window function moves as τt_0 and the width of the window is dilated by the scale factor τ. The ratio between the window width and the window center (or the Q-factor of the window function) remains fixed for all time. This is in contrast to the STFT where the window width is fixed for all time. By properly manipulating (2a), the wavelet transform can also be carried out on the Fourier transform $F(\omega)$ of the original time signal:

$$W_f(\tau,\Omega)=\int F(\omega)\tau^{1/2}H(\tau(\omega-\Omega))d\omega. \tag{2b}$$

1051-8207/92$03.00 © 1992 IEEE

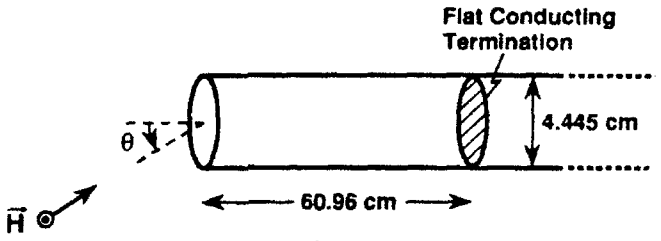


Fig. 1. Geometry of the open-ended circular waveguide cavity.

$H(\omega)$ is the Fourier transform of $h(t)$ and is usually referred to as the "mother wavelet." The operation in (2b) can be interpreted as the decomposition of the frequency signal $F(\omega)$ into a family of shifted and dilated wavelets $H(\tau(\omega - \Omega))$. It is important to point out here that the present definition of the wavelet transform in its time and frequency forms is exactly opposite to the common definition of the wavelet transform used in time-series signal analysis [5].

III. Time-Frequency Representation Of Backscattering Data

The time-frequency representation of the backscattering data from an open-ended cavity is considered. The cavity is an open-ended circular waveguide with a diameter of 4.445 cm. A flat conducting termination exists 60.96 cm inside the waveguide (Fig. 1). To generate the backscattering data, the radar cross section of this target is first computed in the frequency domain. We take into account of the interior cavity contribution using a modal approach [7] and the diffraction contribution from the front rim of the cavity using the asymptotic formula in [8]. It has been previously established that the backscattering data predicted in this manner agree reasonably well with experiments. The time-domain response is then obtained by Fourier transforming the band-limited frequency data (from 2 to 18 GHz).

Fig. 2(a) shows the time-frequency plot of the backscattering data at normal incidence ($\theta = 0°$) using the STFT. In performing the STFT, a 2-GHz Kaiser–Bessel window in the frequency domain is used in equation (1b). Also plotted along the two axes are the time-domain and the frequency-domain responses. It is apparent that the scattering features are much better resolved in the time-frequency domain than in either the time or the frequency domain alone. Both the nondispersive rim diffraction and the two mode spectra due to the TE_{11} (with cutoff at 3.96 GHz) and the TE_{12} (11.45 GHz) mode can be clearly identified. The mode spectra are in fact dispersion curves of the waveguide modes since the phase velocity of each mode is proportional to the travel time. The noncausal noise appearing in the time-frequency plot is caused by the modal approximation used in simulating the backscattering data. If actual measurement data were used, this noise should be absent.

Due to the fixed resolution of the STFT, the scattering features in Fig. 2(a) are smeared out in the time-frequency plane. This problem is overcome by using the wavelet transform that provides a much better representation of the scattering features in the time-frequency plane, as shown in Fig. 2(b). The wavelet

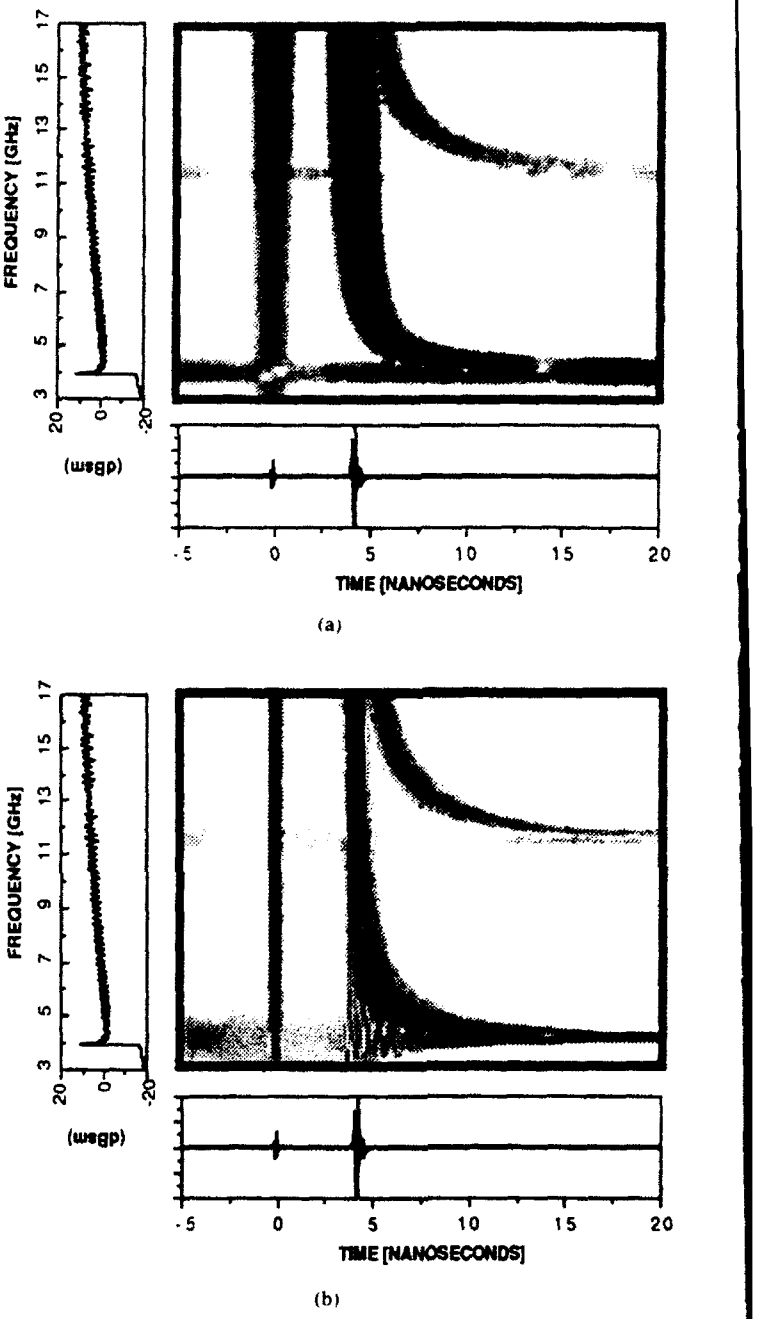


Fig. 2. Time-frequency representation of backscattering data from an open-ended cavity under normal incidence. The grayscale plots of intensity are in decibels with a dynamic range of 40 dB. (a) Short-time Fourier transform (STFT) representation. (b) Wavelet transform representation.

transform is implemented using equation (2a) with the aid of the FFT. The function $h(t)$ is chosen to be a two-sided Kaiser–Bessel window with a Q-factor of 0.3. The $t = 0$ reference of $h(t)$ is located midway between the time events from the rim diffraction and interior contribution (at $\tau = 2.05$ ns). The variable time resolution of the wavelet transform allows sharper time resolution to be achieved during the early-time response and sharper frequency resolution (coarser time resolution) to be achieved during the late-time response. Thus, wavelet transform provides good resolution in identifying the scattering centers and resolving the resonant phenomena of the target while adequately describing the dispersive scattering

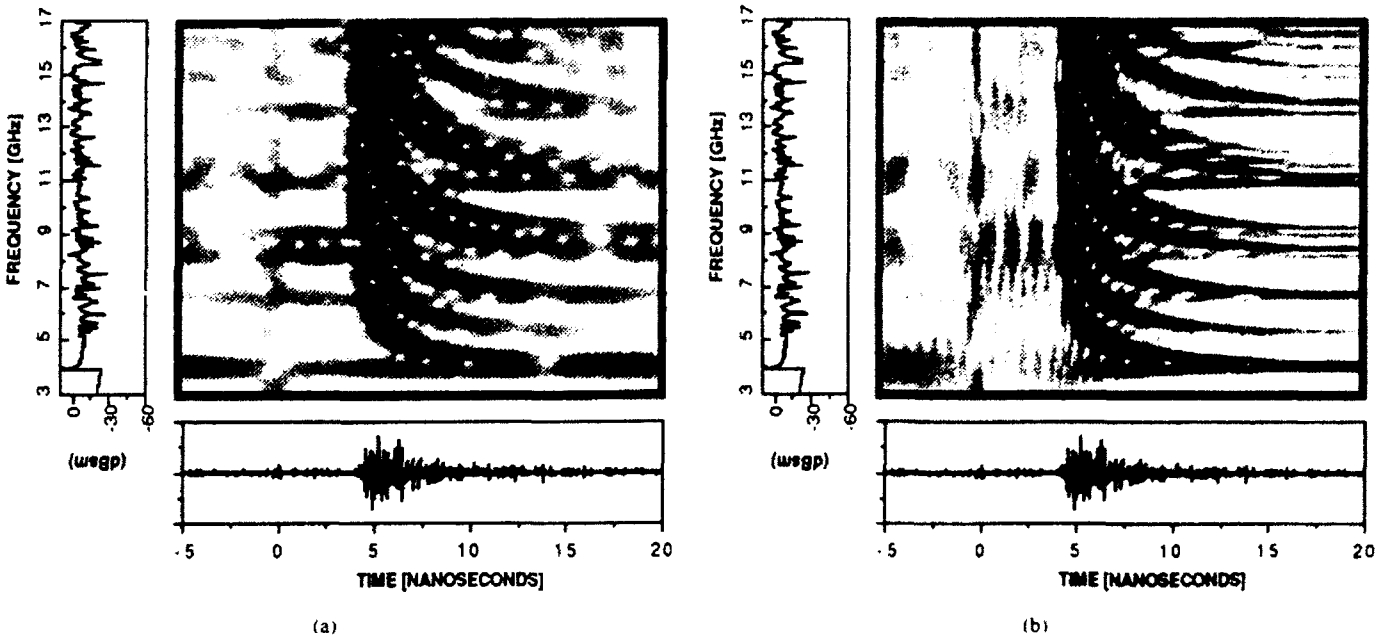


Fig. 3. Time-frequency representation of backscattering data from an open-ended cavity under 45° incidence. The grayscale plots of intensity are in decibels with a dynamic range of 40 dB. (a) Short-time Fourier transform (STFT) representation. (b) Wavelet transform representation.

mechanisms in the intermediate-time region. Figs. 3(a) and 3(b) show the time-frequency plots, generated using the STFT and the wavelet transform, respectively, of the same cavity at 45° incidence. Many more modes are excited by the obliquely incident wave. Consequently the time domain response is much more dispersive. By comparing Figs. 3(a) and 3(b), we find that the wavelet representation again provides a much sharper resolution of the different scattering mechanisms than the STFT.

ACKNOWLEDGMENT

The authors would like to thank Profs. F. X. Bostick and C. M. Maziar for bringing the topic of wavelets to our attention and for many helpful discussions.

REFERENCES

[1] A. Moghaddar and E. K. Walton, "Time-frequency-distribution analysis of dispersive targets," *Workshop on High-Frequency Electromagnetic Modeling of Jet Engine Cavities*, Wright Laboratory, Wright-Patterson AFB, OH, Aug. 1–2, 1991. Also to appear in *IEEE Trans. Antennas Propagat.*

[2] C. E. Heil and D. F. Walnut, "Continuous and discrete wavelet transforms," *SIAM Rev.*, vol. 31, pp. 628–666, Dec. 1989.

[3] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 2091–2110, Dec. 1989.

[4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, Sept. 1990.

[5] J. M. Combes, A. Grossmann and Ph. Tchamitchian, Eds., *Wavelets, Time-Frequency Methods and Phase Spaces*. Berlin: Springer-Verlag, 1989, pp. 2–20.

[6] L. B. Felsen, "Progressing and oscillatory waves for hybrid synthesis of source excited propagation and diffraction," *IEEE Trans. Antennas Propa.*, vol. 32, pp. 775–796, Aug. 1984.

[7] H. Ling, W. Lee, and R. Chou, "High-frequency RCS of open cavities with rectangular and circular cross sections," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 648–654, May 1989.

[8] J. J. Bowman, S. W. Lee, and C. Liang, "High-frequency backscattering from a semi-infinite hollow cylinder," *Proc. IEEE*, vol. 61, pp. 681–682, May 1973.

received signal levels of antennas (d) and (e) are the same as that of antenna (a), i.e., the $\lambda_0/2$ dipole antenna. The loaded monopole mobile antenna has also been tested, and its received signal level is about 6 dB lower. These facts justify the applicability of antennas (d) and (e), i.e., the meander line PCDAs.

4. CONCLUSION

A meander line PCDA of 9.2-cm length with 45.9% SR and 5% bandwidth of 2:1 VSWR has been designed for PCN handset usage operating at center frequency f_0 = 880 MHz. Owing to its planar, lightweight, thin, and small characteristics, the PCDA using meander line is well suited for signal transmission and reception of the PCN handset, in which the PCDA may be built in, and various antenna diversity schemes may be used for fading reduction. If the shortened PCDA is still too long to be accepted, then the monopole version of the same configuration may be constructed to get half a length shorter, but at the price of lower radiation efficiency.

ACKNOWLEGMENT

The authors would like to express deep gratitude to Mr. Chiu Wu-Jyh and Mr. Shi Mu-Biau of Telecommunication Laboratories, Taiwan, Republic of China, for their enthusiastic support for the research of the PCN handset antenna. Finally, special thanks are delivered to Dr. Chu Tah-Hsiung of the National Taiwan University, Republic of China, for his careful review of this article. Without their valuable assistance this article would not have been possible.

REFERENCES

1. I. S. Froves, "Personal Mobile Communication—A Vision of the Future," *Br. Telecommun. Tech. J.*, Vol. 8, No. 1, Jan. 1990, p. 7–11.
2. R. Steel, "Deploying Personal Communication Networks," *IEEE Communication Magazine*, Sept. 1990, pp. 12–15.
3. D. C. Cox, "Personal Communication—A Viewpoint," *IEEE Communication Magazine*, Nov. 1990, pp. 8–20.
4. T. S. Rappaport, "Wireless Personal Communications: Trends and Challenges," *IEEE Antennas Propagat. Magazine*, Vol. 33, No. 5, Oct. 1991, pp. 19–29.
5. H. Kuboyama, Y. Tanaka, K. Fujimoto, and K. Hirasawa, "Experimental Results in Urban and Rural Areas," *IEEE Trans. Veh. Technol.*, Vol. VT-39, No. 2, May 1990, pp. 150–160.
6. D. G. Brennan, "Linear Diversity Combining Techniques," *Proc. IRE.*, Vol. 47, Jan. 1959, pp. 1075–1102.
7. W C.-Y. Lee and Y. S. Yeh, "Polarization Diversity System for Mobile Radio," *IEEE Trans. Commun.*, Vol. COM-20, No. 5, Oct. 1972, pp. 912–923.
8. S. Kozono, T. Tsuruhara, and M. Sakamoto, "Base Station Polarization Diversity Reception for Mobile Radio," *IEEE Trans. Veh. Technol.*, Vol. VT-33, No. 4, Nov. 1984, pp. 301–306.
9. R. G. Vaughan and J. B. Andersen, "Antenna Diversity in Mobile Communication," *IEEE Trans. Veh. Technol.*, Vol. VT-36, No. 4, Nov. 1987, pp. 149–172.
10. R. G. Vaughan, "Polarization Diversity in Mobile Communications," *IEEE Trans. Veh. Technol.*, Vol. VT-39, No. 3, Aug. 1990, pp. 177–186.
11. Y. Ebine and Y. Yamada, "A Vehicular Mounted Vertical Space Diversity Antenna for a Land Mobile Radio," *IEEE Trans. Veh. Technol.*, Vol. VT-40, No. 2, May 1991, pp. 420–425.
12. Special issue on mobile communication antennas, *IEICE Trans. Commun.*, Vol. E74, No. 10, Oct. 1991, Japan.
13. J. R. James, P. S. Hall, and C. Wood, *Microstrop Antenna: Theory and Design*, Peter Peregrinus, New York, 1981.
14. G. Dubost, *Flat Radiating Dipoles and Applications to Arrays*, Research Studies, New York, 1981.
15. H. Oltman and D. A. Huebner, "Electromagnetically Coupled Microstrip Dipoles," *IEEE Trans. Antennas Propagat.*, Vol. AP-29, No. 1, Jan. 1981, pp. 151–157.
16. H. Nakano, S. R. Kerner, and N. Alexopoulos, "Moment Method Solution for Printed Wire Antennas of Arbitrary Configuration," *IEEE Trans. Antennas Propagat.*, Vol. AP-36, No. 12, Dec. 1988, pp. 1667–1674.
17. H. Ragheb and L. Shafai, "Analysis of Arbitrary Shape Printed Line Microstrip Antennas," *IEEE Trans. Antennas Propagat.*, Vol. AP-38, No. 2, Feb. 1990, pp. 269–274.
18. J. Rashed and C. T. Tai, "A New Class of Resonant Antennas," *IEEE Trans. Antennas Propagat.*, Vol. AP-39, No. 9, Sept. 1991, pp. 1428–1430.
19. H. Nakano, H. Tagami, A. Yoshizawa, and J. Yamauchi, "Shortening Ratios of Modified Dipole Antennas," *IEEE Trans. Antennas Propagat.*, Vol. AP-32, No. 4, April 1984, pp. 385–386.
20. S. Silver (Ed.), *Microwave Antenna Theory and Design*, McGraw-Hill, New York, 1949, p. 245.

Received 4-1-92

Microwave and Optical Technology Letters, 5/10, 477–480

© 1992 John Wiley & Sons, Inc.

CCC 0895-2477/92/$4.00

A BOUNDARY INTEGRAL APPROACH TO THE SCATTERING FROM PERIODIC GRATINGS

John Moore and Hao Ling
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, Texas 78712

Urs U. Graf and Daniel T. Jaffe
Department of Astronomy
The University of Texas at Austin
Austin, Texas 78712

KEY TERMS

Scattering, periodic gratings, infrared

ABSTRACT

A boundary integral formulation is presented for computing the scattering from infinite periodic gratings. Periodic boundary conditions and expansion of the fields in terms of Floquet harmonics allow the boundary integral formulation to treat just a single period of the infinite grating. Numerical results for a blazed grating designed for use at infrared wavelengths are presented and shown to be in good agreement with a series of measurements. © *1992 John Wiley & Sons, Inc.*

I. INTRODUCTION

Periodic gratings have long attracted a great deal of interest from both the optics and microwave communities. To this day, they remain a popular choice in a variety of applications, including frequency scanning, angular sensitivity, and polarization sensitivity. Accurate high-frequency solutions to the problem of diffraction from gratings did not become available until the arrival of digital computers. Meecham was the first among many to apply a variational method to the study of perfectly conducting gratings [1]. Later on, Nevière, Cadilhac, and Petit devised a conformal-mapping technique to analyze dielectric gratings [2]. Recent studies of gratings have focused

on design optimization [3] and the characterization of lossy absorbers under oblique incidence [4].

In this article we present a numerical approach to the study of two-dimensional periodic gratings. We first use the boundary integral method to characterize the rather complex contour of a single groove (or period) of the grating. The underlying periodicity of the infinite structure is then accounted for by applying periodic boundary conditions and expanding the fields in terms of Floquet harmonics. Details of this approach are presented next, followed by a presentation of both numerical and experimental results.

II. FORMULATION

The geometry of the grating shown in Figure 1 is a blazed grating used for frequency scanning purposes. In this example, the triangular grooves are spaced a length p apart in the x direction. In addition, a narrow flat spacing of width w is included. A plane wave of either polarization is incident upon the infinite grating, resulting in energy being diffracted in many directions. In general, the period of the grating, p, determines the number of beams or modes of diffracted energy, and the shape of the groove determines the amount of energy diffracted into each mode. Only the two-dimensional case is considered, meaning that neither the structure nor the incident field has z variation. The fundamental task here is to determine the scattered field due to a plane wave incident upon the grating, and from this, evaluate the performance of the grating.

Because all the grooves of the grating are equally spaced a length p apart, it is sufficient to consider just one period of the whole structure. Shown in Figure 2 is a single period of the grating. The approach taken here is to subdivide this single period of the grating into a top region containing just free space and a bottom region containing the contour of the groove. Each of the two regions will be handled separately. For the sake of brevity, only the H-polarization case (magnetic field parallel to the z axis) is described here. The E-polarization formulation closely parallels that of the H polarization.

In Region I, the magnetic field H_z^{I} along the aperture boundary can be expressed in terms of the tangential electric field ($\mathbf{E} \times \hat{y}$) [or equivalently, a surface magnetic current $M_z = \hat{z} \cdot (\mathbf{E} \times \hat{y})$]. Due to the plane-wave incidence and the periodicity of the structure, the expression for H_z^{I} can be written into a sum of Floquet harmonics (or plane-wave spectra) as follows:

$$H_z^{\mathrm{I}}(x) = 2H_z^{\mathrm{inc}}(x) - \sum_{n=-\infty}^{\infty} e^{-j(\beta_x^{\mathrm{inc}} + 2\pi n/p)x} \times \int_0^p \frac{\omega\epsilon}{k_{y_n} p} e^{j(\beta_x^{\mathrm{inc}} + 2\pi n/p)x'} M_z(x')\, dx'. \quad (1)$$

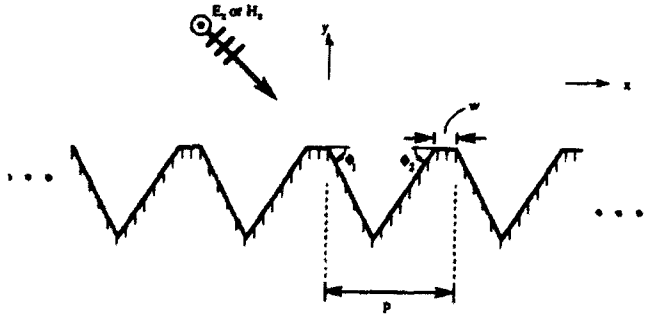


Figure 1 Profile of an infinite grating with triangular grooves

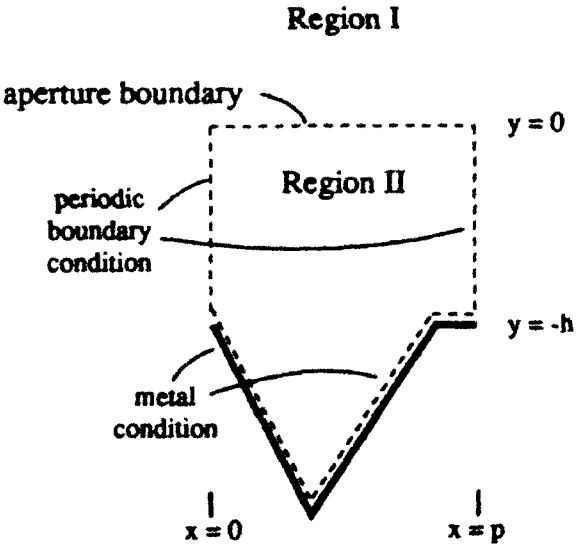


Figure 2 The boundary integral regions for a single period of the grating

In Eq. (1), H_z^{inc} is the magnetic field of the incident plane wave along the aperture boundary, β_x^{inc} is the x-component of the incident wave vector, and $k_{y_n} = \sqrt{k^2 - (\beta_x^{\mathrm{inc}} + 2\pi n/p)^2}$. By applying the standard method-of-moments procedure, Eq. (1) can be expanded via pulse functions and point testing into a matrix equation:

$$[H_z^{\mathrm{I}}] = [2H_z^{\mathrm{inc}}] + [Y_{\mathrm{ap}}^{\mathrm{I}}][M_z], \quad (2)$$

where $[Y_{\mathrm{ap}}^{\mathrm{I}}]$ is an aperture admittance matrix containing the contributions due to the summation in (1).

Region II, unlike Region I, is completely enclosed by the aperture boundary and a number of other boundaries shown in Figure 2. Thus, through the equivalence principle [5], the fields at any point along the boundary may be written in terms of an integral over the tangential fields (or the equivalent surface currents) along the entire boundary:

$$H_z^{\mathrm{II}}(Q) = \int_{\substack{\text{entire} \\ \text{boundary}}} \times \left[\frac{-jk}{2} \cos\theta\, H_1^{(2)}(kr) H_z^{\mathrm{II}} - \frac{\omega\epsilon}{2} H_0^{(2)}(kr) M_z \right] ds. \quad (3)$$

Here $H_z^{\mathrm{II}}(Q)$ is the z-directed magnetic field at a point Q on the boundary, r is the distance between the source on the boundary and point Q, and θ is the incline angle of Q with respect to the inward-pointing normal $\hat{n}$ at the source point. Equation (3) can also be expanded via pulse functions and point testing into a matrix equation and converted to an admittance form:

$$[H_z^{\mathrm{II}}] = [Y^{\mathrm{II}}][M_z], \quad \text{along the entire boundary of Region II.} \quad (4)$$

Because of the additional boundaries comprising Region II, (4) contains more unknowns than (2). Equation (4) can

be reduced to include just those points along the aperture boundary by enforcing the following boundary conditions:

$$M_z = 0, \quad \text{along conducting wall of groove.} \tag{5a}$$

$$H_z(0, y) = H_z(p, y)e^{-j\beta_0^{inc}p}, \quad \text{along the open boundaries.} \tag{5b}$$

$$M_z(0, y) = -M_z(p, y)e^{-j\beta_0^{inc}p}, \quad \text{along the open boundaries.} \tag{5c}$$

(5b) and (5c) are the so-called periodic boundary conditions. When the conditions in (5) are substituted into (4), an admittance relation between the aperture magnetic field and aperture magnetic current is rendered:

$$[H_z^{II}] = [Y_{ap}^{II}][M_z], \tag{6}$$

where $[H_z^{II}]$ and $[M_z]$ are the magnetic field and equivalent surface magnetic currents along the aperture boundary of Region II.

The two vertical boundaries along which the periodic boundary conditions of (5b) and (5c) are applied have been included to enable the E-polarization case to be studied without resorting to higher-order basis functions. Choosing the aperture plane above the grating also results in improved numerical stability.

With the admittance relation for each of the two regions in hand, the unknown currents along the aperture can be found. The total tangential fields along the aperture boundary of Region I are matched to those of Region II. Performing this match using (2) and (6) and taking into account the change of sign between the magnetic currents just above and below the aperture boundary results in

$$M_z = ([Y_{ap}^{II}] + [Y_{ap}^{I}])^{-1}[2H_z^{inc}]. \tag{7}$$

The strengths of the various Floquet harmonics excited by the grating may then be extracted from these currents.

III. RESULTS AND CONCLUSIONS

The formulation presented in the previous section was validated extensively for blazed gratings. One means of observing the frequency scanning characteristics of a blazed grating is through its retroreflection efficiency curves for various diffracted orders. The efficiency of a given order is the fraction of the total incident energy contained in that order. Retroreflection (or backscattering) occurs when the angular deviation between the incident wave and the diffracted beam is zero. Specifically, the angle of retroreflection for a certain order is a function of wavelength and is given by $\theta = \sin^{-1}[(-n/2)/(p/\lambda)]$, where n is the diffracted order, λ is the free-space wavelength, and θ is measured from the normal.

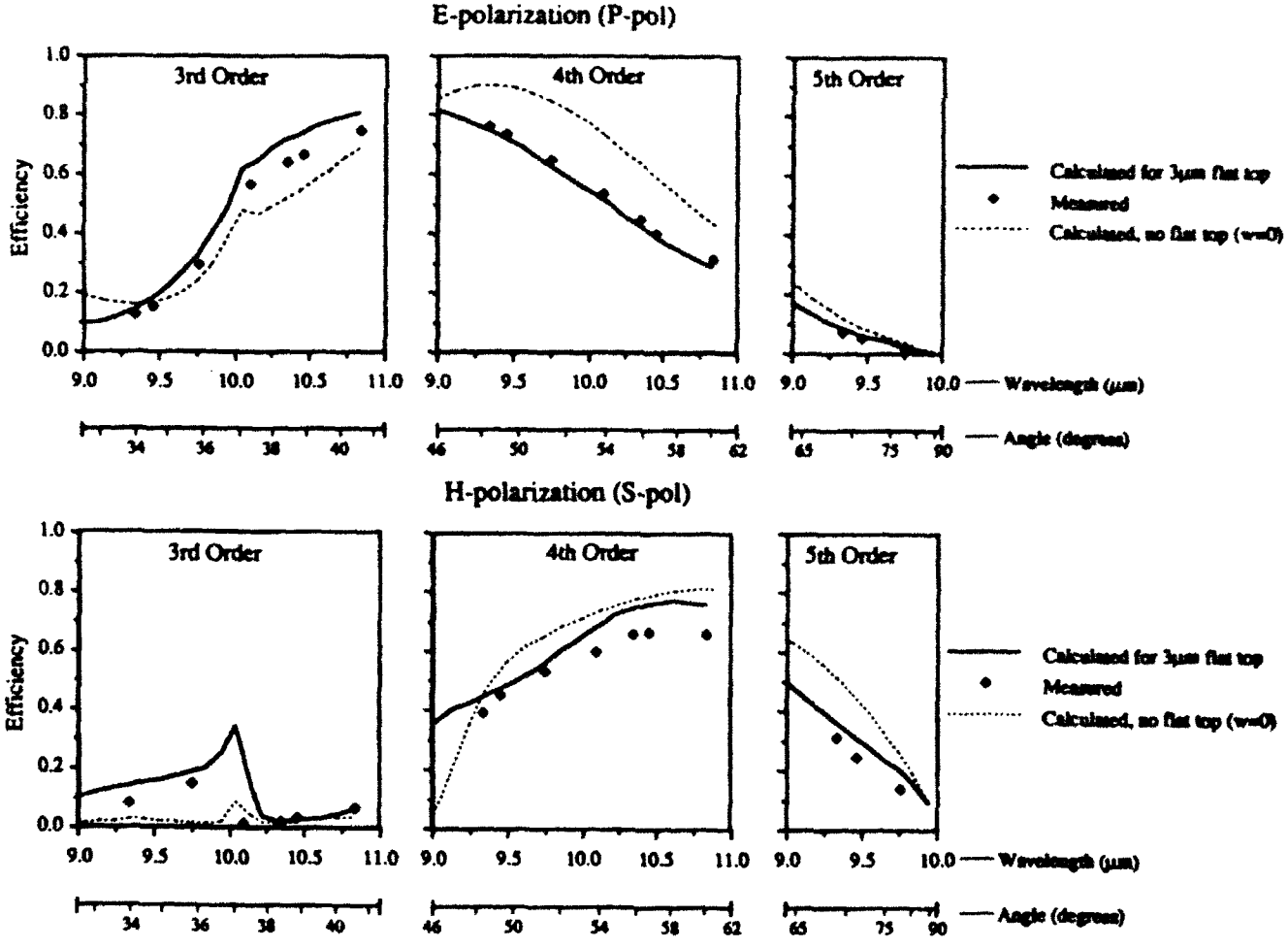


Figure 3 Retroreflection efficiency curves for the triangular grating: (a) E polarization, and (b) H polarization ($\phi_1 = \phi_2 = 54.736°$, $p = 25\ \mu m$)

Excellent agreement is observed between our calculated results and published data in [7] for gratings of various blaze angles for the $n = -1$ order.

Next the calculated results were compared against a set of measurements for a blazed grating designed for infrared spectroscopy. The grating was fabricated by first etching triangular grooves into a silicon substrate and then depositing a thin gold coating onto the etched surface. The nominal dimensions of the grating are $p = 25\ \mu m$ and $\phi_1 = \phi_2 = 54.736°$. Because of the fabrication process for this particular grating, there is a small flat spacing of about $w = 3\ \mu m$ between each groove. Shown in Figure 3 are the retroreflection efficiency curves for both polarizations. Wavelength is restricted to between 9.0 and 11.0 μm, and three diffracted orders are observed ($n = -3, -4, -5$). This corresponds to retroreflection scan angles lying between 34° and 78° from the normal. Agreement between the predicted and measured results is quite good for both polarizations. In particular, agreement for the E-polarization case holds well over a large range of scan angles. To serve as a contrast, the results for an ideal grating (i.e., $w = 0$) are also included in Figure 3. The flexibility of the boundary integral approach and its reasonable computational efficiency in solving periodic structures allow it to serve as a useful tool in the evaluation of novel grating structures.

ACKNOWLEDGMENT

This work is supported by the Air Force Rome Laboratory through the Joint Services Electronics Program and in part by the National Science Foundation (ECS-8657524). We want to thank Gabriel Rebeiz and Ed Kim, who produced the test sample. Dr. Jaffe and Dr. Graf acknowledge support by the David and Lucile Packard Foundation and the W. M. Keck Foundation. The work of Dr. Graf was enabled in part by a NATO fellowship awarded through the German Academic Exchange Service.

REFERENCES

1. W. C. Meecham, "Variational Method for the Calculation of the Distribution of Energy Reflected from a Periodic Surface," *J. Appl. Phys.*, Vol. 27, April 1956, pp. 361–367.
2. M. Nevière, M. Cadilhac, and R. Petit, "Applications of Conformal Mappings to the Diffraction of Electromagnetic Waves by a Grating," *IEEE Trans. Antennas Propagat.*, Vol. AP-21, Jan. 1973, pp. 37–46.
3. F. S. Johansson, "Frequency Scanned Gratings Consisting of Photo-Etched Arrays," *IEEE Trans. Antennas Propagat.*, Vol. AP-37, Aug. 1989, pp. 996–1002.
4. R. Janaswamy, "Oblique Scattering from Lossy Periodic Surfaces with Application to Anechoic Chamber Absorbers," *IEEE Trans. Antennas Propagat.*, Vol. AP-40, Feb. 1992, pp. 162–169.
5. R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill, New York, 1961.
6. H. Ling, "RCS of Waveguide Cavities: A Hybrid Boundary-Integral/Modal Approach," *IEEE Trans. Antennas Propagat.*, Vol. AP-38, Sept. 1989, pp. 1413–1420.
7. E. G. Loewen, M. Nevière, and D. Maystre, "Grating Efficiency Theory as it Applies to Blazed and Holographic Gratings," *Appl. Opt.*, Vol. 16, Oct. 1977, pp. 2711–2721

Received 4-22-92

Microwave and Optical Technology Letters, 5/10, 480–483

© 1992 John Wiley & Sons, Inc.

CCC 0895-2477/92/$4.00

PHASE GRATING: ANALYTICAL FORMULAS FOR THE NEAR FIELD

Victor Arrizón* and J. Ojeda-Castañeda
Instituto Nacional de Astrofísica, Optica y Electrónica
Apdo. Postal 216, Puebla 72000
Pue., México

KEY TERMS

Phase gratings, array illuminators, irradiance shaping, optical interconnects, diffractive optics

ABSTRACT

We present two simple analytical formulas that describe the irradiance distribution, at certain distances, of any phase grating. Our general results are illustrated with two novel examples: the irradiance distribution at the near field of a low-frequency blazé grating, and that of a phase grating with a triangular profile. © *1992 John Wiley & Sons, Inc.*

1. INTRODUCTION

Diffractive optical elements are powerful tools for addressing light beams into several prespecified directions. This addressing capability can be performed with substantial reduction in size, weight, and cost. Furthermore, by employing phase structures, the diffractive elements can work with high light efficiency.

Consequently, phase gratings find many useful applications in modern optical instrumentation, for example, in optical scanning, optical interconnects [1], in setting array illuminators [2, 3], in optical interferometers [4], and for optical decoding [5].

Our aim here is to present two simple, analytical formulas that describe the irradiance distribution at certain distances from any phase grating. Our general results are applied to discuss the irradiance distributions created by the use of a low-frequency, blazé grating, and by the use of a phase grating with triangular phase profile.

In Section 2, we discuss some basic results of diffraction theory. In Section 3, we present our analytical formulas. And in Section 4, we discuss two novel examples.

2. DIFFRACTION THEORY

As is shown in Figure 1, we assume that the phase grating is illuminated by a plane wavefront that propagates along the z axis. Thus, the amplitude just behind the grating is equal to its amplitude transmittance, which is written as

$$t(x, y) = \exp[i\phi(x)]; \qquad = \sum_{m=-\infty}^{\infty} C_m \exp(i2\pi m x/d). \tag{1}$$

In Eq. (1), the phase profile is denoted as $\phi(x)$, and d is the fundamental period of the phase grating.

Under the paraxial approximation, the wavefield behind the phase grating can be expressed as

$$u(x, y, z) = \sum_{m=-\infty}^{\infty} C_m F(m, z) \exp(i2\pi m x/d) \tag{2}$$

*Permanent address: Lab. de Semiconductores, UAP, Apdo. Postal 1651, Puebla, 72000, Pue., México.

International IEEE AP-S Symposium, pp. 305-308, Chicago, IL, July 1992.

NUMERICAL DIFFRACTION COEFFICIENTS FOR THE CONDUCTOR-BACKED DIELECTRIC HALF-PLANE

John Moore* and Hao Ling
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712-1084

INTRODUCTION

The computation of numerical diffraction coefficients, proposed by Hermann [1], holds potential as a means of extending the applicability of the Uniform Theory of Diffraction (UTD) to more general problems. Working in a UTD framework allows for a "building block" approach to characterizing the scattering from complex targets and through this approach provides a clear physical interpretation of the various scattering mechanisms and their interactions. Recently, Hurst [2] extracted the numerical diffraction coefficients of a semi-infinite strip and used the results to characterize the scattering from finite periodic strip gratings.

Scattering from dielectric coated structures is presently of much interest because of the important role played by these structures in radar cross section (RCS) reduction techniques. Use of the impedance boundary condition (IBC) approximation and more recently the generalized impedance boundary condition (GIBC) approximation in the study of coated structures [3-8] has met with much success when applied to thin, high contrast coatings. Though the IBC has proven itself rather robust, its applicability to more general coatings still needs to be validated.

In this paper, a method is presented for extracting the numerical diffraction coefficients for coated edges and material joints. Without resorting to an IBC approximation, an exact boundary integral formulation of the coated edge is used as the basis of numerical coefficient extraction. Contrary to previous work [1,2] in which the diffraction coefficients were extracted by solving finite problems containing multiple scattering centers, we solve the semi-infinite problem. This is accomplished by taking into account the non-decaying physical optics (PO) and surface wave fields such that the final discretization domain involves only a finite region near the diffraction center. Numerical diffraction coefficients are extracted and then employed within a UTD framework to construct solutions to finite sized coated structures. Agreement between this solution and an exact moment method solution for the finite coated structures suggests that this method is a viable approach to the detailed study of the complex scattering physics occurring in coated structures.

BOUNDARY INTEGRAL METHOD

The coated strip configuration is shown in Fig. 1. A coating with material parameters (μ,ε) and thickness d exists on top of an infinitely thin perfectly conducting strip. Our approach entails first dividing the problem geometry into three separate regions (see Fig. 2). By applying the equivalence principle, the field in each region can be related to the tangential fields along the boundary. The resulting set of boundary integral equations can be discretized into a set of matrix equations via the method of moments. Since the boundaries of all three regions extend to infinity, it is necessary to account for the PO fields and surface wave fields (if present) as specific components of the total fields. By placing a couple of semi-infinite bases along each boundary to describe the PO and surface wave fields, the semi-infinite boundaries may be truncated at a finite distance from the edge of the strip where the edge-diffracted fields become negligible.

NUMERICAL DIFFRACTION COEFFICIENTS

Once the boundary fields are found by solving the set of matrix equations, the scattered field is easily determined. The diffracted field is then related to the boundary integral scattered field depending on whether the scattered field at a certain observation

point contains the geometrical optics (GO) incident and reflected fields or surface wave fields. Uniform diffraction coefficients are then calculated by measuring the diffracted field at two different distances for the same incident and observation angle. Each pair of diffracted fields is then fitted to an equation containing two Fresnel transition functions (one corresponding to each shadow boundary) to yield a uniform diffraction coefficient. Surface wave launch coefficients are also found.

RESULTS

Numerical diffraction coefficients were first generated for the uncoated conducting strip. Fig. 3 shows that for both polarizations the numerical diffraction coefficients (NUTD) are in good agreement with those calculated via UTD. A strip with a relatively thick coating ($d=0.2\lambda_m$) was studied next. Numerical diffraction coefficients for this case were found and then used within a UTD framework to predict the diffraction from the finite-width dielectric coated strip shown in Fig. 4 ($w=5\lambda_m$). For E-polarization, only first-order diffractions from each of the two edges are used, and the backscattering echowidth for this case is in good agreement with an exact moment method solution shown in Fig. 5. A UTD solution which used a GIBC approximation [6] did not yield good results for this particular coating and is not shown. The H-polarization case for this coated strip consists of both multiple diffractions along the back side of the strip and surface wave coupling along the front of the strip in addition to the single diffractions from the two edges. Fig. 6(a) shows that considering only the first order diffractions and all multiple diffractions along the back of the strip does not agree well with the exact solution. But by also including the surface wave contributions, as shown in Fig. 6(b), agreement with the exact solution becomes quite good. This method has shown success in predicting not only the diffraction coefficients, but also the surface wave launch coefficients for the semi-infinite coated strip. The same approach can be used to study more complex material joints and their contributions to the overall backscattering from coated structures.

REFERENCES

1. Herrmann, G. F., "Numerical computation of diffraction coefficients," *IEEE Trans. Antennas Propagat.*, vol. AP-35, pp. 53-61, January 1987.
2. Hurst, M. P., "An improved method of calculating numerical diffraction coefficients," *1991 North American Radio Science Meeting, Symposium Digest*, p. 179, June 1991.
3. Senior, T. B. A., "Scattering by resistive strips," *Radio Sci.*, vol. 14, pp. 911-924, September 1979.
4. Volakis, J. L. and T. B. A. Senior, "Application of a class of generalized boundary conditions to scattering by a metal-backed dielectric half-plane," *Proc. IEEE*, vol. 77, pp. 796-804, May 1989.
5. Rojas, R. G. and Z. Al-hekail, "Generalized impedance/resistive boundary conditions for electromagnetic scattering problems," *Radio Sci.*, vol. 24, pp. 1-12, January 1989.
6. Bernard, J. M. L., "Diffraction by a metallic wedge covered with a dielectric material," *J. Wave Motion*, vol. 9, pp. 543-561, 1987.
7. Tiberio, R., G. Pelosi and G. Manara, "A uniform GTD formulation for the diffraction by a wedge with impedance faces," *IEEE Trans. Antennas Propagat.*, vol. AP-33, pp. 867-873, August 1985.
8. Peters, M. E. and E. H. Newman, "TM scattering by a general sheet impedance discontinuity," *IEEE Trans. Antennas Propagat.*, vol. AP-39, pp. 359-366, March 1991.

ACKNOWLEDGEMENT

This work is supported by the Air Force Rome Laboratory through the Joint Services Electronics Program and in part by the National Science Foundation (ECS-8657524).

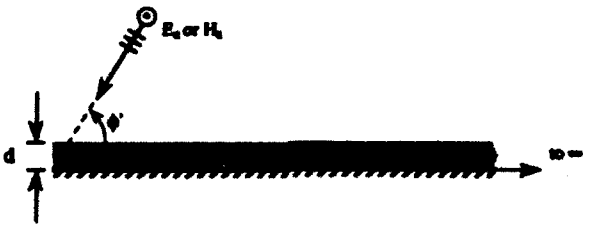


Fig. 1. Geometry of the conductor-backed dielectric strip

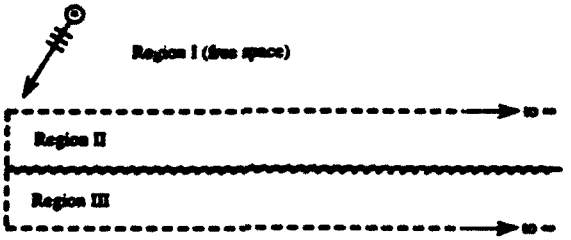


Fig. 2. The boundary integral regions for conductor-backed dielectric strip

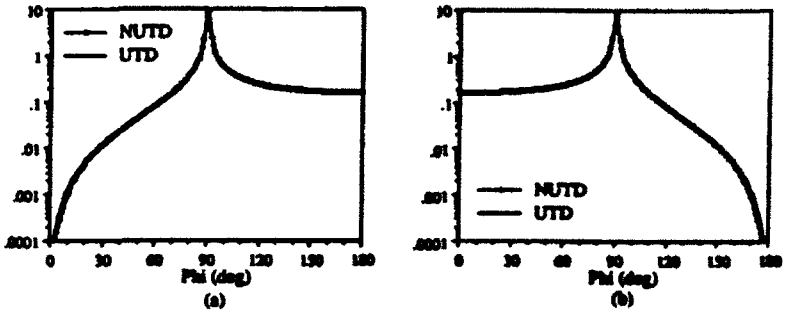


Fig. 3. Diffraction coefficient (Absolute magnitude) for back-scattering from perfect conducting strip (a) E-polarization (b) H-polarization

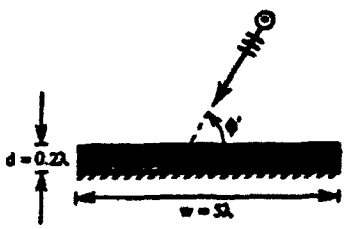


Fig. 4. Geometry of the finite-width conductor-backed dielectric strip

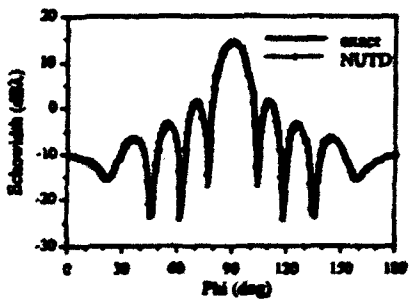


Fig. 5. E-pol back-scanning from a dielectric strip (d=0.2λ, w=5λ, and ε=2)

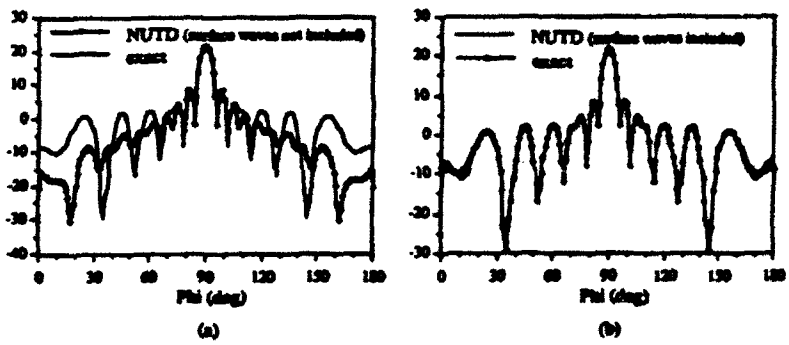


Fig. 6. H-pol back-scanning from a dielectric strip
(a) All multiple interactions EXCEPT surface waves
(b) All multiple interactions INCLUDING surface waves
(d=0.2λ, w=5λ, and ε=2)

International IEEE AP-S Symposium, pp. 1877-1880, Chicago, IL, July 1992.

ANALYSIS OF ELECTROMAGNETIC BACKSCATTERING DATA USING WAVELETS

Hyeongdong Kim* and Hao Ling
Department of Electrical and Computer Engineering
The University of Texas at Austin
Austin, TX 78712-1084

INTRODUCTION

The electromagnetic energy backscattered from an unknown target can provide information useful for classifying and identifying the target. This is commonly accomplished by interpreting the backscattered response in either the time or the frequency domain. For target characteristics which are not immediately apparent in either the time or the frequency domain, the joint time-frequency representation of the radar echo can sometimes provide more insight into the scattering mechanisms. In a recent paper by Moghaddar and Walton[1], the running window Fourier transform (RWFT) (also known as the short-time Fourier transform) was applied to the backscattered frequency response from an open-ended waveguide cavity in order to obtain a joint time-frequency representation of the original signal. In the time-frequency plane, localized events in time (scattering centers), localized events in frequency (resonances), as well as dispersive scattering mechanisms can be simultaneously displayed. Consequently, good insights were gained on the different scattering mechanisms which, when properly interpreted, can be correlated with the scatterer features for target identification applications.

However, there is one intrinsic limitation associated with the RWFT. The scattered signal in the frequency domain is generally comprised of scattering mechanisms with widely different characteristic scales. For example, high-Q resonance phenomena are highly localized (or small-scale) events in frequency. On the other hand contributions from non-dispersive scattering centers, which appear as sharp peaks in time, extend over large frequency ranges and are large-scale events. Since the RWFT uses a fixed window width, it cannot resolve multiple events with different scales simultaneously. In this paper, the wavelet analysis of backscattering data is presented. The wavelet transform has attracted much attention in the signal processing community recently [2-5]. In contrast to the RWFT, the wavelet transform technique has multi-resolution property through the use of multi-scale windows. Therefore, multi-scale events in the frequency data can be more effectively resolved in the time-frequency plane by the wavelet transform. Results for an open-ended waveguide cavity and a dispersive plasma cylinder are presented.

WAVELET ANALYSIS

We define the wavelet transform of a frequency signal $F(\omega)$ as:

$$W_f(\tau,\Omega) = \int F(\omega)\ \tau^{1/2}\ H\left(\tau(\omega-\Omega)\right)\, d\omega \qquad (1a)$$

$H(\omega)$ is usually referred to as the "mother wavelet." Note that the wavelet transform is defined in the frequency domain for our application, contrary to its

0-7803-0728-6/92 $3.00 © 1992 IEEE

usual definition in the time domain. Equation (1a) can be interpreted as the decomposition of the frequency response F(ω) into a family of shifted and dilated wavelets H(τ(ω−Ω)). By shifting H(ω) with a fixed scale parameter τ, the τ-scale mechanisms in the frequency response F(ω) can be extracted and localized. Alternatively, by dilating H(ω) at a fixed Ω, all of the multi-scale events of F(ω) at Ω can be analyzed according to the scale parameter τ. This is the so-called "multi-resolution" property of the wavelet transform and is an important advantage over the RWFT.

By properly manipulating (1a), the wavelet transform can also be carried out on the time domain signal f(t) of the original frequency domain response F(ω):

$$W_f(\tau,\Omega) = \int f(t)\ \tau^{-1/2}\, h\,(t/\tau)\ e^{j\Omega\tau}\ dt \qquad (1b)$$

h(t) is the Fourier transform of H(ω) and must satisfy the "admissibility condition," i.e., h(0)=0. Since (1b) is essentially the Fourier transform of [f(t)h(t/τ)], it is the preferred numerical implementation of the wavelet transform through the use of the FFT. With the definition in (1b), we can also interpret our motivation for using the wavelet transform from the time domain point of view. The backscattered response as a function of time, or the so-called range profile, typically consists of sharp peaks in the early time followed by small ringing in the late time. Very fine time resolution is needed to resolve the various scattering centers during the early time and good frequency resolution (or coarse time resolution) is needed for isolating the target resonances in late time. From (1b), we can see that h(t/τ) is a variable time window with its width determined by τ. In the early time (small τ) the time window width is small and provides good time resolution. In the late time (large τ) the time window width is large and provides good frequency resolution. Therefore, the variable time resolution property of the wavelet transform is ideally suited for analyzing time data which contain early-time scattering center information, intermediate-time dispersive mechanisms and late-time resonances.

WAVELET REPRESENTATION OF BACKSCATTERING DATA

The wavelet representation of the backscattering data from an open-ended waveguide cavity is considered. The cavity is an open-ended circular waveguide with a diameter of 1.75 in. A flat conducting termination exists 2 ft. inside the waveguide. To generate the backscattering data, the radar cross section of this target is first computed in the frequency domain. We take into account of the interior cavity contribution using a modal approach [6] and the diffraction contribution from the front rim of the cavity using the asymptotic formula in [7]. The time-domain response is then obtained by Fourier transforming the band-limited frequency data. Fig. 1 shows the time-frequency wavelet representation of the backscattering data at normal incidence (θ=0°). The wavelet transform is implemented using equation (1b) with the aid of the FFT. The function h(t) is chosen to be a two-sided Kaiser-Bessel window with a Q-factor of 0.3. (The Q-factor is defined as the ratio between the window width and the window center.) Both the non-dispersive rim diffraction (vertical line) and the two mode spectra due to the TE_{11} (with cutoff at 3.96 GHz) and the TE_{12} (with cutoff at 11.45 GHz) mode can be clearly identified.

Next, the backscattering from a dispersive plasma cylinder is studied. The circular plasma cylinder has a radius of 11.94 cm. A cold plasma model [8] is

assumed to model the cylinder as a dispersive dielectric with $\epsilon_r = [1-(f_p/f)^2]$. To generate the backscattering frequency response, the eigenfunction solution [9] for a circular dielectric cylinder is implemented numerically. The resulting time-frequency plot obtained using the wavelet transform is shown in Fig. 2. Two distinct scattering mechanisms can be identified in the wavelet domain: a vertical line due to the exterior specular reflection of the cylinder and a dispersion curve due to the once-internally reflected contribution from the energy penetrated into the cylinder. Below the plasma frequency, the plasma is cutoff to wave propagation and the backscattering comes solely from the strong exterior reflection. Above the plasma frequency, the plasma cylinder is partially penetrable and the once-internally reflected contribution is present. The once-internally reflected wave traverses twice the diameter of the cylinder through the dispersive plasma. Since the phase velocity of the wave in the plasma is proportional to the travel time, the curve in the wavelet plane in fact represents the dispersion diagram of the plasma. From the two examples considered in this paper, it is observed that the wavelet transform provides good resolution in simultaneously identifying the scattering centers and resolving the resonant phenomena of the target, while adequately describing the dispersive scattering mechanisms in the intermediate region.

REFERENCES

[1] A. Moghaddar and E. K. Walton, "Time-frequency-distribution analysis of dispersive targets," Workshop on High Frequency Electromagnetic Modeling of Jet Engine Cavities, Wright Laboratory, Wright-Patterson AFB, Ohio, Aug. 1-2, 1991.

[2] C.E. Heil and D.F. Walnut, "Continuous and discrete wavelet transforms," *SIAM Review*, vol. 31, pp. 628-666, Dec. 1989.

[3] S. Mallat, "Multifrequency channel decompositions of images and wavelet models," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 37, pp. 2091-2110, Dec. 1989.

[4] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961-1005, Sept. 1990.

[5] *Wavelets, Time-Frequency Methods and Phase Spaces*, J. M. Combes, A. Grossmann and Ph. Tchamitchian, eds. Berlin: Springer-Verlag, 1989, pp. 2-20.

[6] H. Ling, S. W. Lee and R. Chou, "High-frequency RCS of open cavities with rectangular and circular cross sections," *IEEE Trans. Antennas Propagat.*, vol. 37, pp. 648-654, May 1989.

[7] J. J. Bowman, S. W. Lee and C. Liang, "High-frequency backscattering from a semi-infinite hollow cylinder," *Proc. IEEE*, vol. 61, pp. 681-682, May 1973.

[8] H. Ling, H. Kim, G. A. Hallock, B. W. Birkner and A. Zaman "Effect of an arcjet plume on satellite reflector performance," *IEEE Trans. Antennas Propagat.*, vol. 39, pp. 1412-1420, Sept. 1991.

[9] H. E. Bussey and J. H. Richmond, "Scattering by a lossy dielectric circular cylindrical multi-layer -- numerical values," *IEEE Trans. Antennas Propagat.*, vol. 23, pp. 723-725, Sept. 1975.

ACKNOWLEDGEMENTS

This work is supported by the Joint Services Electronics Program and in part by National Science Foundation Grant ECS-8657524.

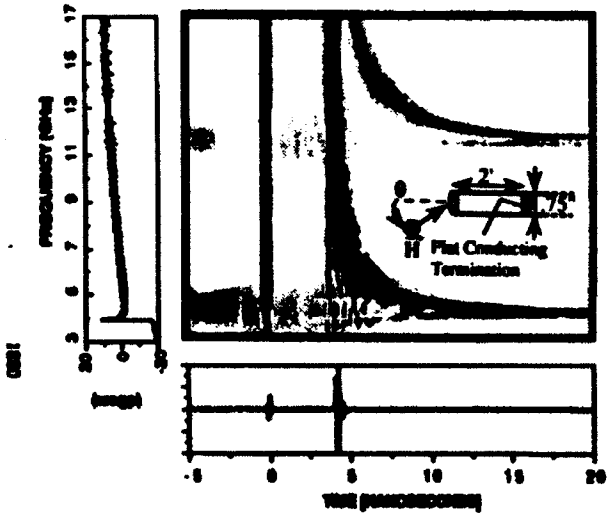


Fig. 1. Wavelet representation of backscattering data from an open-ended waveguide cavity under normal incidence. The grayscale plots of intensity are in decibels with a dynamic range of 40 dB.

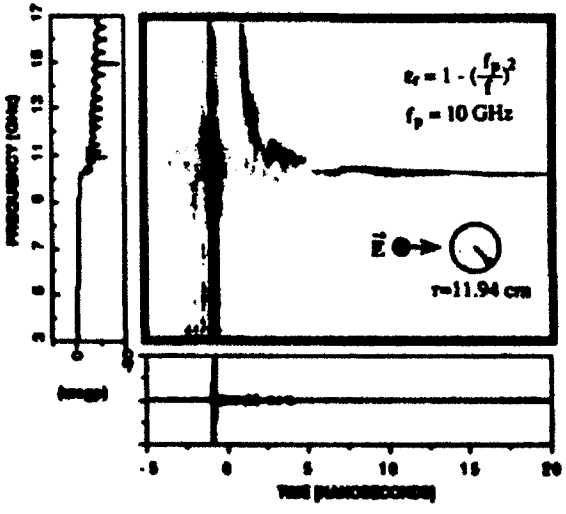


Fig. 2. Wavelet representation of backscattering data from a dispersive plasma cylinder. The grayscale plots of intensity are in decibels with a dynamic range of 48 dB.

Image Interpretation Using Multiple Sensing Modalities

Chen-Chau Chu and J. K. Aggarwal

Abstract—**This paper presents the automatic interpretation using multiple sensors (AIMS) system, which is an automatic image interpretation system using registered laser radar and thermal images. Its objective is to detect and recognize man-made objects at kilometer range in outdoor scenes. The *multisensor fusion* approach is applied to four sensing modalities (range, intensity, velocity, and thermal) to improve both image segmentation and interpretation. Low-level attributes of image segments (regions) are computed by the segmentation modules and then converted to the *KEE* format. The knowledge-based interpretation modules are constructed using *KEE* and Lisp. AIMS applies forward chaining in a bottom-up fashion to derive object-level interpretations from databases generated by the low-level processing modules. Segments are grouped into objects, and objects are classified into predefined categories using selected features. The efficiency of the interpretation process is enhanced by transferring nonsymbolic processing tasks to a concurrent service manager (program). A parallel implementation of the interpretation module on a multiple-input multiple-data (MIMD) machine is also reported. Experimental results using real data are presented.**

Index Terms— **Image interpretation, knowledge-based systems, laser radar, multisensor fusion, thermal imaging.**

I. Introduction

This paper reports on the automatic interpretation using multiple sensors (AIMS) system, which is an automatic system for scene interpretation using registered laser radar (ladar or lidar [1]) *and* thermal images. The goal of the system is to detect and recognize man-made objects (MMO) at kilometer range in outdoor rural scenes. AIMS consists of two building blocks: 1) the segmentation modules for low-level processing and 2) a rule-based interpretation system for high-level reasoning. The MMO's in our test images are mostly ground vehicles. The background is composed of vegetation, ground, and sky. An earlier version of AIMS using only laser radar data was reported in [2].

Manuscript received October 5, 1990; revised July 27, 1991. This work was supported by the DoD Joint Service Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-89-C-0044 and by the Army Research Office under contract DAAL03-91-G-0050. Recommended for acceptance by Associate Editor C. Brown.

C.-C. Chu was with the Computer and Vision Research Center, Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712. He is now with Schlumberger Austin Systems Center, Austin, TX 78726.

J. K. Aggarwal is with the Computer and Vision Research Center, Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712.

IEEE Log Number 9107012.

In this work, the *multisensor fusion* (MSF) approach is used to integrate information derived from multiple modalities. MSF helps to improve both image segmentation (by pixel-level sensor fusion) and image interpretation (by object-level sensor fusion). MSF pursues automatic information integration from different types of sensors, multiple observations through different channels, and different processing techniques. Therefore, vision systems based on MSF can provide a better performance than that of monosensor vision systems.

The knowledge-based systems (KBS) approach has been applied to various machine vision tasks such as image segmentation, object recognition, and scene interpretation [3]–[6]. Some recent research on machine vision is based on range imaging for precise 3-D modeling and understanding [7], [8]. However, both [7] and [8] use indoor range data, which are usually more precise than data obtained from outdoor range imaging because of the much shorter distances involved. Ladar range and intensity data are used in XTRS [9], which is not KBS oriented. XTRS and the earlier implementation of AIMS [2] are not able to use thermal images as input.

The ladar images used in this work have three inherently registered components: range, intensity, and velocity. The thermal images are manually registered with the ladar images (such that the pixel sizes and the viewing orientations are approximately the same). Thermal images have been extensively studied [10], [11] for image segmentation and interpretation. Thermal images usually lack high spatial resolution and convey very little 3-D information about the imaged objects. However, thermal images provide information on the material composition of objects. Therefore, thermal imaging is an effective tool to detect MMO's and complements the capabilities of ladar imaging. Both ladar range data and thermal data are used in [12].

AIMS uses all four available modalities in an *integrated* fashion at both the segmentation level and the interpretation level. Several segmentation methods are implemented for the four sensing modalities used in AIMS to extract different types of information. Three-dimensional geometry and object surface structure are extracted from ladar range data. Intensity data provide object surface reflectivity information. Velocity data indicate moving targets. Thermal images provide information about object temperature and thermal capacitance. AIMS includes an integration module [13] to integrate multiple segmentation maps with user-supplied weights. The image segmentation modules and the low-level integration module use minimal knowledge about the problem domain. Various segmentation (or edge) maps are integrated into a single map [13] before the interpretation starts. Hence, AIMS has more segmentation information for scene interpretation than systems relying only on information derived from a single source or a single feature extractor. AIMS employs

0162-8828/92$03.00 © 1992 IEEE

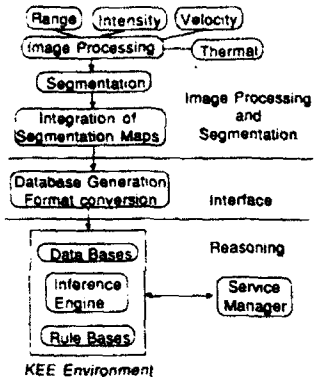


Fig. 1. AIMS System overview.

forward chaining (FC) for a data-driven, bottom-up approach. Inexact reasoning is used as the mechanism for high-level integration. The assertions generated from various rules (knowledge sources) are integrated in the form of consistent interpretation hypotheses and increased confidence factors.

Fig. 1 shows the overall structure of AIMS. The segmentation modules are written in C, whereas the reasoning modules are built using *KEE*[1] and Lisp. *KEE* is a commercial package for expert system development. It provides the inference engine and the rule parser in AIMS. The interpretation process starts by checking attributes extracted by the image segmentation modules. It then labels each segment as part of a man-made object or as the natural background (BG) based on these parameters. Next, segments are grouped into *objects* based on several criteria. Image interpretation rules then generate hypotheses of object interpretations. The hypotheses are strengthened or weakened by examining more evidence.

II. Ladar Image Segmentation

Ladar uses modulated laser light as the illumination source for outdoor long-range sensing. Ladar images are much noisier than images from video, thermal, and indoor laser images. The random refraction and reflection of laser light in the atmosphere and on the object surfaces generate *speckle noise*. This noise is significant in long-distance outdoor range imaging but virtually nonexistent in problems involving indoor range imaging, such as in [7] and [8]. Although many edge detection and segmentation techniques have already been established for video intensity images, few such methods have been established for ladar images. In this work, two region-based segmentation methods (*surface fitting* and *pixel value statistics*) are applied to ladar data. The surface fitting method is designed to highlight object surface geometry, whereas the image statistics method is used to detect differences in pixel value statistical properties. A complete discussion of these two segmentation algorithms and their performances using ladar data is reported in [2] and [14].

Most man-made objects are made of surfaces representable by patches of low-order surfaces. This assertion is practically true when the distance to an object is large compared with its body dimensions as it is in our task domain. Only planar surfaces are used in current AIMS implementation, although the module is capable of fitting surfaces up to bicubic polynomials [15]. However, high-order surface fitting is more sensitive to noise, which is abundant in our problem domain. The surface fitting-based segmentation algorithm employs a region-growing approach. Surfaces are fitted to segments, and segments grow as long as the fitting error is within a specified error bound. Surface normal, distance, and different object surface materials may generate different speckle patterns, which in turn generate image areas with different statistics in pixel values. The statistical approach is also applicable to range and velocity data. For example, the average range value from all pixels within an image segment is a good estimation of its distance to the sensor.

[1] *KEE* is a trade mark of IntelliCorp.

III. Thermal Image Segmentation

Thermal infrared (IR) images have been extensively studied [10], [11] for image segmentation and interpretation because thermal images provide information on the material composition of objects. The physics behind thermal imaging is discussed in [16]. The pixel values in thermal images are usually dominated by the temperatures of imaged objects. The temperatures are determined by various thermal properties of different materials, such as emissivity, the thermal capacitance, and the heat sink/source distinction. Some of these properties have different values for various object surface materials and, hence, can be used to distinguish MMO's from natural background. Thermal sensing can overcome visual camouflage and detect objects under suitable assumptions of thermal behavior (e.g., most objects have similar emissivities). Usually, thermal images provide a reliable differentiation between metallic objects, concrete, water bodies, and vegetation.

Quantitative analysis of thermal images is difficult, however, because the heat transfer process is complicated and the atmosphere introduces strong absorption and random refraction of heat fluxes. Consequently, thermal images lack contrast and spatial resolutions to discern details. Therefore, thermal imaging is an effective tool to detect MMO's, but it is less useful for interpretation tasks and quantitative analysis. Fortunately, contrasts between materials with distinct thermal capacities are usually strong enough to separate metal and concrete from vegetation, soil, and wood. Besides, as opposed to thermal imagery, ladar can acquire structure and shape information to complement thermal imaging.

Thermal data are used in a qualitative way in AIMS. High-temperature pixels indicate heat sources, such as engines and exhaust pipes, or materials with small heat capacitances, such as metals. Low-temperature pixels indicate heat sinks, such as vegetation, or materials with large heat capacitances, such as water, and low-temperature areas, such as a shadow. At the pixel level, thermal images provide region segmentation maps, which are then integrated with other information sources. At the object level, thermal data are used to detect metallic objects and heat sources. Using thermal images jointly with range data helps to detect covered cavities, e.g., the passenger compartment of a truck. Glass windows of vehicles are usually cooler than neighboring metallic parts. This conjecture can be verified with weak laser return intensity data and low-error planar surface fit. Hot exhaust pipes and engine blocks can confirm moving targets with segmentation derived from ladar velocity data. Most thermal imaging-based research on target recognition are signature oriented. However, AIMS is not limited to the signature-based recognition strategy since ladar imaging provides 3-D structural information and better spatial resolution.

Thermal images have different characteristics from video images. Hence, different low-level processing techniques are needed. For example, the Burns edge detection algorithm [17] does not work well in thermal images because of overblurring. The quantitative analysis technique described in [10] is not applicable in the desired work domain of AIMS. First, registered video intensity images are needed, whereas AIMS currently uses only ladar intensity, which is significantly different from ordinary video intensity imaging. Second, targets in the test images are on the order of 1 km away from the camera. Therefore, the blurring of thermal images is too much to facilitate the computation of heat flux pixel-by-pixel as in [10].

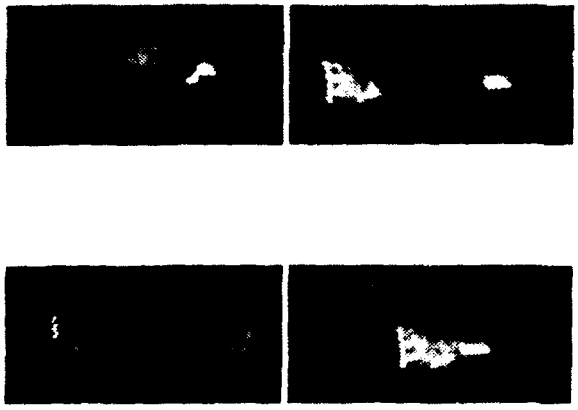


Fig. 2. Thermal images source.

Finally, surface normals are assumed perfectly known in [10], but they have to be estimated from noisy ladar range data in the work domain of AIMS. Fig. 2 shows some typical thermal images (each 128 × 256) used in this research. The top row contains a truck (left) and a truck and a personnel carrier (right). The bottom row contains two trucks and a bulletin board (left) and one truck occluding the personnel carrier (right).

A popular approach for IR segmentation is *dynamic thresholding*. It distinguishes between background and MMO's using the histogram of pixel values, assuming that pixel values are derived from a bimodal distribution function. The IR images used in this research satisfy this assumption. Usually, a large cluster exists for low-temperature pixels (vegetations), and a second cluster, which is much smaller, contains high-temperature pixels (metallic objects). In the test images, MMO's usually occupy less than 15% of the total number of image pixels and exhibit higher temperatures than those of the background vegetation. All the images have open field or trees/woods as a background. The field is covered with grass. When the central portion of IR images is resampled with the same spatial resolution as the ladar images, usually 90% of the pixels belong to the background in a resampled 128 × 256 window. For an object 5m long and 2.5m high, the broadside view takes about 12% of the window.

One can assume that all the different thermal characteristics of background vegetation can be approximated by a Gaussian distribution of pixel values. The Gaussian bell of background vegetation is located at the low end of the histogram because hardly anything is cooler than the background vegetation (except shadows and the sky). The peak of this Gaussian bell is also the peak of the entire histogram because the background dominates the entire image (Fig. 3). A segmentation scheme is designed based on the above-mentioned observations and assumptions. The standard deviation σ of the Gaussian distribution consisting of the low-temperature pixels is determined by solving

$$0.5 = \exp\left(\frac{-(x_0 - \mu)^2}{2\sigma^2}\right) \tag{1}$$

where x_0 is the 3-db width of the Gaussian bell. In a Gaussian distribution, the *mean* μ is the same as the *mode* of the distribution and, hence, is easily determined as the peak of the histogram. x_0 is estimated by tracing the histogram to find the 50% drop-off point from the peak of the histogram. Note that μ is not determined as the average of the entire thermal image. All pixels with gray values covered by the range of $[0, \mu + \sigma]$ are classified as background; all pixels with gray values in the range of $[\mu + 3\sigma, 255]$ are considered to be MMO. Therefore, the thresholds are determined dynamically for individual images. Pixels with gray values in between are then determined by their proximity to classified pixels. However, only regions that are large enough are established as segments to reduce noise effects.

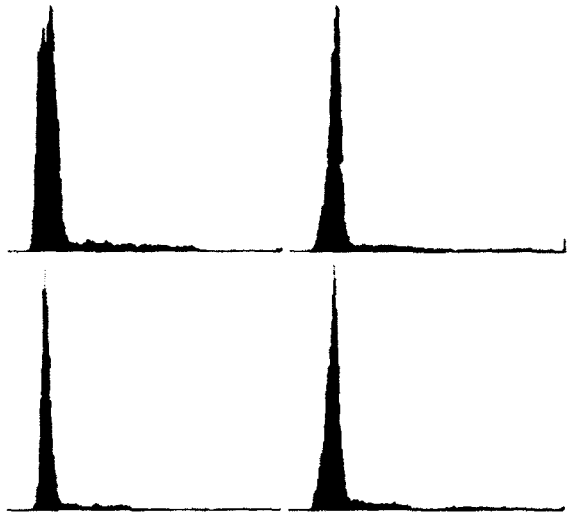


Fig. 3. Histograms for thermal images in Fig. 2.

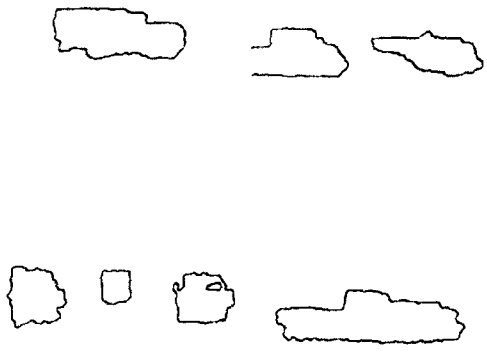


Fig. 4. Thermal image segmentation results from images in Fig. 2.

Fig. 3 shows the histograms of four thermal images shown in Fig. 2. All of them satisfy the assumption of small target surface areas in a large background full of vegetation and soil. The histograms clearly indicate the Gaussian bell distribution at the lower end (left) of the plot. Fig. 4 shows the segmentation results. The execution time of the dynamic thresholding algorithm is about 5 s on an IBM RT and is the fastest method among all segmentation methods employed in this work. Note that the segmentation derived from thermal data usually cannot detect the occlusion contours between two objects of similar thermal properties. In addition, the same segmentation algorithm does not perform as well when it is applied to urban scenes with many concrete structures (high thermal capacitance) in the scenes. In such cases, the histograms no longer satisfy the assumptions.

IV. The Integration of Segmentation Maps

High-quality image segmentation is a crucial intermediate stage before image interpretation. The strong noise makes the interpretation of ladar images even more difficult. Although various techniques have been proposed for different sensors, each one of them has respective merits and weakness. Different methods operating on multiple data sources generate different segmentation maps. These maps may have errors and possibly contradict one another. AIMS includes an integration module that is very different from and

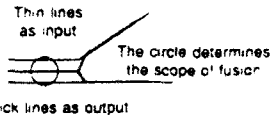


Fig. 5. Low-level integration of segmentation cues.

more flexible than previous works developed for similar purposes [18]–[20]. For example, it is helpful to apply different weights on various input segmentation maps because quality and reliability of different segmentation methods and data sources may be different.

The low-level module (i.e., using no high-level knowledge) in AIMS [13] integrates both region and edge segmentation maps, which are produced by various methods operating on different sensor outputs. The *maximum likelihood* criterion is used to generate an initial solution of edge pixel positions from samples collected in local windows. With the assumption of white Gaussian noise, the solution s_x is in the form of a weighted average from m observed samples x_i:

$$s_x = \frac{\sum_{i=1}^{m} (1/\sigma_i^2) x_i}{\sum_{i=1}^{m} (1/\sigma_i^2)} \tag{2}$$

and similarly for s_y. User-supplied weights may be considered to be the effects of the precision of various observation processes and can be combined with σ_i's. The size of the local window thus decides the spatial scope of the low-level fusion, and it plays a role similar to the *channel resolution width*. Solutions produced by neighboring windows are linked to preserve contour connectivity (Fig. 5). The resultant edge map is represented by a node-spring model. The edge map is then modified to minimize the stored potential energy within the node-spring model. Finally, the edge map is converted to a region map, and small regions are removed.

The current implementation of this integration module is independent of the image sources and segmentation techniques. Both region and edge maps are allowed. In addition, different weights can be applied to various information sources for differences in data quality and reliability. In general, range data are not as noisy as their intensity counterparts and, therefore, are given higher weights. Velocity data provide useful segmentation information only if moving targets are in the scene. Therefore, the weight on velocity segmentation depends on the segmentation outcome of individual images. Edge maps derived from ladar data are usually of a lower quality and, thus, receive less weights. However, edge detection usually provides better cues than region segmentation for ladar intensity data and jump edges in ladar range data. Thermal segmentation maps usually produce precise MMO/BG boundaries; hence, a heavy weight is given. However, the integration algorithm is robust enough even if strong weights are mistakenly given to the *wrong* information sources [13]. If video intensity images (e.g., from TV camera) are available, they can also contribute to the integration process.

A set of utility programs collects the values for various attributes, such as region size and standard deviation, using the original images and the integrated segmentation map. These data are converted to the representation format of *KEE* by the *database generator*, and the database is then transferred to the interpretation modules [2]. The database generator serves as the interface from the low-level segmentation block to the high-level interpretation block.

V. The Interpretation System

The interpretation modules in AIMS consist of four major components: 1) the *inference mechanism* provided by *KEE*; 2) the *rule bases* and supplemental Lisp code, which contain the knowledge for image interpretation; 3) the *data bases*, which are produced by the database generator; and 4) the *service manager*, which executes numerical and graphics tasks for AIMS.

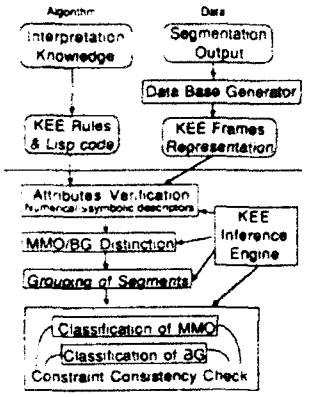


Fig. 6. Block diagram for the interpretation system.

The interpretation strategy of our work follows the three-step paradigm [21] of data abstraction, heuristic classification, and refined classification. First, numerical parameters are converted into qualitative descriptors. Second, these descriptors are used to generate intermediate classifications of segments as MMO's or background. Third, segments are grouped into objects, and these objects are further classified into one of the predefined categories. Fig. 6 shows the block diagram for the interpretation module and its operation. A more detailed description of the complete interpretation system is presented in [22].

The rules in AIMS are organized into six groups: 1) data preparation, 2) the coarse classification of segments into MMO/BG using low-level attributes, 3) segment grouping using geometric information, 4) the classification of BG segments/objects, 5) the classification of MMO segments objects, and 6) consistency verification of neighboring segment objects. These groups of rules are sequentially invoked in the FC mode. At any given time, only one group of rules is active in the *match-resolve-fire* cycle. However, stages 4) and 5) can operate in parallel. The conflict resolution strategies in AIMS are *rule weighting* and *FIFO*. The partition of rule bases reduces the matching overhead of rule selection and provides indirect control over the breadth-first search implied in FC. BC rules will be added in the future to adopt the *hypothesize-and-verify* approach for focused searches. Thus, when a hypothesis with a strong confidence is posted, the reasoning process can switch into the BC mode to verify that hypothesis.

Despite the flexibility of *KEE*, three major issues are identified as its weak spots: 1) execution efficiency, 2) low-level data access during high-level reasoning, and 3) interface capability and feedback to low-level processes. Lisp-based development systems, such as *KEE*, are convenient tools to execute symbolic reasoning tasks and handle explicitly encoded knowledge. However, these systems usually do so at the price of software overhead. The slowdown occurs for two main reasons. First, most such packages are built on multiple layers of software and, therefore, are very inefficient. Second, image interpretation is not a task that consists solely of symbolic processing. Our solution is a program written in C (the service manager), which runs concurrently with the interpretation modules. The purpose of the service manager is to help AIMS run low-level tasks (I/O, video display, etc.) efficiently on the designated development platform. The interpretation module sends a message to the service manager for the desired service. The service manager interprets and executes the commands and feeds the results back.

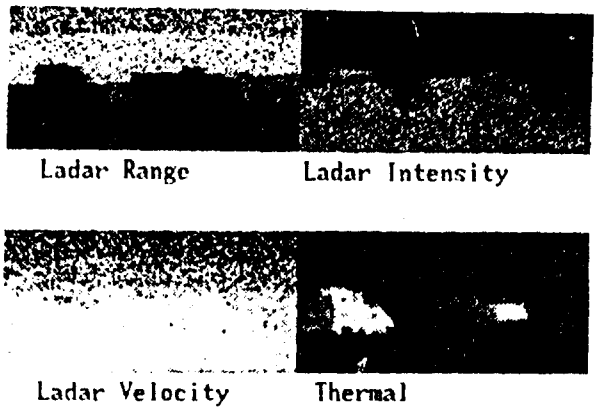


Fig. 7. Source images for one truck and one personnel carrier.

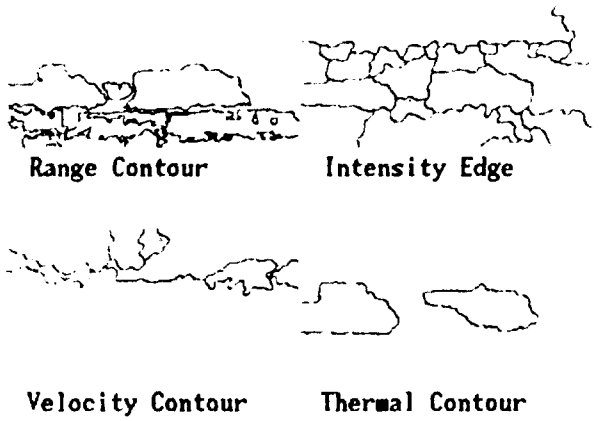


Fig. 8. Segmentation maps for Example 1.

VI. Experimental Results

The system is implemented on an IBM RT PC running AIX. Figs. 7 and 10 contain the original ladar range, intensity, velocity, and registered thermal image. Figs. 8 and 11 are various segmentation maps fed into the low-level integration module. Figs. 9 and 12 are the integrated segmentation maps (top) and the interpretation results (bottom). The integrated segmentation maps show region boundaries in white contours overlaid on the range image. For the interpretation results, white regions are detected targets, and black areas are segments that do not have a high-confidence interpretation hypothesis. Some of the black areas are actually classified as GROUND or SKY. However, the confidence factors for such classifications fall below a threshold (0.4) and are considered too weak to report. Light gray marks GROUND, and dark gray marks SKY.

The first example (Figs. 7–9) consists of a 2.5-ton truck and an armored personnel carrier (APC) at 922 m. The final segmentation delineates the entire truck as a single segment, whereas the APC is split into two segments and then grouped as a single object. The interpretation process classifies both segments as MMO's and recognizes them as a TRUCK and an APC, respectively. Because the truck is occluded by the image frame boundary, the truck driver's cabin could be mistaken for a tank turret. However, the analysis of ladar intensity images at the possible turret position reveals that the left target is a truck because of the very weak return from that portion. The weak return compared with its surroundings indicates heterogeneity in the material composition of the upper central portion

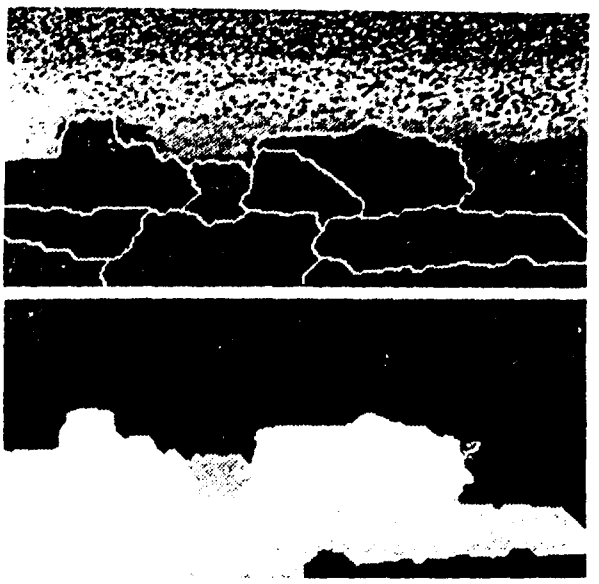


Fig. 9. Integrated segmentation and interpretation for Example 1.

of the target. An open space is identified at the tail of the APC, which supports the interpretation as APC by comparing the thermal and ladar range images. Note that the thermal segmentation module does not include the cavity area as part of APC because of the lower temperature of that area. Since the cavity is large enough, the APC is split into two segments in the integrated segmentation map. If thermal information is unavailable, then the recognition process must rely solely on contour and structural analysis. The body dimensions of the APC are correctly estimated using the spatial resolution of the radar system and the estimated target distance. The rotations are estimated as 28.1° for the APC and 26.1° for the truck. This is consistent with the documentation, which states that both targets are parallel, with a rotation of about 30° from the viewing direction. Most background segments are classified as GROUND. Some of them are also classified as SKY with weaker CF's because of weak ladar intensity return and low temperature.

The second example (Figs. 10–12) consists of two trucks and one bulletin board between them at 939 m. The truck on the right is not moving, whereas the truck on the left is moving toward the ladar receiver. There is a strong indication of a moving target for the left truck in the velocity image. The speed is estimated correctly by the average velocity value. The segmentation module successfully marks out the three objects as MMO's corresponding to the two trucks and the central bulletin board. Both trucks receive the correct classification as TRUCKs. However, the corresponding CF's are weak. The reason is that the current implementation of the contour analysis rules has only a limited knowledge base to recognize vehicles from a broadside view. Therefore, only interpretation rules that are not strongly dependent on contours provide the correct classification. The turret-analysis rule used in the previous example suggests that the upper structure of the two targets may be trucks. Thermal rules detect the exhaust pipe of the left truck and determine that the central target is unlikely to be a heat source. The width of the bounding rectangle of the target is matched against the width, and not the length, of a potential target. The surface fitting error for range data indicates that the central MMO segment is almost flat. Its shape is rectangular, and its size (2.07 m high and 1.56 m wide) is smaller than that of most vehicles, yet it is taller than a JEEP. The size, smaller surface fitting error, moderately high temperature, and the contour can be used as evidence to hypothesize and support the claim that the central segment

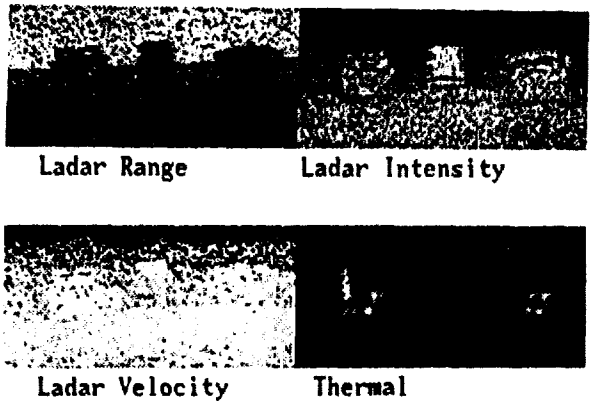


Fig. 10. Source images for two trucks and a bulletin board.

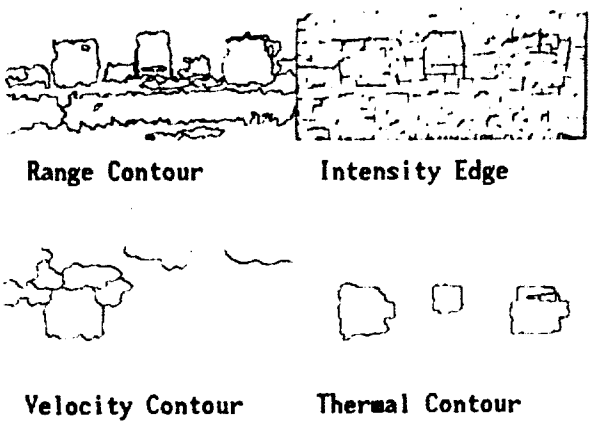


Fig. 11. Four segmentation maps for Example 2.



Fig. 12. Integrated segmentation and interpretation for Example 2.

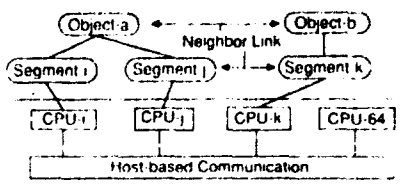


Fig. 13. Processor mapping used in the parallel implementation.

is a BULLETIN BOARD. The other two MMO's have surface fitting errors of about $\pm 3.5m$, which is approximately half the length of a truck. The widths of both are estimated to be 2.4 and 2.35 m (actually one pixel in difference), which is approximately the width of a 2.5-ton truck. The height of the vehicles argues against a JEEP. The dz/dy gradient is used to estimate that the rotations (of the long axis) of both trucks are about 10° from the viewing direction. This measurement is consistent with the documentation, which states that the trucks are parallel, and one of them was backing up toward the ladar.

The data collection and database generation modules between the segmentation block and the interpretation block take about 2 min of CPU time. The interpretation block takes about 15 min of wall clock time (368 s of CPU time due to extensive memory swapping) to interpret one set of images. The CPU time would have been longer (roughly 30 min WCT and 19 min CPU) if the C-based service manager is not employed, and hence, all the low-level processing tasks must be coded as *KEE* and Lisp functions.

VII. Parallel Implementation of the Interpretation System

After the interpretation subsystem stabilizes, it is ported to an AT&T for further improvement in run time statistics. Each of the 64 nodes in the available PIXEL machine has a DSP32 chip as the CPU, 36 KB memory for code and 256 KB for data. Consequently, only a subset of the rule base is ported to the PIXEL machine due to the limited memory space. Instead of using OPS-5 and the Rete network [23], a strategy of one processor per image segment is chosen (Fig. 13). The issue here is that the PIXEL machine adopts a message passing-based architecture. Fine-grained parallelism as emphasized in [23] would be difficult to explore. Therefore, it is preferable to duplicate the software and the shared data to reduce the amount of message communications.

Data extracted from Fig. 7 are translated into the DSP32 format and downloaded with the integrated segmentation map. The rules are implemented together with the inference engine (IE) as one program in C to save the time needed to construct a general-purpose rule parser. There are ten rules implemented as three sections. The three sections are asserted sequentially in the FC mode. Conflict resolution is determined by rule weighting and by FIFO when rules have equal weights. Rules are of various degrees of complexity. In general, they are more numerical intensive than those used for symbolic matching in OPS-5. For example, one rule scans the entire segmentation map to determine the bounding rectangle and the symmetry of the contour shape of regions as part of its left-hand-side (LHS) operations. Another rule generates all 108 messages in its LHS (reporting the interpretation at the current node to its neighbors) and processes all of them (examining the interpretation at neighboring segments for consistency) in its right hand side (RHS). The workload of processing these messages is distributed over many nodes, of course. The 24 data segments (based on Fig. 7) are distributed among 24 nodes. The remaining 40 nodes are not used in the experiment. A data frame describing one region in the test image is about 300 bytes. An average rule takes about 1.4 KB SRAM space, whereas data, the IE, and other housekeeping codes take about 21 KB SRAM space.

The total run time averages (in 20 runs) 7.2 s wall clock time (WCT), including 6.91 s host-CPU time (SUN-3 280) using the UNIX *time* command. With a PIXEL machine program downloaded but without actually running it, the timing statistics are 6.815 s CPU host and 7.0 WCT (averaged from 20 runs). Therefore, the actual run time with code and data ready in individual nodes is 0.1–0.2 s

More precise measurements are not done because no system tool is available. Most of the host CPU time is consumed by data download and system initialization. Total rule firing is 137, and at least 240 WME changes because some rules change several elements. The passage and processing of 108 messages generate no measurable increase in both the host CPU time and the WCT (by disabling the single rule that generates all the messages). These statistics result in an equivalent rule-firing rate of 700–1400 rules/s. This rule-firing speed is close to the 903 rules/s *projected* for NON-VON [24]. In comparison, the complete rule base of 47 rules accelerated by a C-based server running on an IBM RT takes 15 min WCT and 368 s CPU time. The IBM RT is roughly 2.5 times as fast as a VAX-11/780, whereas a VAX-11/780 has a rule-firing rate of 2–3 rules/s.

Gupta makes a good point [23] that the highly inefficient implementation of OPS-5 by Lisp and RBS built on OPS-5 severely limits the use of such systems and the research progress. Similar situations exist in vision research when the recognition/interpretation tasks are implemented using the RBS approach. It is reported [23] that C-based implementation alone (with some optimization effort) provides a speedup of 10–20 over a Lisp-based implementation in a uniprocessor environment. If the same factor is applicable to the simulation results in this experiment, then the speedup due to the reported parallel implementation is $(368/0.2) \times (10/47)/20 \approx 19.59$ without considering the speed difference between the CPU's. The language factor as 20 may be overestimated because the 368 s run time is obtained from the mixed-language implementation, which uses C to a great extent [22]. The ratio of raw speed between a DSP32 chip and an RT CPU is roughly 5 to 8 in the sense of nominal MIPS. If the CPU factor is also considered, then the parallel implementation (roughly) achieves linear (24) speedup. If the image contains more segments (up to 64, which is the number of physical PE nodes in the PIXEL machine), the run time on the PIXEL machine should increase very little. In comparison, a uniprocessor implementation will use proportionately more time (w.r.t. the number of segments). However, the simulation time is too short for a precise run-time measurement.

VIII. Conclusion

We presented a knowledge-based multisensor system (AIMS) to interpret registered laser radar (ladar) and thermal images. The objective is to detect and recognize man-made objects at kilometer range in outdoor scenes. AIMS consists of rule-based reasoning modules and the segmentation modules. The *multisensor fusion* approach is applied at both the segmentation and reasoning levels. The low-level integration module fuses segmentation cues from multiple image sources and processing techniques to generate an improved segmentation map. The reasoning system integrates high-level information from various modalities and knowledge sources in the form of consistent interpretation hypotheses and increased confidence factors. AIMS uses forward chaining to drive the interpretation process in a bottom-up fashion. The reasoning process follows the order of data abstraction (feature extraction), heuristic classification (target detection), and refinement verification (target recognition). The tasks at different levels of the machine vision paradigm are performed in AIMS using different software tools and methodologies. A subset of the interpretation subsystem of AIMS is implemented on a multicomputer MIMD environment and achieves almost linear speedup.

Two factors, including 1) multisensor fusion and 2) the hypotheses integration using inexact reasoning, enable AIMS to tolerate intermediate errors by verifying image segmentation and initial interpretation incrementally. Thermal imaging is very effective in detecting metallic objects and vegetation, and it complements the shape/structure information provided by ladar imaging. A high-quality segmentation map is obtained using the map integration module in AIMS. Consequently, low-level data can be correctly aggregated into object-level attributes, which is essential for region-based image interpretation. The rule-based interpretation system cooperates with low-level modules for efficiency and incremental verification. A parallel implementation of a subset of the complete interpretation rule bases provides similar results much faster than the uniprocessor mixed-language implementation. The speed-up factor is about 130 in wall clock time and 1800 in CPU time. AIMS performs well on real images and recognizes unoccluded man-made objects at kilometer range. Further testing is still under way. The current implementation of AIMS is only a prototype, and it must acquire additional knowledge to recognize more objects. When the problem domain changes, different sets of object models and recognition rules have to be built, and different sets of features may be needed for recognition.

Acknowledgment

The authors would like to thank the Army Electro-Optical and Night Vision Center for providing the test images.

References

[1] C. G. Bachman, *Laser Radar Systems and Techniques*. Dedham, MA: Artech House, 1979.

[2] C. Chu and J. K. Aggarwal, "Interpretation of laser radar images by a knowledge-based system," *J. Machine Vision Applications*, no. 4, pp. 145–163, 1991.

[3] D. M. McKeown, Jr., W. A. Harvey, and J. McDermott, "Rule-based interpretation of aerial imagery," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-7, no. 5, pp. 570–585, Sept. 1985.

[4] J. A. Mulder, A. K. Mackworth, and W. S. Havens, "Knowledge structuring and constraint satisfaction: the Mapsee approach," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, no. 9, pp. 866–879, Nov. 1988.

[5] Y. Ohta, *Knowledge-Based Interpretation of Outdoor Natural Color Scenes*. Marshfield, MA: Pitman, 1985.

[6] B. A. Draper *et al.*, "Tools and experiments in the knowledge-directed interpretation of road scenes," in *Proc. DARPA Image Understanding Workshop*, 1987, pp. 179–193.

[7] A. K. Jain and R. Hoffman, "Evidence-based recognition of 3-D objects," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, no. 6, pp. 783–802, Nov. 1988.

[8] T. J. Fan, G. Medioni, and R. Nevatia, "3D object recognition using surface descriptions," in *Proc. DARPA Image Understanding Workshop*, 1988, pp. 383–397.

[9] D. E. Dudgeon, J. G. Verly, and R. T. Delanoy, "An experimental target recognition system for laser radar imagery," in *Proc. DARPA Image Understanding Workshop*, 1989, pp. 479–506.

[10] N. Nandhakumar and J. K. Aggarwal, "Integrated analysis of thermal and visual images for scene interpretation," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 10, no. 4, pp. 469–481, July 1988.

[11] B. Bhanu and R. D. Holben, "Model-based segmentation of FLIR images," *IEEE Trans. Aerospace Electron. Syst.*, vol. 26, no. 1, pp. 2–11, Jan. 1990.

[12] C. W. Tong, S. K. Rogers, J. P. Mills, and M. K. Kabrisky, "Multisensor data fusion of laser radar and forward looking infrared (FLIR) for target segmentation and enhancement," in *Proc. SPIE*, May 1987, vol. 782, pp. 10–19.

[13] C. Chu and J. K. Aggarwal, "The integration of region and edge-based segmentation," in *Proc. Third Int. Conf. Comput. Vision* (Osaka, Japan), Dec. 4–7, 1990, pp. 117–120.

[14] C. Chu, N. Nandhakumar, and J. K. Aggarwal, "Image segmentation using laser radar data," *Patt. Recogn.*, vol. 23, no. 6, pp. 569–581, 1990.

[15] C. Chu and A. C. Bovik, "Visual surface reconstruction using minimax approximation," *Patt. Recogn.*, vol. 21, no. 4, pp. 303–312, 1988.

[16] F. P. Incropera and D. P. De Witt, *Fundamentals of Heat Transfer*. New York: Wiley, 1981.

[17] J. B. Burns, A. R. Hanson, and E. M. Riseman, "Extracting straight lines," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-8, no. 4, pp. 425–455, July 1986.

Pattern Recognition, Vol. 25, No. 8, pp. 835–853, 1992
Printed in Great Britain

0031-3203/92 $5.00 + .00
Pergamon Press Ltd
© 1992 Pattern Recognition Society

APPLYING PERCEPTUAL ORGANIZATION TO THE DETECTION OF MAN-MADE OBJECTS IN NON-URBAN SCENES

H. Q. Lu† and J. K. Aggarwal‡
Computer and Vision Research Center, Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712, U.S.A.

(*Received* 19 *April* 1991; *in revised form* 16 *December* 1991; *received for publication* 9 *January* 1992)

Abstract—A new approach for the detection of large man-made objects in a non-urban area using a single monochrome image is presented. In this study, the man-made objects may be unspecified and the appearance of the objects is unpredictable. Prominent features that distinguish man-made objects from natural objects are identified. A computational framework of applying perceptual organization and using the prominent features is presented. Techniques are developed to group low level image features hierarchically into a *region-of-interest* (ROI) likely to enclose man-made objects or a substantial part of the man-made objects. These techniques include feature extraction, primitive structure formation, and segmentation. Some of these methods are novel and others present unique properties and advantages compared to previous related works. Experimental results are presented using real images that include several different man-made objects in complex backgrounds or a natural scene without man-made objects. It is shown that the located ROIs properly enclose the man-made objects in the scenes.

Object recognition Perceptual organization Geometric structures Feature extraction Feature representation

1. INTRODUCTION

The computer perception of man-made objects in non-urban scenes is a challenging task in computer vision research. It also presents additional complexities and difficulties beyond the computer perception of objects in a well-controlled laboratory or factory environment. This paper presents a new approach for the automatic detection of large man-made objects in outdoor non-urban scenes. The methodology presented here is based on *perceptual organization*. The approach hierarchically organizes low level image features to higher level structures to find a region of interest in a given image.

The environment we consider for image acquisition is a non-urban area in daylight hours. Large man-made objects, such as bridges and electric transmission towers, may be present in the area among natural objects, such as trees, bushes, and vegetation. The man-made objects are unspecified and their appearance is unpredictable. Thus, we do not know what man-made objects may appear or whether there is a man-made object in the scene. Given a single monochrome image of such a scene, the goal is to automatically detect large man-made objects in the image. Because of the complexity and variation of man-made objects and the uncertainty of the natural environment, the intermediate goal is to find in the image a *region-of-interest* (ROI) most likely to enclose man-made objects or a substantial part of the man-made objects.

Most computer vision research on object recognition has focused on problems with specified objects in controlled environments. For example, many of the current vision systems try to recognize objects in a given image with a uniform background and one or more objects whose exact models are known to the system.[1] Even with such seemingly well-defined problems, many obstacles exist and considerable research efforts are being made. Locating man-made objects in a natural outdoor environment adds yet another dimension of complexity. One cannot arrange the natural environment. Natural objects, such as trees, vegetation, rivers, rocks, and clouds, co-exist in the scene with the possible man-made objects. As Fischler and Strat[2] observed, it is seldom possible to establish complete boundaries between objects of interest in natural scenes, and very few natural objects have compact shape descriptions. Therefore, the problem of detecting man-made objects in a natural environment is, in general, much more difficult than the object recognition tasks in a well-controlled environment.

Many researchers have investigated the automatic interpretation of outdoor natural scenes. Some of them have investigated the detection of man-made objects[3–7] other than buildings and roads. These studies, in general, use additional information about the scene, such as color or range, or have better knowledge (models) about the objects. Most of the

† Present address: Schlumberger Austin Systems Center, Austin, Texas, U.S.A.
‡ Author to whom all correspondence should be addressed.

other work focuses on the interpretation of aerial images or outdoor robot navigation. In the first area, techniques have been developed for detecting complex buildings and roads in which buildings are modeled as combinations of rectangles with uniform intensity and roads as parallel curves.[8-12] For outdoor robot navigation, most vision-related work concerns road following[13,14] and position estimation.[15,16] Usually, a sequence of images is used for the navigation problems. Thus, when interpreting each image, a good initial estimation can be obtained from the interpretation of the previous images.[13,15]

Unlike the previous studies cited above, this investigation has the following features. (1) Man-made objects which may appear in the scene are unspecified. (2) These man-made objects may have more complicated structures than the buildings which appear in aerial images and cannot be easily modeled using rectangles. (3) It is hard to predict the presence of man-made objects in the scene. (4) One monochrome image is used, and no color or range information is provided. In addition, the change in appearance of the man-made objects caused by changing view-direction is a more severe problem than in top-view aerial images.

Generally speaking, humans can detect man-made objects easily. Psychologists have found that *perceptual organization* or *perceptual grouping* plays an important role in human perception. Perceptual organization refers to the human's visual ability to derive relevant groupings or structures from input images without prior knowledge of their contents.[17] For example, people can easily detect symmetry, collinearity, and parallelism. Thus, if we can derive similar groupings or structures computationally from an input image, that will be very useful to the detection of man-made objects, especially since we do not have any prior knowledge of the image's contents.

Based on the principles of perceptual organization, this paper presents a new approach for the detection of large man-made objects in a non-urban area. The system finds in the image an ROI likely to enclose man-made objects or a substantial part of the man-made objects. Minimal knowledge and information about the objects and scenes are used in this work. Since it is desirable to have a general approach able to handle a variety of man-made objects, we try to find prominent features distinguishing man-made objects from natural objects. We then present a computational framework for applying perceptual organization and using the discriminative features to find an ROI in an image. Other researchers have also applied perceptual organization to various computer vision tasks.[3,8,17-23] Their work has important impact on our research. However, these studies concentrated on grouping features and recognizing objects using simple generic models or exact models for specified classes of objects. Such approaches do not apply to our problem in which the objects of interest are unspecified and, in general, have complicated structures.

The new framework presented in this paper is implemented in three modules which correspond to the three levels of the grouping process. The first module, *feature extraction*, detects linear edges from the input image and then extracts image features from the edges. The image features include *linear structures* (LS) and coterminous lines (*coterminations* (CT)). Unlike other collinearization methods,[9,21,24] our method of extracting linear structures performs line *folding* and line extension simultaneously. Coterminations are represented by a graph called a *CT graph*. The second module, *primitive structure formation*, extracts primitive structures (PS) in the image by grouping image features. The primitive structures include parallel PS and polygon PS. The definition and extraction of the parallel PS consider more general situations than most of the previous relevant work using parallel lines.[3,8,23] An efficient and systematic procedure is developed to find polygon PS using the CT graph. This algorithm is executed in polynomial-time and, hence, is much more efficient than the direct search, which would take exponential-time. The primitive structures are then represented by a graph called a *PS graph* that describes the spatial relationships among the PS. Such a graph facilitates efficient higher level processing. The third module, *segmentation*, segments the given image to obtain an ROI most likely to enclose man-made objects or a substantial part of the man-made objects. Using the PS graph, we develop a procedure to group related PS, to eliminate isolated PS, and to segment the image into regions occupied by the grouped PS and a background. This graphic approach is efficient and allows a parallel implementation. The largest region of the grouped PS is then evaluated based on the area of the region and the statistics of the PS. If this region is determined to be significant, then it is deemed to be the ROI of the given image.

The approach presented in this paper is different from the work presented by Reynolds and Beveridge.[23] In reference (23), geometric relations among lines are extracted and represented by a graph. Our method is a hierarchical grouping process. We define primitive structures and identify their relationships. The advantages are the abilities to extract a variety of geometric structures, to establish higher level relationships among image features, and to use regional information.

We present several examples, including different kinds of objects in non-urban scenes and a natural scene without any man-made objects. We show that the ROI enclose the man-made objects in the scenes. These examples illustrate the ability of this approach to locate useful regions of interest in complex real images. The identification of ROI is important and may be necessary for the initial screening of a large environment to detect man-made objects, since the search space can be reduced to the located ROI. When specific object classes are given, the primitive structures composing the ROI may be used to match structures of the object models instead of matching individual features. The ROI may also be used in human–machine systems to find the focus-of-attention so that human

expertise can make a final determination. This is applicable to real-time operations, such as assisting an aircraft pilot by looking in alternate directions and providing ROI, and to off-line processing involving a large number of images.

The rest of the paper is organized as follows. Section 2 briefly reviews previous research relevant to this work. Section 3 overviews our approach. Section 4 describes feature extraction. Section 5 defines primitive structures and presents techniques to derive PS. Section 6 concentrates on the region of interest. Section 7 gives implementation examples, and, finally, Section 8 concludes the paper.

2. RELATED WORK

This section briefly reviews previous work relevant to our research. The review mainly includes the work of applying perceptual organization to computer vision tasks and the detection of man-made objects in outdoor natural scenes.

Perceptual organization has been studied since the early part of this century. Gestalt psychologists studied a large number of grouping phenomena of human vision and roughly categorized them into several *Gestalt Laws* (or *grouping rules*) which are summarized as follows:[17,25] (1) *proximity*: closer elements tend to be grouped together; (2) *similarity*: similar elements tend to be grouped together; (3) *continuation*: elements that lie along a common line or smooth curve are grouped together; (4) *closure*: curves tend to be completed to enclose a region; (5) *symmetry*: elements symmetric about some axis are grouped together. Gestalt psychologists believed these factors to be important in the perception of structures.

In recent years, perceptual organization has been introduced into computational vision research and its functional role has been addressed.[17,26] Lowe[17] argues that the most important functions of perceptual organization include segmentation, three-space inference, and the indexing of world knowledge. All of these lead to the reduction of search space for object recognition. McCafferty[18] formulates perceptual organization as an energy minimization problem. He quantifies the Gestalt Laws by defining individual energy terms. However, the selection of the weightings for these energy terms presents problems.

Recently, perceptual organization has been applied to solve practical computer vision problems.[3,8,19,20,23] Mohan and Nevatia[20] apply perceptual organization to segment images into visible object surfaces. They also investigate the detection and description of complex buildings in aerial images.[8] Assuming that roofs are the essential building structure seen in the image, they model the roof as a combination of rectangles. Baker *et al.*[3,4] present an approach for the detection of concrete bridges. The straight line segments, once detected, are grouped into parallel lines. Intrinsic rectangles are extracted from the parallel lines. Color cues are then used to restrict the candidate artifacts and to produce confidence measures. However, the rectangle-type model may be unsuitable to man-made objects with more complex structures in outdoor scenes. In this investigation, we deal with objects with complex structures as well as those with simpler structures.

Reynolds and Beveridge[23] examine the problem of searching for geometric structures in natural scene images. Straight lines are grouped using the geometric relations of collinearity, parallelism, orthogonality, and spatial proximity. The connected components of a graph representing lines and their relations are illustrated to correspond to significant geometric structures in the image. However, some components may contain many different image events. In addition, this method may not find some simple geometric structures, such as parallelograms other than rectangles. Unlike reference (23), our method groups low level features into primitive structures. The spatial relationships among the primitive structures are used to find the ROI most likely to enclose man-made objects or a substantial part of the man-made objects. The advantages of finding primitive structures and identifying their relationships are the abilities to extract a variety of geometric structures, to establish higher level relationships among image features, and to use regional information.

Jacobs[19] presents a system called GROPER, which recognizes two-dimensional (2D) objects using a library of many different objects. GROPER applies perceptual organization to reduce the search space for matching scene objects with models. GROPER is designed for a simplified world containing only 2D polygonal objects, whereas the objects encountered in our research are 3D objects with complex structures.

Various techniques are presented in the literature for applying perceptual organization to group lower level image features, such as edge points, into higher level structures, such as straight lines and curves, and to detect junction, collinearity, parallelism, and symmetry.[9,21,22,24,27] Although similar to some of these techniques, the methods developed in our research for the first two levels of grouping have unique properties and advantages. We discuss the differences and advantages after describing each of the methods.

There are other works that concern the interpretation of natural scenes in addition to those using perceptual organization. Brooks[5] presents the identification of aircraft in aerial images using ACRONYM, a model-based vision system. In ACRONYM, generic object classes and specific objects are represented by volumetric models using generalized cones along with sets of constraints relating to model parameters. Fua and Hanson[10–12,28] present a sequence of papers on extracting features and locating *general cultural objects* such as buildings in aerial images. A new approach, based on information theory, to evaluate the correspondence between generic models and shape hypotheses in an image is presented in references (10, 28). Generic models for buildings are formulated and experimental results on several aerial images are presented. Beveridge

et al.[29] present a method for identifying known 2D models in imperfect line data. The method is applied to complex outdoor scenes and good matches are demonstrated. However, the above approaches are unsuitable to detection problems where the potential objects are unspecified, as in the case of our research. Chu *et al.*[6,7] present a system called AIMS to detect and recognize man-made objects in outdoor scenes. Multiple sensing modalities (range, intensity, velocity, and thermal) are integrated in AIMS to improve both low-level image segmentation and high-level image interpretation.

In summary, previous work has concentrated on extracting groups of features; using simple generic models for specified classes of objects; recognizing objects with exact models; or using additional sensing information. Our task is to detect objects with minimal knowledge and information about the objects and scenes. The objects are unspecified and may have much more complicated structures than the objects considered in the previous research.

3. OVERVIEW OF THE APPROACH

This section describes the basic concepts and presents an overview of our approach. The approach essentially organizes those features indicating man-made objects into structures, and finds the image region in which related structures reside.

The goal of our research is to detect man-made objects from images of natural scenes. Figure 1 shows an example of such an image. Since the objects are not particularly specified, features must be found that distinguish man-made objects from natural objects in an image. Two of the most prominent characteristics of man-made objects are the apparent regularity and relationship of their components. Most man-made objects have linear structures or linear boundaries. The linear structures form certain regular patterns such as rectangles, parallels, and polygons. These regular patterns are usually related to each other and form the man-made objects. After line detection, much of such regularity and relationship remain in the resulting image. In comparison, most natural objects do not have linear structures, and lines extractable from their images are usually randomly distributed.

To detect man-made objects, geometric structures with regularity and relationships must be extracted from the image. Hence, the framework of our approach includes three phases: (1) extracting image features, (2) finding regularities and relationships among these features, and (3) identifying the region occupied by the

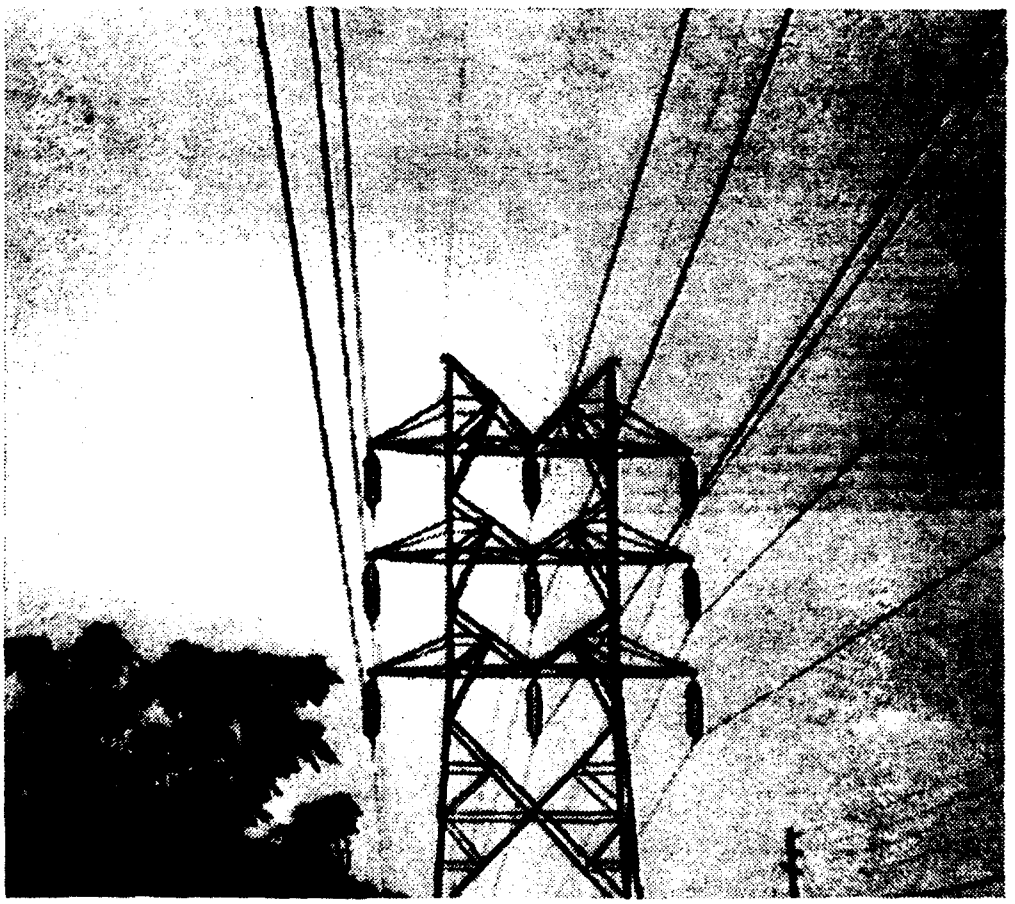


Fig. 1. An image of an electric transmission tower.

related regular structures. This indicates a three level hierarchical grouping process. An important issue in the grouping is *non-accidentalness*. Lowe[17] argues that perceptual groupings are useful when they are unlikely to have arisen by an accidental viewpoint or position and, therefore, are likely to reflect meaningful structures. Our computational approach to perceptual grouping should agree with this non-accidentalness principle.

The first level of grouping extracts image features. We currently consider two kinds of features: *linear structures* (LS) and *coterminations* (CT). A linear structure in an image is a representation of a set of approximately collinear line segments which are close and likely to come from the same linear structure in the scene. Extracting LS reflects the *proximity*, *collinearity*, and *continuation* of the perceptual grouping.

A cotermination is a set of lines terminating at a common point or a small common region. In practice, coterminous lines in a scene may not terminate at a common point in the image because of the imaging, digitizing, and line extraction processes. Instead, they may terminate in a small common region. The cotermination is an important relation. According to the *proximity* rule of perceptual organization, the human visual system easily groups coterminous lines. In fact, it is suggested that the major function of eye movements is to determine coterminous edges.[30] Cotermination is a non-accidental relationship and, hence, reflects significant structural information. It is also view-invariant in a wide range of viewpoints and can be used for 3D inference.[17] The CT are represented by a graph called a *CT graph*.

The second level grouping organizes features that exhibit regularity and relationship into larger structures called *primitive structures* (PS). Each PS may be evidence indicating man-made objects. In addition, a PS implies more constraints among line segments. Two kinds of PS are currently considered: *parallel PS* and *polygon PS*. A parallel PS is a set of parallel lines satisfying certain conditions. A large number of man-made objects contain parallel structures and human vision can rapidly identify parallel lines.[17] A parallel relation is a non-accidental relationship and can be used to infer relations in three-space.[17] A polygon PS is a closed figure that consists of line segments and satisfies certain criteria. It is also a significant image relation. According to the *closure* rule of perceptual organization, human vision tends to complete curves so that they form enclosed regions.[17] Extracting closed figures corresponds to this feature of human vision. Polygon PS are non-accidental image events since the coterminations forming them are non-accidental. Hence, they represent significant structures in an image. They are also higher level structures than lines and coterminations. Polygon PS can also be used for 3D inference. For example, a closed figure in an image suggests a closed structure in 3D space.

The PS are represented by a graph called a *PS graph* that describes the spatial relationships among the PS. Such a graph facilitates efficient higher level processing. Using the PS graph, the third phase of the framework groups spatially closed PS, eliminates the isolated ones, and segments the image into regions occupied by the grouped PS and a background. The largest region of the grouped PS is then evaluated based on the area of the region and the statistics of the PS. If this region is determined to be significant, it is most likely to enclose man-made objects or a substantial part of the man-made objects. Then it is called the ROI. The rationale for this level of grouping is the following. (1) Spatially closed PS are likely to be related and to reflect meaningful structures. For example, an electric transmission tower is a connected entity and, hence, the PS resulting from the image of the tower are spatially closed. On the other hand, spatially closed PS are more likely to be perceptually grouped according to the *proximity* grouping rule. (2) Some PS may be caused by accidental image relations of natural objects. For example, line segments extracted from a cluster of tree leaves may accidentally form a parallel PS. Such PS tend to be randomly and sparsely distributed and are unlikely to form meaningful structures. (3) Man-made objects usually consist of regular structures related structurally and spatially.

In summary, Fig. 2 illustrates the overall data representation, relationship, and flow in this framework. Down-arrows in the figure represent "consist of", up-arrows denote "derive", and lines express "represent". "Par-PS" and "Pol-PS" represent parallel PS and polygon PS, respectively. Figure 2 shows that CT and parallel PS consists of lines; CT are represented by the CT graph from which polygon PS are derived; and all PS are represented by the PS graph from which the ROI is derived. The data abstractions of the three level grouping are also indicated in Fig. 2.

The term *primitive structure* is also used in reference (26) to represent a larger class of entities, including edges, regions, parallelism, symmetry, repetition, and so forth. However, in this paper primitive structure represent the regular patterns formed by straight line

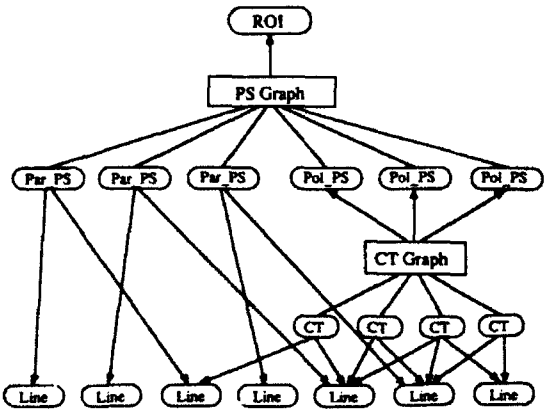


Fig. 2. The data representation and data relationship in the framework.

segments. Basically, we consider the finer classification of the structural entities, since these entities are not grouped at the same level. For example, a line segment is a grouping of edge points, whereas a parallel is a grouping of line segments.

The implementation of the framework presented in this paper consists of three modules: feature extraction, PS formation, and segmentation. Each module implements one phase of the framework. The next three sections describe the techniques developed for the implementation.

4. FEATURE EXTRACTION

This module first extracts linear edges (or line segments) from the input image. These edges are the basic information used for the grouping process. Linear structures and coterminations are then extracted from the linear edges.

4.1. *Linear structures*

We used an existing method[31] to detect straight line segments from the intensity image. For practical reasons, the extracted line segments may not reflect well the linear structures in the image. A post-processing is required to obtain the linear structures.

For this purpose, a *representative line* to a set of closely bunched and similarly oriented linear edges should be found, since these edges represent the linear structure of an object at a higher granularity level than the edges themselves.[8] For example, we want to extract the linear structure implied by a set of line segments shown in Fig. 3.

However, most of the collinearization techniques, such as references (9, 21, 24), are unsuitable for extracting linear structures, since they only link approximate collinear lines by examining the neighborhoods of the end points of each line; that is, they perform a line extension. Our objective of extracting linear structures is quite similar to that presented by Mohan and Nevatia.[8] In reference (8), the space around each line segment is *folded* onto the segment repeatedly to obtain a *single* line representing the grouped line segments. However, this folding technique does not perform a line extension. The techniques developed in this paper perform both folding and line extension simultaneously.

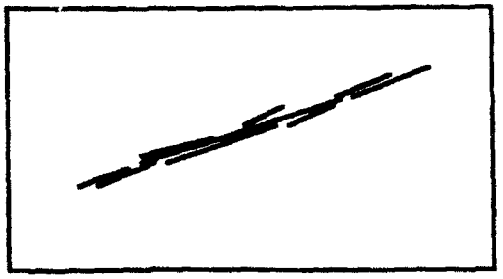


Fig. 3. Extracting a linear structure from a set of line segments.

According to Gestalt Laws of proximity, similarity, and continuation, a set of proximate lines with similar orientations tends to be perceived as a continuous line. The lines are likely to come from the same linear structure in the scene and, hence, should be merged into one line. Two techniques, the *neighborhood method* and the *classification method*, are developed to extract linear structures. The neighborhood method iteratively groups and merges similarly oriented lines in a neighborhood of each line segment. The classification method groups line segments by classifying them according to orientation, collinearity, and proximity.[32] In our current system, the two LS extraction modules are concatenated. The result of the LS extraction is a set of lines including the representative lines of the grouped line segments and the un-grouped line segments.

More details of the neighborhood method are presented below. We first introduce three definitions. Two lines have *similar* orientations if the angle between the two lines is less than a threshold, called the *similarity-angle*. The *neighborhood* of a line segment L is a symmetric elongated region with L as the *medial axis* of the region.[33] Two line segments are *close* if at least one end point of one line segment is in the neighborhood of the other line segment.

The idea of the grouping process is as follows. The neighborhood of each line is searched to find all lines with orientations similar to the current line, called the *base line*. The resulting set of lines, including the base line, is then replaced by a *representative line*. The process continues until no replacement occurs.

To reduce the search space, a line segment is represented by its two end points and is indexed by the image pixels corresponding to the end points. When searching for lines close to a base line, the neighborhood of the base line is searched. Hence, only those lines whose end points fall into this neighborhood are examined. After a set of lines S is found with respect to a base line L, with $L \in S$, a representative line L_r of S is computed. L_r passes through the point that is the geometric center of the line segments in S. The orientation of L_r is the length-weighted average of the orientations of the lines in S. To determine the end points of L_r, all the end points of the line segments in S are orthogonally projected onto L_r. The two furthest apart projection points are the end points of L_r. L_r replaces the lines in S. The process continues until no merge occurs. It always terminates after a finite number of iterations, since there is a finite number of lines and since this number declines in each iteration.

4.2. *Coterminations and their representation*

A *cotermination* (CT) is a set of lines terminating at a common point. For practical reasons, we only require lines in a CT to terminate in a small common region. To extract CT from a set of lines, we use a

two-step procedure. First, *elementary* CT are extracted. Then, closed elementary CT are grouped if they satisfy certain conditions. This procedure is similar to the one in reference (21) except that we allow more than two lines in an elementary CT. In addition, we derive a representation for CT and their relationships.

An elementary CT is a set C of lines terminating in a small common region: each pair of lines $\{L_i, L_j\} \in C$ satisfies the following conditions (see Fig. 4):

(1) $\delta \leq \theta \leq \pi - \delta$;

(2) $\max\{d_i, d_j\} \leq \frac{w}{2}$;

where θ is the angle between L_i and L_j, δ the *similarity-angle*, and $w \times w$ the size of a neighborhood. Let e_i and e_j be the end points of L_i and L_j that fall into the small common region; and let s be the intersection of the two lines. Then d_i and d_j are the distances from e_i and e_j to s, respectively. The above conditions are the same as the first three conditions in reference (21). Condition (1) excludes lines that are approximately collinear or that form a sharp angle. These situations are considered in the linear structure extraction, and such lines are unlikely to correspond to a CT in a scene. When an elementary CT is extracted, we say that line $L_i(L_j)$ joins the CT at $e_i(e_j)$. Similar to extracting LS, lines are represented and indexed by their end points to reduce the search space. For each line, only the neighborhoods of its end points are searched instead of searching all the other lines.

The extracted elementary CT are further grouped into CT. An elementary CT C initiates a CT. Other elementary CT are examined if they have a common line with C and if the line joins them and C at the same end point. For example, Fig. 5(a) has three elementary CT. C_1 contains lines L_1 and L_2, C_2 contains lines L_2 and L_3, and C_3 contains lines L_2 and L_4. If C_1 initiates a CT, C_2 is examined but C_3 is not because although

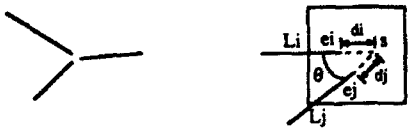


Fig. 4. A cotermination is a set of line segments terminating in a small common region.

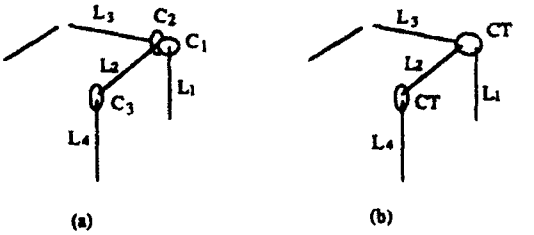


Fig. 5. Elementary CT are grouped into CT.

C_1 and C_3 have a common line L_2, L_2 joins them at two different end points. Let an elementary CT C_i be examined for a possible grouping with C. C_i is grouped with C if all the end points at which lines join C and those at which lines join C_i fall into a common region of size $w \times w$. All the elementary CT are thus processed, resulting in a set of CT. Some of the CT are just elementary CT while others are unions of elementary CT.

The CT and their relationships are represented by a graph called a *CT graph*. A representation is necessary to derive efficient and systematic procedures for higher level processing that locates and uses CT. In the CT graph, each vertex represents a CT in the image. Two vertices are connected by an edge in the graph if the two corresponding CT share a common line segment at the two different end points of the line. Figure 6 shows a set of line segments with six CT and the corresponding CT graph. The CT graph enables us to extract higher level structures.

5. PRIMITIVE STRUCTURE FORMATION

This section defines *primitive structures* (PS) and describes techniques of grouping features into PS. Two kinds of PS are discussed: parallel PS and polygon PS.

5.1. *Parallel primitive structure*

The distinction between the parallel PS defined in this paper and parallel lines as defined by previous researchers[3,8,23] is the requirement for the overlap between parallel lines. Usually, only a certain overlapping using an orthogonal projection is required.[3,8,23] For example, in Fig. 7(a), the overlapping between the line segments L_1 and L_2 using an orthogonal projection

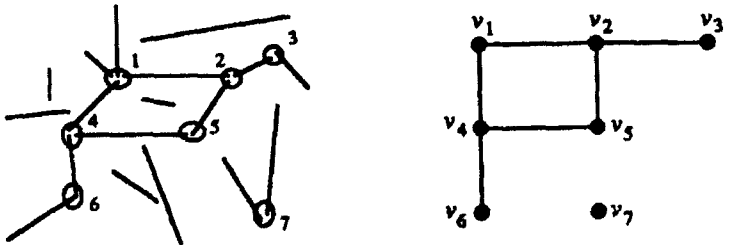


Fig. 6. A set of line segments, CT, and their CT graph.

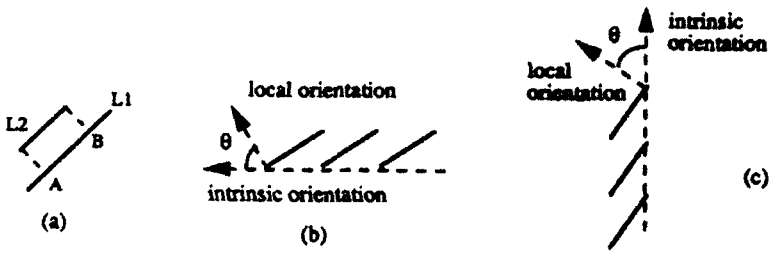


Fig. 7. Lines with overlapping in different directions.

is the line segment AB. However, according to such a definition, when the *intrinsic orientation* of a set of similarly oriented lines[34] is different from the local orientation of each line, these lines may not be grouped. As a result, many apparent parallel lines, such as those shown in Figs 7(b) and (c), will not be identified when there is an angular difference, θ, between the intrinsic and local orientations and when the overlapping between lines is small using the orthogonal projection. Such situations arise very often in practice. For example, a set of parallel lines in 3D space may fall into this situation under the 2D projection of the imaging process. To solve this problem, we define two additional overlapping conditions in two perpendicular directions. The idea here is similar to the *θ-aggregation* of Marr's earlier work.[34]

The parallel PS is a set of lines, $S = \{L_1, L_2, \ldots, L_m\}$, $m \geq 2$, that have *similar* orientations. In addition, for each line $L_i \in S$, there exists a line $L_j \in S$ such that:

(1) L_i and L_j have *similar lengths*.

(2) L_i and L_j have a *sufficient overlap* in one of the three projections, i.e.

$$\frac{\mathcal{O}(\mathcal{P}_x(L_i), \mathcal{P}_x(L_j))}{\mathcal{L}(\mathcal{P}_x(L_k))} > \delta_1$$

where $\mathcal{P}_x(\cdot)$ is the projection of a line onto the x-axis, $\mathcal{L}(\cdot)$ the length of a line, $\mathcal{O}(L, L') = \mathcal{L}(L \cap L')$ the length of the overlap, and L_k the shorter line of L_i and L_j; or

$$\frac{\mathcal{O}(\mathcal{P}_y(L_i), \mathcal{P}_y(L_j))}{\mathcal{L}(\mathcal{P}_y(L_k))} > \delta_1$$

where $\mathcal{P}_y(\cdot)$ is the projection of a line onto the y-axis; or

$$\frac{\mathcal{O}(\mathcal{P}_o(L_i), L_j)}{\mathcal{L}(\mathcal{P}_o(L_i))} > \delta_1$$

where L_i is assumed to be the shorter line and $\mathcal{P}_o(L_i)$ the orthogonal projection of L_i onto L_j.

(3) L_i and L_j are *relatively* close.

The above conditions are based on the perceptual organization rules of proximity, parallelism, and similarity. Condition (1) requires that two line segments have a similar length. Two lines with very different lengths are unlikely to come from a parallel structure and are unlikely to be perceptually grouped. Condition (2) indicates that two line segments in a PS should have a sufficient overlap. With the defined set of overlapping conditions, lines in Figs 7(a)–(c) can all be properly grouped. Condition (3) restricts the relative distance between the two line segments.

The line segments satisfying the above conditions are grouped into parallel PS. To avoid a brute-force search, line segments with similar orientations are first classified into clusters. A further grouping is performed within each cluster to find all parallel PS.

5.2. *Polygon primitive structure*

To extract polygon primitive structures, we first find closed figure (polygons) formed by line segments using coterminations. We then check them to determine if they satisfy the conditions of a polygon PS. Coterminations may form arbitrary polygons. Since our purpose of finding polygons is to extract significant image structures reflecting man-made objects, we want to retain closed figures with "regular" shapes as polygon PS and to discard those with "irregular" shapes. We also want to include a variety of polygons. For example, polygons in Fig. 8 will be retained and those in Fig. 9 will be discarded.

Formally, we define a polygon PS as follows. A closed figure P is a polygon PS if:

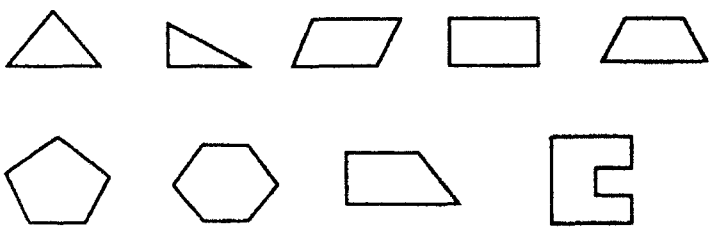


Fig. 8. These polygons are accepted as polygon PS.



Fig. 9. These polygons are discarded.

(1) P is a simple polygon, that is, the edges of P do not intersect among themselves;

(2) P is relatively compact: $\mathcal{F}(P) \leq \delta_2 \mathcal{F}(\mathcal{C}(P))$, where $\mathcal{F}(\cdot)$ is defined as[33]

$$\mathcal{F}(P) = \frac{perimeter^2(P)}{area(P)}$$

$\mathcal{C}(\cdot)$ is the convex hull of P, and $\delta_2 \geq 1$ is a constant;

(3) P does not have many cavities: $n_i \leq \delta_3 n_c$, where n_i is the number of vertices of P inside $\mathcal{C}(P)$, n_c the number of vertices of P on $\mathcal{C}(P)$, and $\delta_3 \leq 1$ is a constant;

(4) the number of edges on P does not exceed a given threshold.

Finding closed figures can be done by tracing the lines or CT. Starting from an end point of a line and going along the given lines or CT, if we can come back to the starting point, a closed figure is found. (Practically, as in extracting CT, we cannot expect a point connection between lines and to trace back to a point. Instead, we should consider small regions.) However, the time-complexity of this kind of direct search is exponential. Alternatively, we propose a novel polynomial-time algorithm based on Graph Theory.[35,36] (Appendix A details this algorithm.) The approach finds a set of independent and sufficient closed figures using the CT graph. The closed figures are then checked with the conditions for the polygon PS. The ones satisfying the conditions are accepted as polygon PS and the rest are discarded.

5.3. *The representation*

To facilitate higher level processing, the primitive structures, including parallel PS and polygon PS, are represented by a graph called the *PS graph*. Each node in the PS graph corresponds to one PS. Let $P_i = \mathcal{C}(PS_i)$ and $P_j = \mathcal{C}(PS_j)$ be convex hulls of two PS, PS_i and PS_j $(i \neq j)$, respectively. Then, the two nodes corresponding to PS_i and PS_j are connected in the PS graph if and only if P_i and P_j intersect, that is, $P_i \cap P_j \neq \phi$. Thus, the spatial relationships among PS are represented explicitly using the PS graph. For example, in Fig. 10(a), lines 1–3 form PS_1, lines 4 and 5 form PS_2, lines 3, 4, and 6 form PS_3, and lines 7 and 8 form PS_4. Figure 10(b) shows the convex hulls of these PS. Three convex hulls, P_1, P_2, and P_3 overlap each other and, hence, the corresponding nodes in the PS graph are connected. Figure 10(c) shows the PS graph.

The PS graph can be generalized if we define a spatial distance between two PS. For example, the distance between PS_i and PS_j $(i \neq j)$ may be defined as the minimum distance between P_i and P_j if $P_i \cap P_j = 0$, and as zero otherwise. Then the two nodes corresponding to PS_i and PS_j are connected in the PS graph if and only if the distance between PS_i and PS_j is smaller than or equal to a given threshold. The advantage of such a generalization is the ability to include *closeness* in the next level of grouping. The price to be paid is the increased computation.

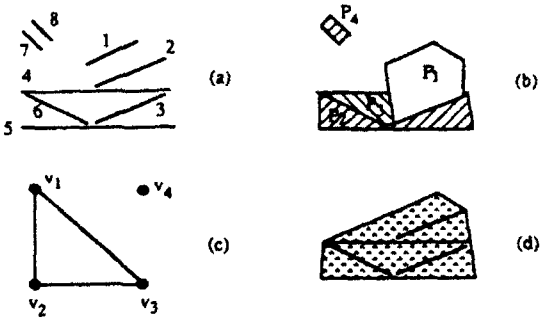


Fig. 10. Four primitive structures and their PS graph.

6. SEGMENTATION

The segmentation module segments the given image to obtain ROI that is most likely to enclose man-made objects or a substantial part of the man-made objects. The significance of this segmentation is then evaluated.

6.1. *Grouping related PS*

As discussed earlier, spatially closed PS are likely to be related and to reflect meaningful structures, whereas randomly and sparsely distributed PS are likely to be formed accidentally. Therefore, for man-made object detection, we want to find spatially related PS and to eliminate accidental PS. Each PS occupies a region in the image. For example, a PS containing two parallel lines occupies a trapezoidal region. The regions of spatially proximate PS tend to overlap, touch, or be close, whereas the regions of sparsely located PS are isolated. Hence, we can segment an image by grouping PS whose regions intersect or are close. The largest image region occupied by the grouped PS is considered as the ROI if the significance of this region exceeds a given threshold.

An algorithm is presented to achieve the segmentation using the PS graph. The region of a PS is defined as the convex hull of the line segments in the PS. The connectivity of the nodes in the PS graph, hence, reflects the intersection and the closeness of the regions of the PS. Each connected component with more than one node of the PS graph corresponds to a set of PS whose regions intersect or are close. Therefore, segmenting the image into regions that include those occupied by the grouped spatially related PS and a background can be accomplished by finding connected components of the PS graph.

Given the PS graph, we first find the connected components of the graph. For each component, the corresponding PS are grouped into a larger structure. This structure is again called a PS for the simplicity of the algorithm description. A new PS graph is then derived for the new PS obtained from the grouping. The connected components of the graph are found, leading to further grouping of the PS. This process continues until no PS can be grouped. Our experiments show that only a small number of iterations are actually needed. More details of this algorithm are given in Appendix B. This process results in an image segmentation that includes regions occupied by the grouped PS and a background. Currently, the largest region $\mathcal{R}$ of the grouped PS is evaluated for consideration as the ROI.

6.2. *Significance evaluation*

The significance of the region $\mathcal{R}$ obtained in the segmentation is evaluated. If $\mathcal{R}$ is significant, then it is the ROI most likely to enclose man-made objects or a substantial part of the man-made objects. Naturally, the area of $\mathcal{R}$ is a measure of significance. If the area is very small, $\mathcal{R}$ is not an ROI. Since PS are likely to arise from man-made objects, the probability of PS falling into $\mathcal{R}$ and the probability of lines forming PS in $\mathcal{R}$ are other measures of significance.

Let A be the event that lines form a PS, B be the event that lines fall outside $\mathcal{R}$, and C be the event that lines fall inside $\mathcal{R}$. Let N, N_B, and N_C be the numbers of lines in the image, outside $\mathcal{R}$, and inside $\mathcal{R}$, respectively; and let M, M_B, and M_C be the numbers of lines forming PS in the image, outside $\mathcal{R}$, and inside $\mathcal{R}$, respectively. Then the probability of lines forming PS is $P(A) = M/N$. The probability of a PS being inside $\mathcal{R}$ is $P(C|A) = M_C/M$. The probability of lines in $\mathcal{R}$ forming PS is $P(A|C) = M_C/N_C$. The probability of lines outside $\mathcal{R}$ forming PS is $P(A|B) = M_B/N_B$. Since most PS are expected to come from man-made objects, and since an ROI should be the region occupied by these PS, the probability that a PS falls into the ROI should be high and the probability that lines form PS in the ROI should be higher than that outside the ROI. Therefore, the significance of $\mathcal{R}$ is determined using the following three conditions:

(1) $P(C|A) \geq \max\{0.5, \mathcal{A}_r\} + \varepsilon$;
(2) $P(A|C) \geq P(A|B) + \varepsilon$;
(3) $\mathcal{A}_r > \mathcal{A}_0$;

where $\mathcal{A}_r = Area(\mathcal{R})/Area(Image)$, $\mathcal{A}_0$ is a threshold, the ε is a small number.

Each test result for the above conditions is represented by a function defined in the interval $[0, 1]$, indicating how well $\mathcal{R}$ satisfies the condition. The first two functions p_1 and p_2 are of the same form representing the first two conditions. Let the left-hand sides of conditions (1) and (2) be f and the right-hand sides be g. Then p_i, $i = 1, 2$, are defined as

$$p_i = \begin{cases} 1 & f \geq g \\ \dfrac{f-(g-\Delta)}{\Delta} & g-\Delta \leq f \leq g \\ 0 & f < g-\Delta \end{cases}$$

where Δ is a number controlling the drop of p_i from 1 to 0. The last function p_3 is defined as

$$p_3 = \begin{cases} 1 & \mathcal{A}_r \geq \mathcal{A}_0 \\ \dfrac{\mathcal{A}_r}{\mathcal{A}_0} & \mathcal{A}_r < \mathcal{A}_0 \end{cases}.$$

The significance of region $\mathcal{R}$ is defined as

$$p = \frac{\sum_{i=1}^{3} \omega_i p_i}{\sum_{i=1}^{3} \omega_i}$$

where ω_is are weights representing the relative importance of the conditions. Hence, p is defined in the interval $[0, 1]$. We select the equal weights in the following examples. If p of $\mathcal{R}$ is equal or close to 1, $\mathcal{R}$ is the ROI. Otherwise, $\mathcal{R}$ is rejected.

In summary, we have presented techniques to group low level image features hierarchically into an ROI likely to enclose man-made objects or a substantial part of the man-made objects. These techniques include feature extraction, primitive structure formation, and segmentation. Some of these techniques are novel and others present unique properties and advantages compared to previous related work.

7. EXPERIMENTAL RESULTS

This section presents several examples of finding ROI from given images. In these examples, the *similarity-angle* is 5 deg and δ_1 for overlapping conditions is 60%. A single monochrome image is used for each of the examples. All the image sizes are 480 × 512. The simple PS graph is used, that is, the intersection of convex hulls instead of the distance determines the node connectivity.

Figure 1 shows an image containing an electric transmission tower. The image is processed by Burns' algorithm to generate line segments (Fig. 11). These lines enter the Feature Extraction module. Very short lines (less than four pixels long) are eliminated; lines likely to come from the same linear structures are merged into one line; and coterminations are extracted. Figure 12 shows the lines resulting from the LS extraction. Figure 13 shows the coterminations. The straight lines and CT then enter the PS Formation module to identify primitive structures. Figure 14 illustrates the parallel groups thus obtained, and Fig. 15 shows polygon PS. Figure 16 shows the located ROI overlapped on the line image (Fig. 12). The significance of the ROI is 1.0. The ROI is bounded by a polygon displayed with a bold outline. From Fig. 16, we see that the tower

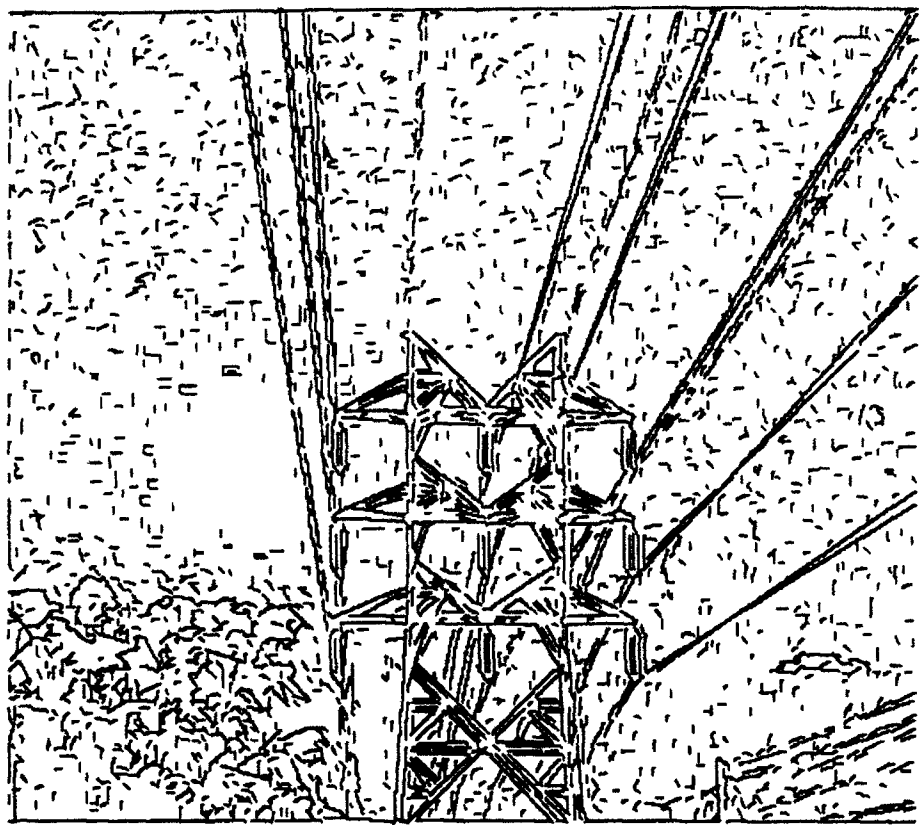


Fig. 11. Linear edges detected from the tower image.

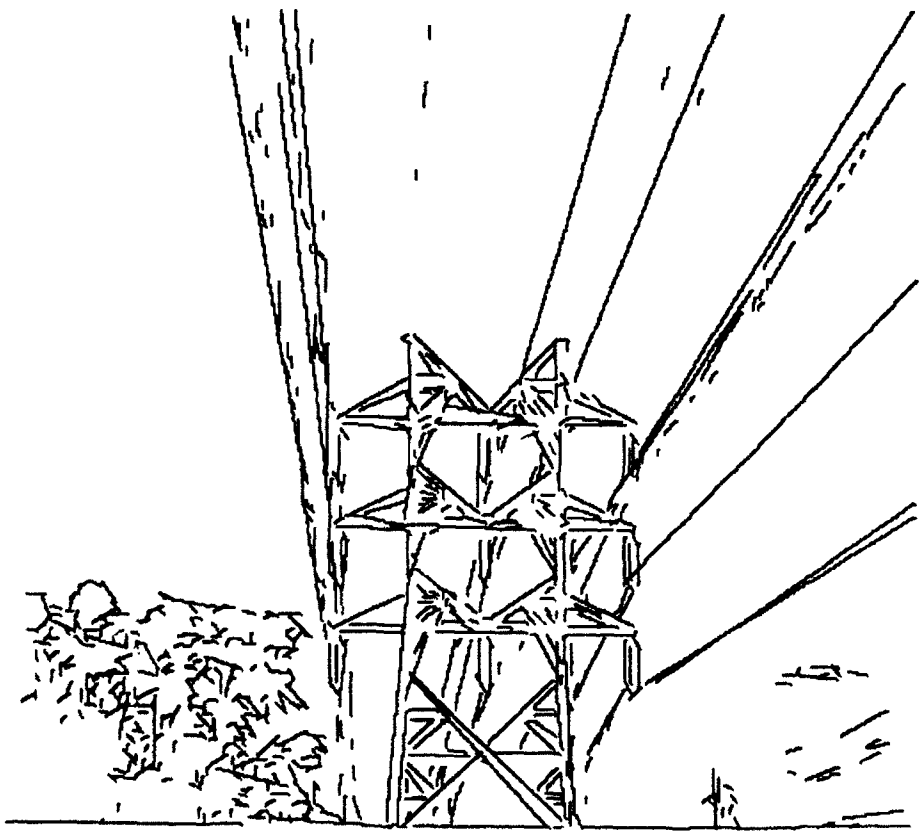


Fig. 12. Line segments after the LS extraction for the tower image.

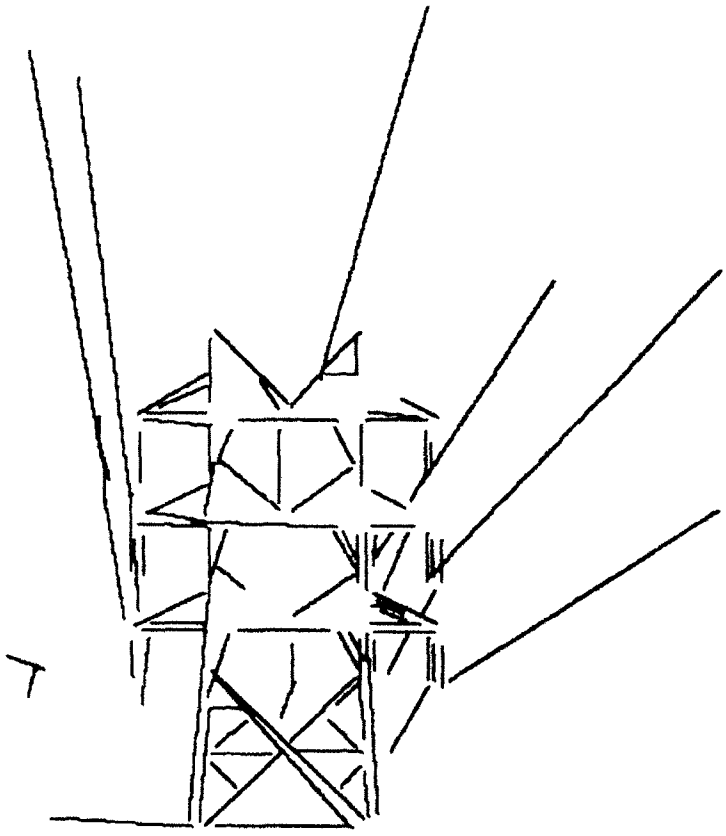


Fig. 13. Coterminations extracted from the tower image.

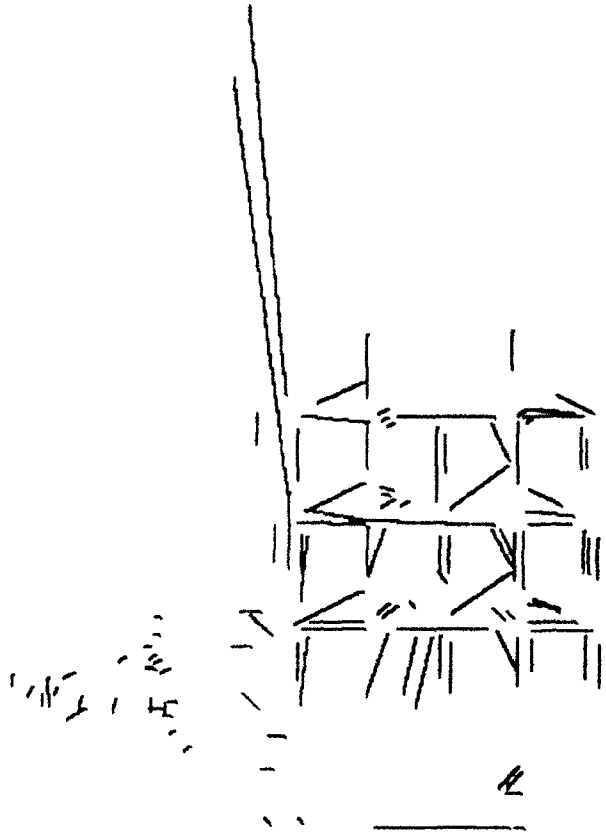


Fig. 14. Parallel lines of the tower image.

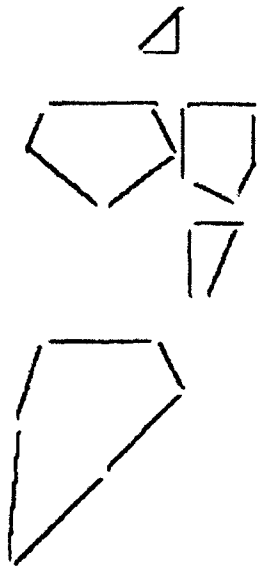


Fig. 15. Extracted polygon PS of the tower image.

and some of the transmission lines are properly included inside the polygon. Other transmission lines are not included because they do not form primitive structures. Although the cluster of tree leaves is very close to the tower, it is properly separated from the ROI. We can also see the random distribution of the linear edges detected from the region of leaves.

Figure 17 shows another tower image and Fig. 18 is the located ROI overlaid on the lines after LS extraction. The polygon with the bold outline is the ROI. The significance of the ROI is 1.0. The tower and most of the transmission lines are properly enclosed inside the polygon.

Figure 19 shows an image of a concrete bridge, trees, and a river. The located ROI is shown in Fig. 20. The lines in Fig. 20 are the output of the LS Extraction. The polygon with the bold outline is the ROI. The significance of the ROI is 1.0. From Figs 19 and 20, we can see that most of the bridge is enclosed inside the ROI except for its far end near the image boundary. Trees and the river are properly excluded from the ROI.

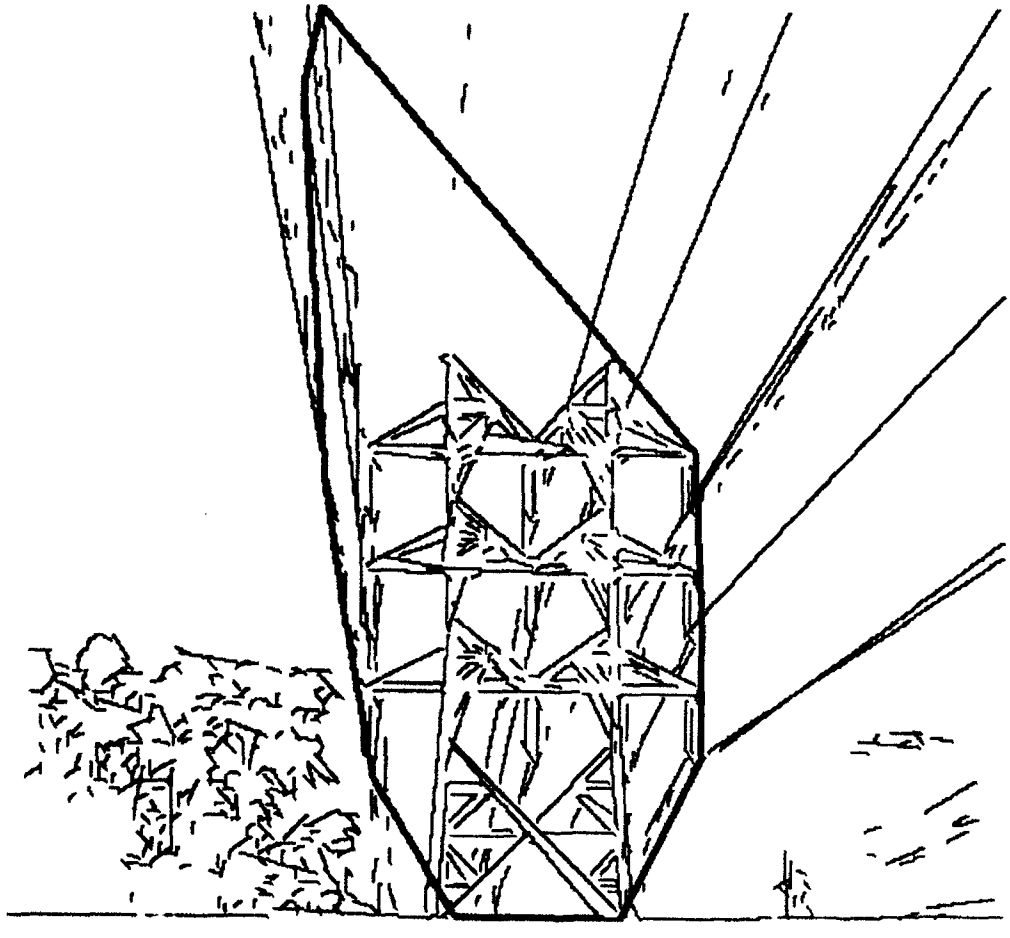


Fig. 16. The region of interest for the tower image.

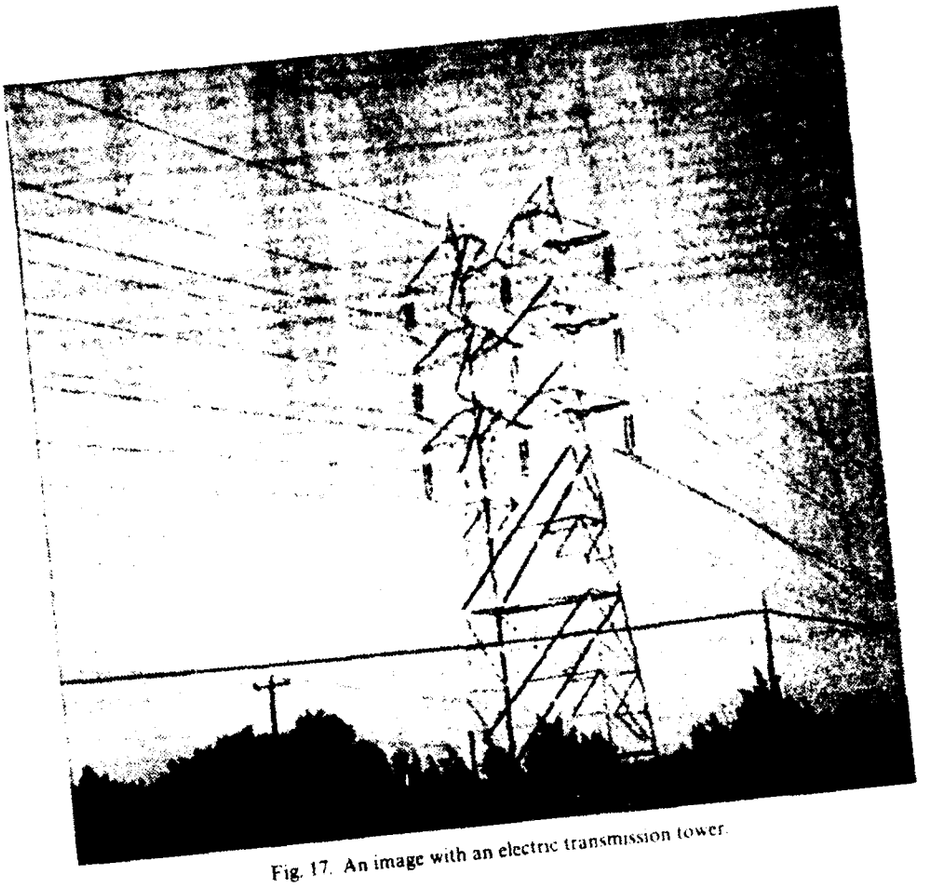


Fig. 17. An image with an electric transmission tower.

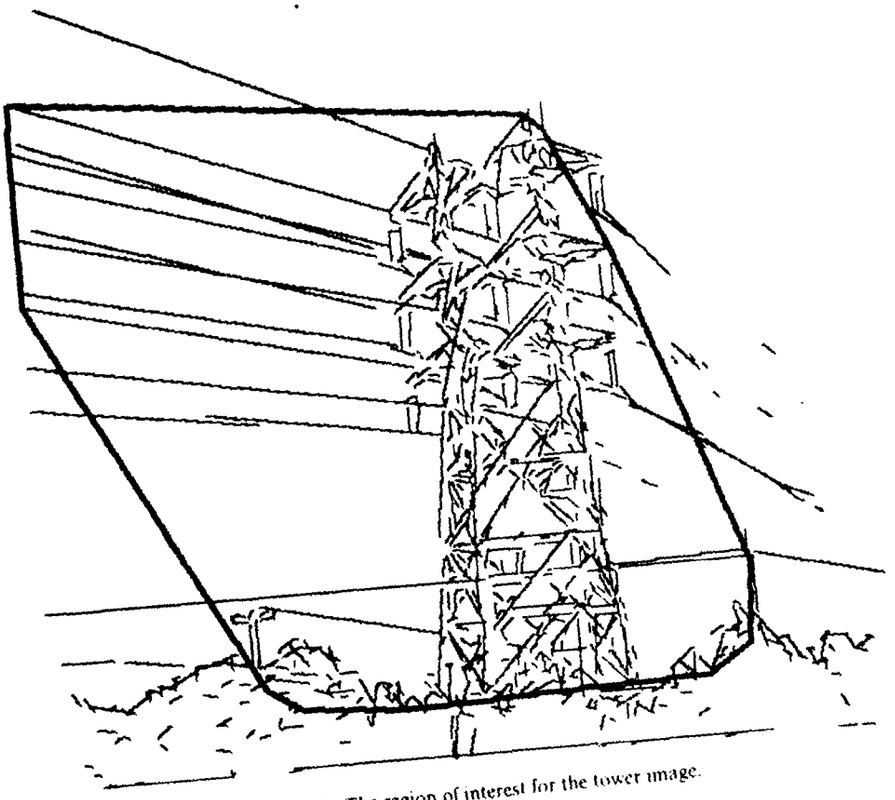


Fig. 18. The region of interest for the tower image.

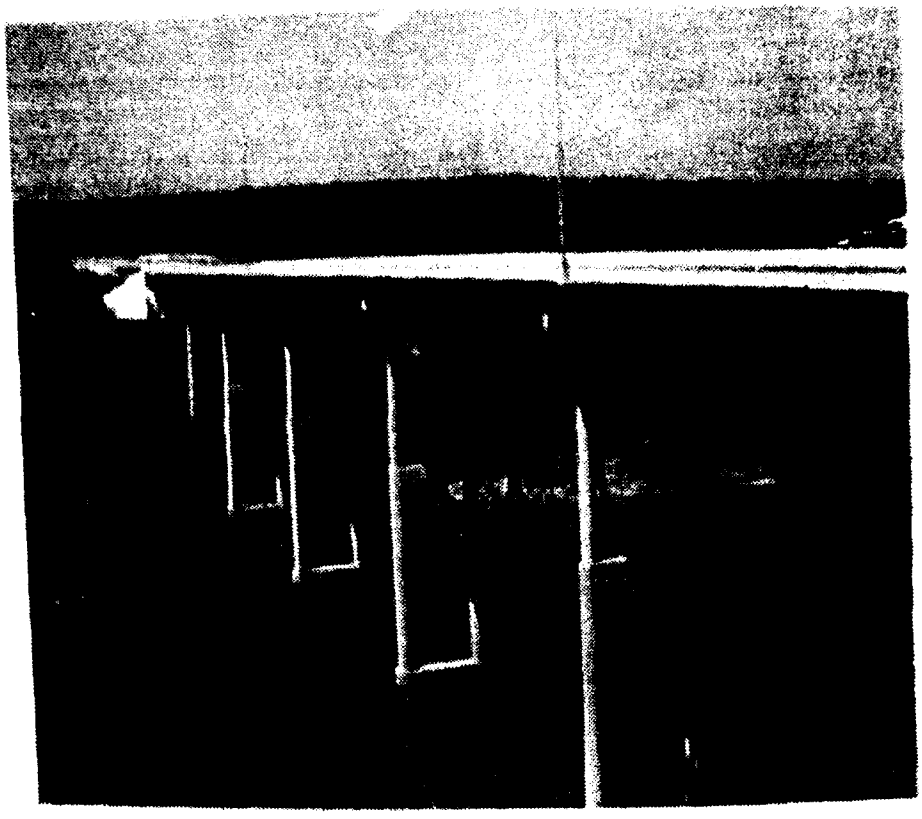


Fig. 19. An image with a bridge.

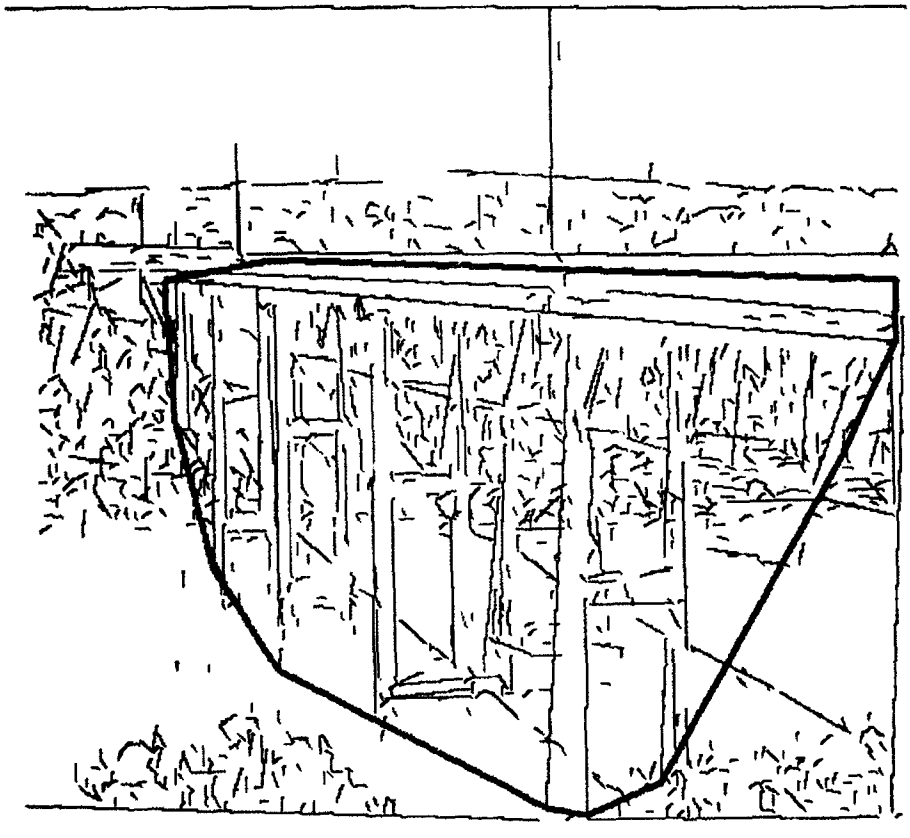


Fig. 20. The region of interest for the bridge image.

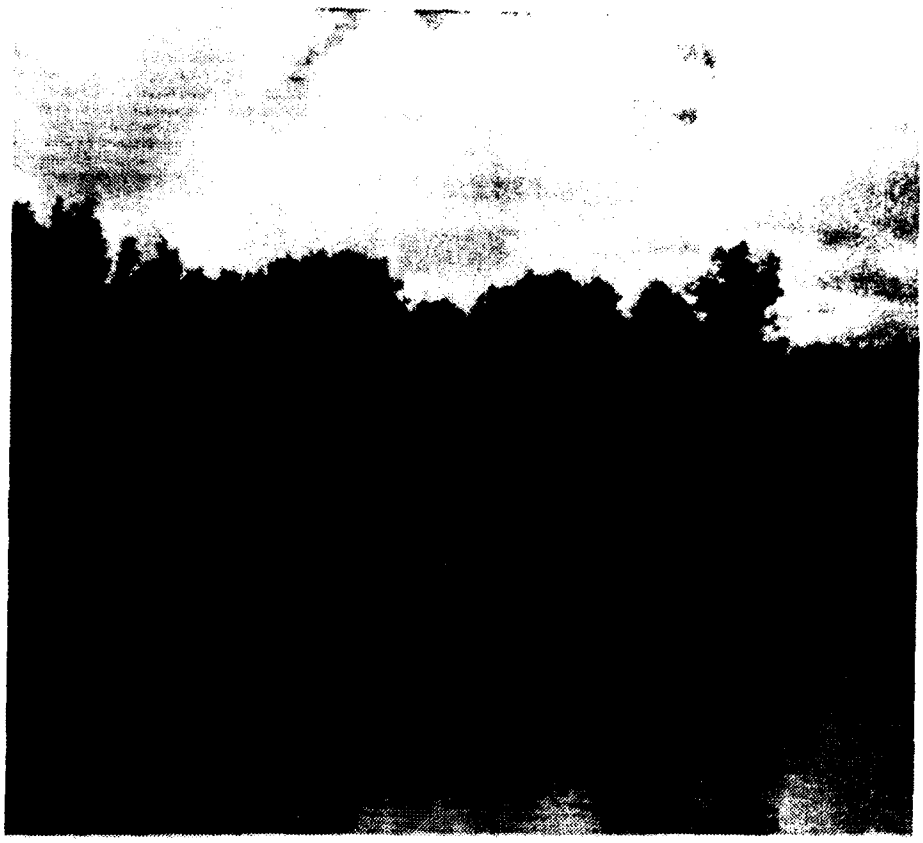


Fig. 21. An image of a natural scene.



Fig. 22. The located polygon is too small to be an POI

Figure 21 is a natural scene of trees, sky, and a pond. The polygon with a bold outline shown in Fig. 22 is the largest region obtained by the grouping process. The region is not accepted as an ROI because its significance is very low ($p = 0.37$). Therefore, no ROI is found in this image. In fact, there are no large man-made objects in the image.

8. CONCLUSION

This paper presents a new approach for the detection of large man-made objects in a non-urban area using a single monochrome image. The research shows how minimal knowledge and information about the domain can best be used for a computer vision task. Prominent features are identified that distinguish man-made objects from natural objects. We propose a computational framework to locate an ROI most likely to enclose man-made objects or a substantial part of the man-made object. The framework is based on the principles of perceptual organization. Techniques are developed to group low level image features hierarchically into the ROI. Some of these techniques are novel, while others present unique properties and advantages compared with previous related works.

The paper presents various examples, including images of different kinds of man-made objects in complex backgrounds as well as an image of a natural scene without man-made objects, to show the approach's effectiveness. As the examples show, this approach is capable of locating a useful ROI in complex real images. Hence, the search space is reduced from the whole image to the ROI. Further analysis of the ROI may lead to the identification of an object or to the rejection of the existence of man-made objects in the scene. Therefore, this technique of locating the ROI may be used for an initial screening of a large number of images for automatic object recognition or for a human–machine system. For an automatic system, when specific object classes are given and models are established, primitive structures composing the ROI can be matched to object models instead of matching individual features. This will considerably reduce the search space for matching since more constraints are applied. For a human–machine system, the ROI can be used as a focus-of-attention for human expertise to further examine the image.

We have currently used straight lines, coterminations, parallel PS, and polygon PS in the grouping process. Hence, man-made objects without such structures are not detected. Future research will consider features and primitive structures related to curves, in addition to straight lines.

Acknowledgement—This research was supported in part by the Army Research Office (contract DAAL03-91-G-0050) and JSEP (AFSC) (contract F49620-89-C-0044), and in part by the General Dynamics (purchase order WPC001023).

REFERENCES

1. R. T. Chin and C. R. Dyer, Model-based recognition in robot vision, *ACM Comput. Surveys* **18**, 67–108 (1986).
2. M. A. Fischler and T. M. Strat, Recognizing objects in a natural environment: a contextual vision system (CVS), *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 774–796, May (1989).
3. D. C. Baker, S. S. Hwang and J. K. Aggarwal, Detection and segmentation of man-made objects in outdoor scenes: concrete bridges, *J. Opt. Soc. Am. A* **6**, 938–950 (1989).
4. D. C. Baker, Incremental segmentation for interpretation if man-made objects in color images of non-urban scenes: concrete bridges, Ph.D. Thesis, The University of Texas at Austin, May (1989).
5. R. A. Brooks, Model-based three-dimensional interpretations of two-dimensional images, *IEEE Trans. Pattern Analysis Mach. Intell.* **5**, 140–150 (1983).
6. C. C. Chu and J. K. Aggarwal, Interpretation of laser radar images by a knowledge-based system, *Mach. Vision Applic.* **4**, 145–163 (1991).
7. C. C. Chu, N. Nandhakumar and J. K. Aggarwal, Image segmentation using laser radar data, *Pattern Recognition* **23**, 569–581 (1990).
8. R. Mohan and R. Nevatia, Using perceptual organization to extract 3-D structures, *IEEE Trans. Pattern Analysis Mach. Intell.* **11**, 1121–1139 (1989).
9. R. Nevatia and K. R. Babu, Linear feature extraction and description, *Comput. Graphics Image Process.* **13**, 257–269 (1980).
10. P. Fua and A. J. Hanson, Objective functions for feature discrimination: applications to semiautomated and automated feature extraction, *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 676–694, May (1989).
11. P. Fua and A. Hanson, Using generic geometric models for intelligent shape extraction, *Proc. DARPA Image Understanding Workshop*, Los Angeles, California, pp. 227–233, February (1987).
12. P. Fua and A. J. Hanson, Resegmentation using generic shape: locating general cultural objects, *Pattern Recognition Lett.* **5**, 243–252 (1987).
13. D. Kuan, G. Phipps and A.-C. Hsueh, Autonomous robotic vehicle road following, *IEEE Trans. Pattern Analysis Mach. Intell.* **10**, 648–658 (1988).
14. C. Thorpe, M. H. Hebert, T. Kanade and S. A. Shafer, Vision and navigation for the Carnegie–Mellon Navlab, *IEEE Trans. Pattern Analysis Mach. Intell.* **10**, 362–373 (1988).
15. R. Kumar, Determination of camera location and orientation, *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 870–881, May (1989).
16. R. Talluri and J. K. Aggarwal, Positional estimation for a mobile robot in an outdoor environment, *IEEE Workshop on Intelligent Robots and Systems*, Tsuchiura, Ibaraki, Japan, pp. 159–166, July (1990).
17. D. G. Lowe, *Perceptual Organization and Visual Recognition*. Kluwer Academic, Hingham, Masschusetts (1985).
18. J. D. McCafferty, *Human and Machine Vision—Computing Perceptual Organisation*. Ellis Horwood, West Sussex, U.K. (1990).
19. D. W. Jacobs, GROPER: a grouping based recognition system for two dimensional objects, *Proc. IEEE Computer Society Workshop on Computer Vision*, Miami Beach, Florida, pp. 164–169 (1987).
20. R. Mohan and R. Nevatia, Perceptual organization for segmentation and description, *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 415–424, May (1989).
21. R. Horaud, F. Veillon and T. Skordas, Finding geometric and relational structures in an image, *Proc. European Conf. on Computer Vision*, Antibes, France, pp. 374–384, April (1990).

22. J. Dolan and R. Weiss, Perceptual grouping of curved lines, *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 1135–1145, May (1989).
23. G. Reynolds and J. R. Beveridge, Searching for geometric structure in images of natural scenes, *Proc. DARPA Image Understanding Workshop*, Los Angeles, California, pp. 257–271, February (1987).
24. R. Weiss and M. Boldt, Geometric grouping applied to straight lines, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Miami Beach, Florida, pp. 489–495, June (1986).
25. M. Wertheimer, Principles of perceptual organization, *Readings in Perception*, D. Beardslee and M. Wertheimer, eds, pp. 115–135 (1958).
26. A. P. Witkin and J. M. Tenenbaum, On perceptual organization, *From Pixels to Predicates*, A. P. Pentland, ed., pp. 149–169. Ablex, Norwood, New Jersey (1986).
27. J. Princen, J. Illingworth and J. Kittler, A hierarchical approach to line extraction, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, San Diego, California, pp. 92–97, June (1989).
28. P. Fua and A. J. Hanson, Objective functions for feature discrimination: theory, *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 443–460, May (1989).
29. J. R. Beveridge, R. Weiss and E. M. Riseman, Optimization of 2-dimensional model matching, *Proc. DARPA Image Understanding Workshop*, Palo Alto, California, pp. 815–830, May (1989).
30. T. O. Binford, Inferring surfaces from images, *Artif. Intell.* **17**, 205–244 (1981).
31. J. B. Burns, A. R. Hanson and D. M. Riseman, Extracting straight lines, *IEEE Trans. Pattern Analysis Mach. Intell.* **8**, 425–455 (1986).
32. H. Q. Lu and J. K. Aggarwal, Extracting geometric structures and a region of interest for man-made object detection, *The 20th AIPR Workshop on Computer Vision Applications*, McLean, Virginia, October (1991).
33. D. H. Ballard and C. M. Brown, *Computer Vision*. Prentice-Hall, Englewood Cliffs, New Jersey (1982).
34. D. Marr, Early processing of visual information, *Phil. Trans. R. Soc. London* **275**, 483–524 (1976).
35. A. Gibbons, *Algorithmic Graph Theory*. Cambridge University Press, London (1985).
36. N. Deo, *Graph Theory with Applications to Engineering and Computer Science*. Prentice-Hall, Englewood Cliffs, New Jersey (1974).

APPENDIX A. FINDING CLOSED FIGURES

In order to extract polygon primitive structures, we need to find closed figures (polygons) formed by line segments. This appendix derives a method of finding a set of independent and sufficient closed figures using a CT graph. From the definition of the CT graph, it is clear that a one-to-one correspondence exists between the closed figures of line segments in an image and the *circuits* in the CT graph. Thus, the problem of finding closed figures can be solved by finding circuits in the graph. There may be many closed figures in an image, and some of them may be redundant. For example, two polygons sharing one common line form a third polygon, which is dependent on the other two. Since our objective is to find primitive structures, we want to find a set of closed figures that have a simple form (that is, the number of edges is small), are independent (to avoid redundancy), and are sufficient (to be able to generate all closed figures). According to graph theory, to satisfy the latter two requirements, we just need to find a set of *fundamental circuits* in the CT graph.

The set of fundamental circuits is not unique. Corresponding to each *spanning tree* of a graph, there is a set of fundamental circuits. To meet the simplicity requirement, we want to find a set of fundamental circuits such that the number of edges in the set is small. This may be formulated as a combinatorial optimization problem: given a graph, find a spanning tree such that the number of edges in the set of fundamental circuits corresponding to this tree is minimized. We can require the total number of edges to be minimized or the number of edges in the largest circuit to be minimized. If the graph is planar and if the graph keeps the geometry of the line segments in the image, the optimal solution is the set of *meshes* (*faces*). However, a CT graph may be non-planar because line segments may cross each other in an image. To our knowledge, there is no efficient algorithm available to the above optimization problem for a non-planar graph. Alternatively, we develop a heuristic solution, which may not be mathematically optimal but is efficient.

The basic idea is to find the "*bushiest*" spanning tree. For example, Fig. A1 shows a graph and two of the spanning trees of the graph. All the fundamental circuits corresponding to tree (a) have three edges whereas those corresponding to tree (b) have 3, 4, 5, 6, 7, and 6 edges, respectively. In this example, tree (a) is the bushiest spanning tree among all the spanning trees, and the corresponding fundamental circuits have the minimum number of edges. To formally define the *bushiest spanning tree*, we first define the *weight* of an edge in the graph.

Let $G = (V, E)$ be a CT graph, where V and E are the set of vertices and the set of edges of G, respectively. Let $e_{ij} \in E$ be an edge connecting vertices $v_i, v_j \in V$. Then the weight of e_{ij} is defined as: $w(e_{ij}) = d(v_i) + d(v_j)$, where $d(\cdot)$ is the *degree* of a vertex, that is, the number of edges *incident* with the vertex. The weight of a spanning tree $\mathcal{T}$ is the sum of the weights of all the branches in $\mathcal{T}$. A spanning tree with the largest weight *in G is called the maximal spanning tree*. We define the bushiest spanning tree as the maximal spanning tree. Let us look at the example in Fig. A1 again. The numbers on each edge of the graph indicate the weights of the edges. Obviously, tree (a) is the maximal spanning tree and is the bushiest one. The maximal spanning tree can be efficiently found by slightly modifying a minimal spanning tree algorithm.[35] Given a spanning tree of a graph, the set of fundamental circuits corresponding to this tree can be found in polynomial-time.[35]

The procedure for finding a set of closed figures satisfying our requirements is now straightforward. After extracting CT, the CT graph is established that is represented by an *adjacency matrix*. We then find the connected components of the graph and process each subgraph corresponding to each component separately, using the following steps. First the weights for all the edges in the subgraph are calculated. Then the maximal spanning tree of the subgraph is found. Finally, we find the fundamental circuits corresponding to the spanning tree. Each fundamental circuit represents a closed figure in the image, where edges on this circuit correspond to line segments on the closed figure.

The above procedure is efficient. Let n be the number of vertices in G. Then the time for finding connected components is $O(n)$. Let G_i be a connected component of G, and let n_i and m_i be the number of vertices and the number of edges in G_i, respectively. The time for finding the edge weights is $O(m_i)$, and that for finding the maximal spanning tree is $O(n_i^2)$. Finding fundamental circuits requires $O(n_i^2)$-time for a sparse graph and $O(n_i^3)$-time generally. (A CT graph is most likely to be a sparse graph.) Therefore, our algorithm is usually executed in $O(n^2)$-time and, at worst, in $O(n^3)$-time. It is, hence, a polynomial-time algorithm while the direct search takes exponential-time. Another advantage of using a CT graph to find closed figures is that the existence of closed figures among a given set of line segments can be readily determined. In fact, there are circuits in a graph only if the number of edges is greater than or equal to the number of vertices in the graph.

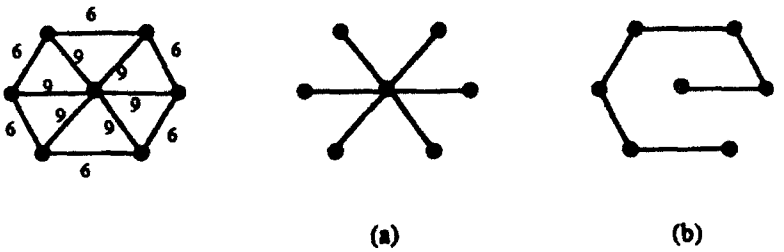


Fig. A1. A graph and two of its spanning trees.

APPENDIX B. GROUPING PRIMITIVE STRUCTURES

This appendix presents an algorithm that groups PS using the PS graph. Let $\mathcal{P}^0 = \{P_i^0 = \mathcal{G}(PS_i), i = 1, 2, \ldots, n_0\}$ be the set of convex hulls of the PS. Let $\phi: \mathcal{P} \mapsto \mathcal{G}$ be the function that maps the convex hull set $\mathcal{P}$ to the PS graph. Let $\mathcal{G}^0 = (\mathcal{V}^0, \mathcal{E}^0) = \phi(\mathcal{P}^0)$ be the PS graph of the PS in the image, where $\mathcal{V}^0$ is the set of vertices, and $\mathcal{E}^0$ the set of links. An algorithm described in reference (36) is applied to find the connected components of a graph.

Let $\mathcal{U}^0 = \{\mathcal{U}_i^0, i = 1, 2, \ldots, m_0\}$ be the set of connected components of $\mathcal{G}^0$. Then $\mathcal{U}^0 \subseteq \mathcal{G}^0$, and $\phi^{-1}(\mathcal{U}_i^0) \subseteq \mathcal{P}^0$ is either one convex hull or a set of convex hulls that intersect or are close. In the latter situation, the PS whose regions belong to $\phi^{-1}(\mathcal{U}_i^0)$ are grouped into a larger structure PS_i^1. The region of the PS_i^1 is

$$P_i^1 = \mathcal{G}(PS_i^1) = \mathcal{G}\left(\bigcup_{\forall P_j^0 \in \phi^{-1}(\mathcal{U}_i^0)} P_j^0\right).$$

For example, in Fig. 10, the regions (convex hulls) of the three PS intersect. Figure 10(d) shows the region of the grouped PS when these three PS are grouped. If $\mathcal{U}_i^0$ has only one vertex, that is, $\phi^{-1}(\mathcal{U}_i^0)$ represents one PS, PS_j, then $PS_i^1 = PS_j$, and $P_i^1 = P_j^0$. All the connected components of $\mathcal{G}^0$ are thus processed to produce possible groupings of PS and, hence, a segmentation of the image.

A new PS graph $\mathcal{G}^1 = (\mathcal{V}^1, \mathcal{E}^1) = \phi(\mathcal{P}^1)$ is then established, where $\mathcal{P}^1 = \{P_i^1 = \mathcal{G}(PS_i^1), i = 1, 2, \ldots, n_1\}$. The connected components of $\mathcal{G}^1$ are found, leading to further grouping of PS. This process continues until no PS can be further grouped. That is, the iteration stops at $\mathcal{G}^k = (\mathcal{V}^k, \mathcal{E}^k)$ with $\mathcal{E}^k = \varnothing$. Our experiments show that only a small number of iterations is actually needed.

About the Author—PROFESSOR J. K. AGGARWAL has served on the faculty of the University of Texas at Austin College of Engineering since 1964 and is now the Cullen Professor of Electrical and Computer Engineering and Director of the Computer and Vision Research Center. His research fields include computer vision, parallel processing of images, and pattern recognition. He has been a Fellow of IEEE since 1976. Dr Aggarwal is the author or editor of 6 books and 20 book chapters; author of 130 journal papers, 11 re-printed articles, over 137 proceedings papers and abstracts, and over 50 technical reports; he has made over 200 technical presentations and edited 8 proceedings of conferences or special issues of journals. He served as Chairman of the IEEE Computer Society Technical Committee on Pattern Analysis and Machine Intelligence (1987–1989); Director of the NATO Advanced Research Workshop on Multisensor Fusion for Computer Vision, Grenoble, France (1989); and was Chairman of the Computer Vision Program Committee, 10th International Conference on Pattern Recognition, Atlantic City (1990). Currently, he serves as IEEE Computer Society representative on International Association for Pattern Recognition and is the Treasurer of International Association for Pattern Recognition.

About the Author—HONG-QIAN (KAREN) LU received a Ph.D. degree from the Department of Electrical Engineering at the University of Pittsburgh, in 1989. From 1989 to 1991, she was a research associate with the Computer and Vision Research Center at the University of Texas at Austin. Currently, she is a Project Engineer with the Schlumberger Austin Systems Center. Her interests include computer vision, digital image/signal processing, and computer graphics.

A MOBILE ROBOT FOR VISUAL MEASUREMENTS IN ARCHITECTURAL APPLICATIONS

Xavier Lebègue and J. K. Aggarwal

Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712, USA

ABSTRACT

This paper describes a method to perform distance measurements inside buildings using a single video camera mounted on a mobile robot. This algorithm is designed specifically to create architectural floor plans with accurate measurements. Several issues associated with architectural surveying are discussed in detail, such as the visibility of features and the accuracy in calibration. Results of the edge reconstruction in an indoor scene are presented and compared to an architect's drawing of the building. Applications of this method include the verification of constructed buildings, the update or creation of civil engineering CAD models, and the acquisition of data for architectural graphics simulations.

INTRODUCTION

The perception system of most mobile robots is geared towards navigation, where qualitative information is more important than a high accuracy. We propose here a different task: the automatic metrical surveying of a building using a single camera mounted on a mobile robot. The robot provides a sequence of images registered with odometric estimates of the motion between frames. The vision algorithm reconstructs the 3-D scene by using the known motion of the camera. In order to extract the most useful features for reconstruction, we concentrate on segments that have a particular orientation in the 3-D scene. In most buildings, there are three prominent 3-D orientations: the vertical and two horizontal 3-D orientations perpendicular to each other. This assumption is often used with indoor scenes [1, 2, 7, 8, 9], and will hold for the rest of this paper.

We begin by describing algorithms to extract useful line segments from each image and to reconstruct the 3-D scene. We then discuss several issues of particular importance for practical applications in architectural surveying, such as the visibility of segments in typical indoor scenes. The following section concentrates on accuracy issues and explains the calibration procedure for the camera and the robot. Finally, we provide a comparison between reconstructed edges and the corresponding architectural floor plan.

3-D RECONSTRUCTION

Line segments corresponding to particular orientations in 3-D are extracted from each image using a special line detector based on vanishing points [4]. The detection and interpretation process provides a 3-D *orientation* hypothesis for each 2-D segment. Vertical segments are of particular importance for floor plans, but other particular 3-D orientations can be processed as well. Segments that do not have any of the predefined 3-D orientations are ignored.

The 3-D *position* of segments is estimated from a sequence of images using a method based on Kalman filtering [5, 6]. The position of each 3-D segment is represented by an estimate and an associated Gaussian uncertainty. The floor plan can then be constructed from these estimates. In our approach, each 3-D orientation is treated separately. This is possible since the line detector indicates the 3-D orientation of the segments. Matching is therefore simplified, as is the complexity of prediction and update: all the computation is done in planes perpendicular to each one of the predefined 3-D orientations.

APPLICATION TO ARCHITECTURE

The application of our reconstruction algorithms to architectural surveying imposes a few practical constraints.

Field of view: In a typical indoor scene, there are relatively few *architecturally significant* features, such as room corners and doorways. Those features are usually far apart. However, the robot should see as many features as possible at any time. For this reason, we equipped the camera with a wide-angle lens (6 mm focal distance for a 2/3 inch CCD camera). The advantage is that more interesting edges can be tracked at once. In long and narrow corridors, segments can be observed not *only from* a distance, but also as the robot passes by them. Those segments are therefore acquired under very different angles, thus reducing the uncertainty in their reconstruction. With a wide-angle lens, the forward-looking camera can still see edges lying almost along a perpendicular to the direction of motion. For those edges, the precision of reconstruction is similar to that of a stereovision system with widely spaced cameras. The accuracy is then very high, assuming that the exact motion of the camera is known.

The drawback to using a wide-angle lens is the associated barrel distortion (see Figure 1). In the next section,

we will explain how this distortion is calibrated and corrected.

Motion determination: We have assumed so far that the exact motion of the camera was given to the 3-D reconstruction algorithm. Although the robot's odometers provide a position and a heading, these measurements cannot be used alone, since odometers drift without bounds. Odometry may be adequate for estimating the motion between a few images, but it is insufficient for the long sequences necessary in mapping an entire building.

The robot is equipped with an odometer on its right and left driving wheels. In practice, translation measurements are very good because they are derived from the average of the two odometers. However, rotations (changes in robot heading) drift much faster because they rely on the difference between the two odometers. For this reason, we correct the odometric heading with vision every time an image is processed. By extracting the exact position of the vanishing points in the image, the orientation of the camera with respect to the scene is computed precisely. The heading of the robot is estimated, as well as the roll and pitch errors due to uneven floor surface. This technique prevents the heading from drifting without bounds. The position of the robot, obtained by integrating small displacements, can still drift because of errors in translations, but they build up much slower than errors in rotations.

Using this technique, the robot's positional error was reduced by several orders of magnitude after a rectangular trajectory of 125 meters around a building floor (see [5] for details).

ACCURACY AND CALIBRATION

For any practical application to architectural surveying, the precision of measurement is paramount. An accurate calibration procedure is necessary. In this section, we describe the calibration procedure for the intrinsic parameters of the camera, including the barrel distortion, as well as for the extrinsic parameters (camera-robot relationship).

The Optical axis: We first determine the intersection of the image plane and the optical axis, expressed in pixel units in the coordinate system of the frame grabber. The method we use is optical. After closing the iris of the camera, a low-power laser beam is shone through a perforated white screen into the center of the lens. The beam is partially reflected by the lens onto the screen. The goal is to align the laser beam and the optical axis of the lens. By carefully adjusting the pan and tilt of both the laser and the camera, the beam is made to reflect onto itself. At that point, an attenuating filter is placed in the beam and the iris is slightly opened. A bright spot becomes visible in the image, indicating the location of the optical axis on the retina. This method is not recommended with non-CCD cameras.

The projection model: The next calibration step consists of identifying the focal length and the distortion of the lens. The distortion is zero on the optical axis, and non-negligible near the borders of the image rectangle. As a simplification, we assume that the image is formed in two steps. Firstly, the scene is projected onto the image plane by a standard pinhole perspective projection. Secondly, the image is distorted in the image plane by a 2-D to 2-D function. This assumption means that the barrel distortion does not depend on the distance between the lens and the scene, and holds if the diameter of the lens is not too large.

The standard perspective projection is completely defined by the location of the optical axis in the image (determined in the previous subsection) and the focal length (to be determined later for points near the optical axis, where the distortion is negligible). The second step, a 2-D distortion in the image plane will then modeled by calibration. This model is used to correct the distortion in digitized images [3].

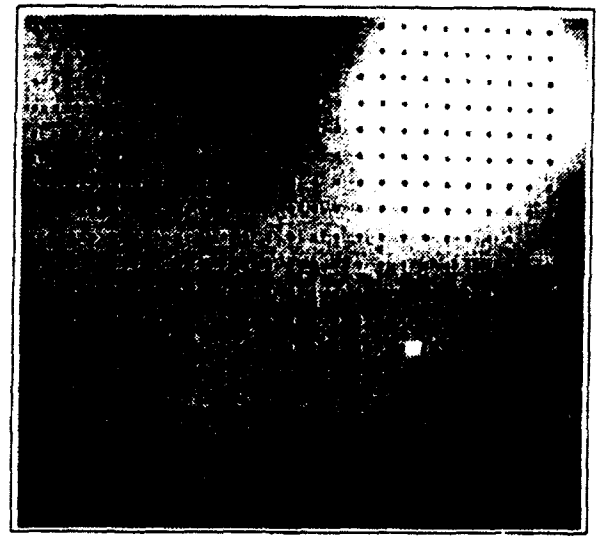


Figure 1: **The calibration pattern distorted by the wide-angle lens**

The calibration pattern: We first measure the location of a set of data points covering the image. To obtain the data points, we place the camera in front of a large calibration pattern on a wall (see Figure 1). The pattern consists of a grid of black dots placed every 10 cm. For best results, the distance between the camera and the pattern should approximate the typical distance to objects when the robot is in operation. In practice, this leads to a calibration pattern as large as the scene itself, if possible. The size of our pattern is 2.4 x 1.8 meters.

We first need to ensure that the image plane is parallel to the pattern, so that the perspective projection will not interfere with the 2-D distortion in the image plane. To achieve this constraint, we place a mirror flat on the pattern. We then adjust the pan and tilt of the camera until the reflection of the camera by the mirror coincides with the location of the optical center in the image plane. To facilitate this operation, we digitally superimpose cross hairs centered on the the optical axis over the live image from the camera. Using the cross hairs, any roll in the camera's orientation is also eliminated at this stage. Finally, the image of the pattern is digitized and the data points are automatically extracted.

The focal length: The next intrinsic parameter to be determined is the focal length, expressed in horizontal and vertical pixel units. The focal length is computed for points lying close to the center of the image, where the barrel distortion is negligible. The distance D between the camera and the calibration pattern is physically measured, and expressed in millimeters. Let L be the distance (in millimeters) between two points close to the center of

the pattern, and l the distance (in pixels) between their perspective projections. Neglecting the barrel distortion, the focal length f is given in pixels by $f = lD/L$. Since the image pixels are not necessary square, the focal length needs to be expressed both in horizontal and vertical pixel units.

Barrel distortion: At this point, all the parameters for a conventional perspective projection have been determined. However, most points of the pattern do not project exactly to their theoretical location on the image plane because of the barrel distortion. We automatically measure the resulting 2-D displacement vectors in the image plane, as the difference between the theoretical and observed coordinates of projected dots. The 2-D distortion function is estimated by linear interpolation between data points [3]. Digitized images can then be corrected to eliminate barrel distortion. The intensity of every pixel in the corrected image is found by looking up the intensity at the appropriate location in the original distorted image. In order to speed-up the distortion correction, a look-up table is computed at the time of calibration to translate corrected coordinates into raw coordinates.

Extrinsic parameters: We then measure the rotation and translation of the camera relative to the robot's coordinate system. The translations are physically measured. The roll is forced to zero by aligning cross hairs digitally superimposed on the image to the horizontal axis of the calibration pattern, as described previously. Next, the pan and tilt are adjusted until the robot is able to back away from the calibration pattern in a straight line while keeping the same point of the pattern under the cross hairs of the optical axis. This process results in zero pan and tilt. If a non-zero pan or tilt is desired, the camera is rotated from the zero-angle position by a controlled angle.

RESULTS

Figure 2 shows a short sequence of images acquired by the robot. The corresponding line segment images are given in Figure 3. The result of the reconstruction of vertical edges using these four images is presented in Figure 4, and compared to the architectural floor plan of the building. Vertical edges are indicated by small crosses. The last position of the robot is represented by a small square in the middle of the corridor. The grid has a one-meter spacing. The door in the top right corner was closed, and some segments on it are reconstructed. A few segments with no architectural interest are picked up. For example, two of them lie on the wall at the bottom of the figure. The segments in the top-left corner of the figure appear to have been shifted to the right. After measuring the actual dimensions in the scene, we found that the corner of the corridor was built approximately 5 cm to the right of the position shown on the architectural drawing. The robot therefore correctly indicated this discrepancy, although it overestimated it slightly.

CONCLUSION

We proposed a method for the task of surveying buildings using mobile robots. We addressed several important issues resulting from the particular requirements of architectural applications, and we presented results obtained in the reconstruction of indoor scenes. Finally, we compared those results to the architect's drawing of the floor plan. Improvements to our approach are possible in several areas, to further increase the precision of measurements and to produce CAD-like building maps. We believe that automatic architectural surveying using mobile robots is a feasible task in the fairly near future. Applications will include the verification of constructed buildings, the update or creation of civil engineering CAD models, the acquisition of data for architectural graphics simulations, and the creation of maps for use by other robots.

Acknowledgments

This research was supported in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-89-C-0044, and in part by the Army Research Office under contract DAAL03-91-G-0050.

References

[1] S. T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4):435–462, November 1983.

[2] P. Belluta, G. Collini, A. Verri, and V. Torre. Navigation by tracking vanishing points. In *Working Notes of AAAI Robot Navigation Symposium*, pages 6–10, Stanford University, March 1989.

[3] D. A. Butler and P. K. Pierson. A distortion-correction scheme for industrial machine-vision applications. *IEEE Trans. on Robotics and Automation*, 7(4):546–551, August 1991.

[4] X. Lebègue and J. K. Aggarwal. Detecting 3-D parallel lines for perceptual organization. In *Proc. Second European Conf. on Computer Vision*, pages 720–724, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[5] X. Lebègue and J. K. Aggarwal. Extraction and interpretation of semantically significant line segments for a mobile robot. In *Proc. IEEE Int. Conf. Robotics and Automation*, pages 1778–1785, Nice, France, May 1992.

[6] X. Lebègue and J. K. Aggarwal. Semantically significant line segments for a mobile robot... ROBOTEX. In *Video Proc. IEEE Int. Conf. Robotics and Automation*, Nice, France, May 1992.

[7] P. Olivieri, M. Gatti, M. Straforini, and V. Torre. A method for the 3D reconstruction of indoor scenes from monocular images. In *Proc. Second European Conf. on Computer Vision*, pages 696–700, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[8] T. Shakunaga. 3-D corridor scene modeling from a single view under natural lighting conditions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):293–298, February 1992.

[9] M. Straforini, C. Coelho, M. Campani, and V. Torre. The recovery and understanding of a line drawing from indoor scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(2):298–303, February 1992.

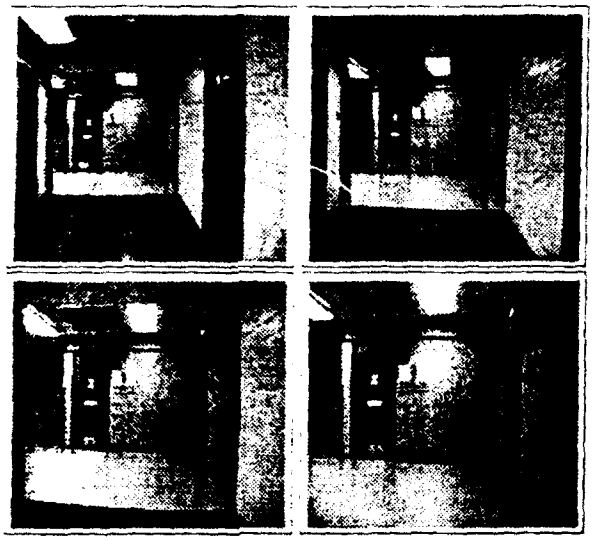


Figure 2: The four images used in the reconstruction

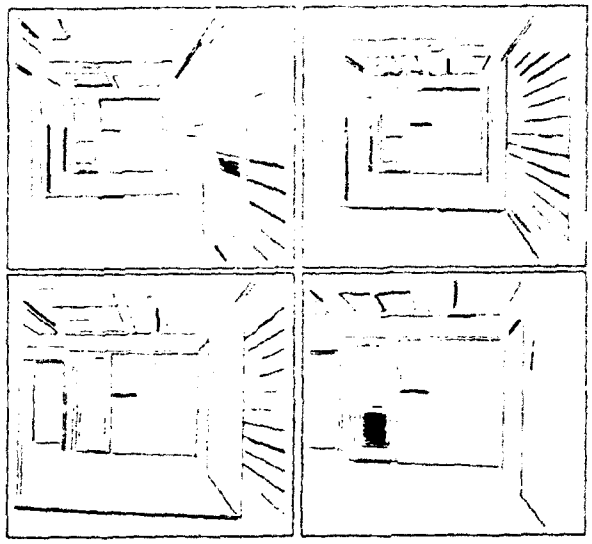


Figure 3: The corresponding line segment images

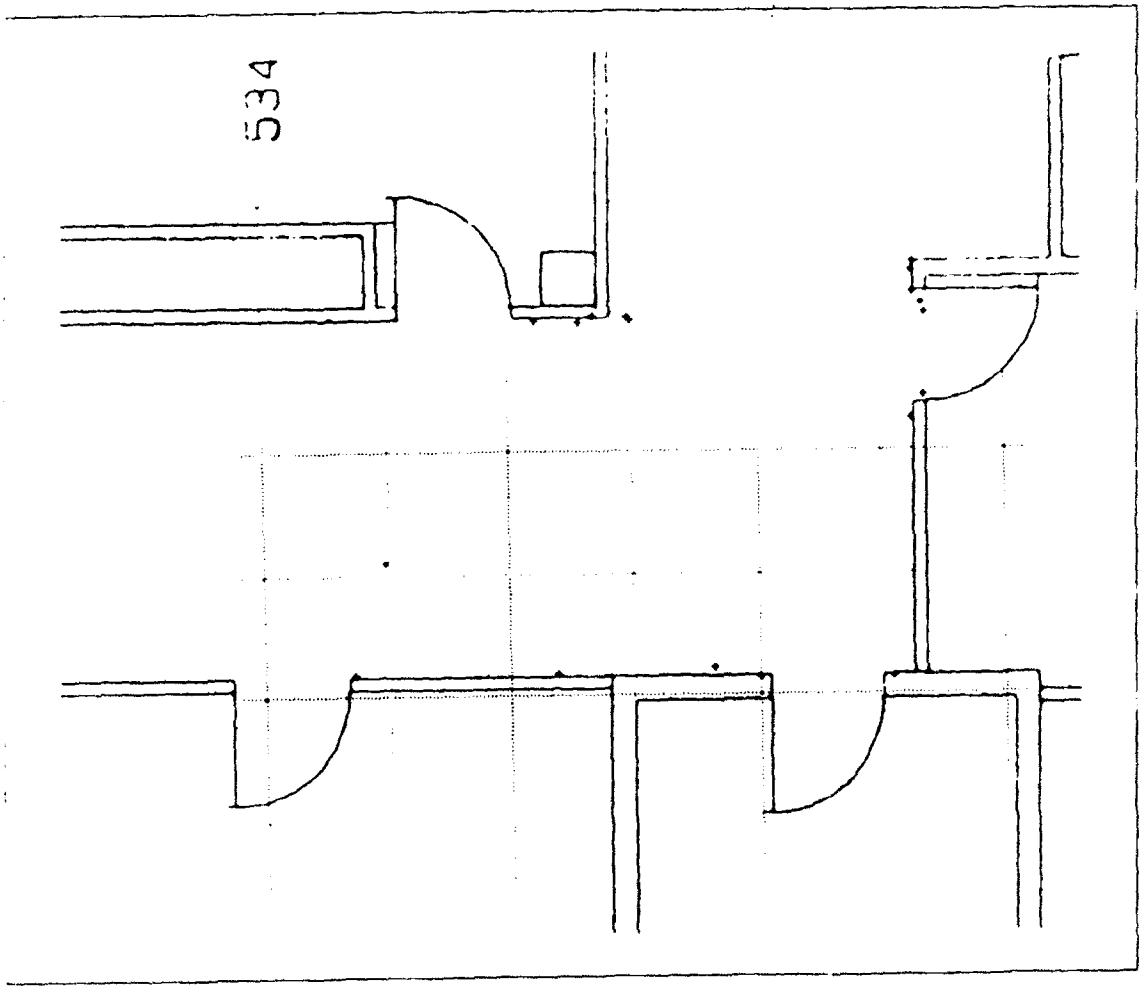


Figure 4: The reconstructed vertical edges overlaid on the floor plan

MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

CORRESPONDENCE OF SURFACES IN A SEQUENCE OF RANGE IMAGES FOR MOTION ESTIMATION AND TRACKING

Bikash Sabata and J. K. Aggarwal

Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712, USA

ABSTRACT

The key issue in motion estimation and tracking an object over a sequence of images is establishing correspondence between the features of the object in the different images of the sequence. For range image sequences, this problem translates into finding a match between the surface segments in a pair of range images of the scene. This paper considers the problem of establishing correspondences between surfaces in a sequence of range images. We present a novel procedure for finding correspondence and show the results on real range image sequences. A graph search procedure forms the basis for the algorithm that computes the correspondence between surfaces. The solution uses geometrical and topological information derived from the scenes to direct the search procedure. Fundamental to our strategy to match features over a sequence of range images is a hypergraph representation of the scenes. Two scenes are modeled as hypergraphs and the hyperedges are matched using a sub-graph isomorphism algorithm. The hierarchical representation of hypergraphs not only reduces the search space significantly, but also facilitates the encoding of the topological and geometrical information. Further, we present a sub-hypergraph isomorphism procedure to establish the correspondences between the surface patches and demonstrate the algorithm on different types of real range image sequences. We present results that show that the algorithm is robust and performs well in presence of occlusions and incorrect segmentations.

INTRODUCTION

The key issue in motion estimation and tracking an object over a sequence of images is establishing correspondence between the features of the object in the different images of the sequence. In this paper we deal with the tracking of objects in a sequence of range images to estimate the motion of the camera (range sensor) in the environment. Range images sense the surface of the objects, so it is natural to use surface segments as the features of interest; this translates the tracking of objects into finding a match between the surface segments in a pair of range images of the scene. This paper considers the finding of correspondences between surfaces in a sequence of range images. Finding correspondence or a match between features is not isolated to object tracking, but is also central to other computer vision tasks including navigation, object recognition, target tracking, and map building. We present a novel procedure for establishing correspondence and show the results on real range image sequences.

A graph search procedure forms the basis for the algorithm that computes the correspondence between surfaces. The solution uses geometrical and topological information derived from the scenes to direct the search procedure. In general, the input to the matching algorithm is the output from a segmentation algorithm that partitions the image into surface segments. The performance of the matching depends greatly on the results of the segmentation algorithms. Incorrect segmentation causes poor estimation of the surface parameters and affects the performance of the matching algorithm. We address this issue and obtain a solution that is robust and able to handle occlusions of surfaces, noise in data, and incorrect segmentation from a segmentation algorithm. In the present implementation, we assume that the images have planar, cylindrical and conical surfaces; however, the procedure is general enough to be extended to other surface classes.

The question of finding correspondences between features has been studied extensively (see [1, 3, 4, 5]) but, most of these approaches deal with matching a scene to a model of the object. The fundamental difference between model-to-scene matching and scene-to-scene matching is that in the former, the model description of the object is complete, and to that we match the incomplete description of the object obtained from the scene. However, in the case of scene-to-scene matching, both descriptions of the object are incomplete and we must find a match between two incomplete descriptions. By incomplete, we mean that all the features are not present in the description of the object because of occlusions and sensor errors. This difference makes it impossible to use the strategies obtained for object recognition in the domain of object tracking; new strategies based on the constraints of the problem have to be designed.

Fundamental to our strategy to match features over a sequence of range images is a hypergraph representation of the scenes. The two scenes are modeled as hypergraphs and the hyperedges are matched using a sub-graph isomorphism algorithm. To reduce the complexity of the matching task, heuristics derived from the topological and the geometrical information available from the scene are used to direct the search. The hierarchical representation of hypergraphs not only reduces the search space significantly, but also facilitates the encoding of the topological and geometrical information. Hyperedges are formed by grouping the surface features, which reduces the search space. Using a priori knowledge arising out of the physical constraints of laser scanning, a fast matching algorithm is designed.

HYPERGRAPH REPRESENTATION

Hypergraphs are generalizations of graphs. The edge is generalized as a hyperedge, where a set of vertices forms the hyperedge, instead of just two vertices forming the edge. The group of vertices forming the hyperedge may share some common property. Hypergraphs have been used earlier in vision and robotics applications [11, 12], but have not found widespread usefulness. We present a new definition of the hyperedge and a novel method for constructing the hypergraphs that makes it a powerful tool for vision applications.

Attributed hypergraphs are a concise way of representing objects such that both quantitative and qualitative information are encoded in the representation. Formally;

Definition 1 *The Hypergraph [2] is defined as an ordered pair $H = (X, E)$ where $X = \{x_1, x_2, \cdots, x_n\}$ is a finite set of attributed vertices and $E = \{e_1, e_2, \cdots, e_m\}$ are the hyperedges of the hypergraph. The set E is a family of subsets of X (i.e. each e_i is a subset of X) such that*

1. $e_i \neq \emptyset, i = 1, \cdots, m$
2. $\bigcup_{i=1}^{m} e_i = X$.

A graph is a hypergraph whose hyperedges have cardinality of two. To each hyperedge, we associate an attribute set that maps the vertices (belonging to the hyperedge) to an attributed graph.

Each surface patch in the range image forms an attributed vertex. The attribute values are the surface property values. For each pair of surfaces that are connected, an attributed arc is formed. The attributes of the arc describe the interfacing edge and the relative geometrical information between the two surfaces. Groups of the attributed vertices (surface patches) form an hyperedge, and with each hyperedge we associate an attributed graph that describes the topology of the component attributed vertices (surface patches).

The set of vertices that form the hyperedge should represent a topologically significant feature in the graph so that the matching task is guided by the topology of the scene. Cliques in the graph are significant features that are rich in information. Physically, the cliques represent groups of surfaces that are adjacent to each other. Since a clique provides a larger attribute set and many geometrical properties, the probability of a false positive match (between two cliques) is reduced significantly. Each clique forms a hyperedge in the hypergraph and the attributed graph describing the clique is the associated attribute of the hyperedge. Figure 1 illustrates the formation of a hypergraph from a scene.

The complexity of computing the cliques in a graph is exponential, so the formation the hypergraph will be exponential. However, the physics of the range imaging process restricts the size of the cliques in the scenes that we observe. It can be shown that the size of the clique is restricted to be four [8]. Once the upper bound on the size of the cliques is known, the complexity of computing the cliques becomes O(n).

THE MATCHING PROCEDURE

This section presents the matching procedure used to derive the surface correspondences in a sequence of range images. The heart of the procedure is a *directed tree search* algorithm that tests various hypotheses and rejects the impossible ones. Finally the interpretation that gives the largest match is selected as the solution. Constrained tree search algorithms have been used in many applications [4, 5, 7]. Data pairings are formed by a depth first search of an *interpretation tree*. Each node of the tree represents a possible pairing. The first data (surface patch) is taken from the first scene and paired with each of the data in the second scene. These form the nodes in the first level of the tree. To account for missing surface segments due to occlusions, the data is also paired with a *wild card* $\star$. Subsequent levels of the tree correspond to pairings of other vertices. Each branch of the tree represents a partial matching of the scenes. The constraints are used to prune the search tree and thus reduce the search space.

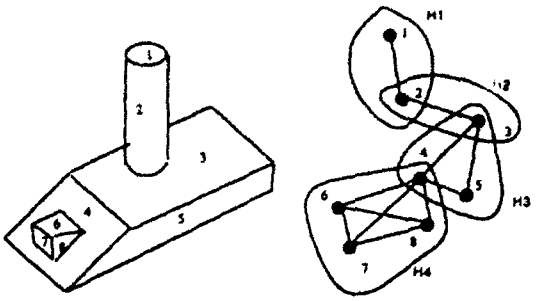


Figure 1: *An object and its corresponding Hypergraph representation*

We present a variation to the constrained tree search, in which the search is directed based on the current hypothesis. The directed search, coupled with the termination conditions, further reduces the search space. The key idea is to use the topological constraints of the scene to determine the next most likely match, and to accept or reject the matches based on the geometrical constraints.

The features used in the matching process are surface segments. We assume that a segmentation algorithm [9, 10] segments the range image into surface patches and the surface parameters are computed. The interfacing edge between the surface patches are detected and their properties are computed. The properties of the edge segments used are (1) the edge type (straight line or curved), (2) the edge length, and (3) the depth discontinuity. The depth discontinuity across the edge implies that one surface may be occluding (partially or completely) another surface. The information about occlusion is also incorporated in the attribute list of the surface patches.

The constraints used are similar to the *unary* and *binary* constraints developed by Grimson and Lozano-Perez [6]. The only unary constraint we use is the surface type classification (planar, cylindrical, conical, etc.). Other properties, used in model based object recognition, such as area, perimeter, compactness, etc., are very sensitive to occlusion, and since occlusion may occur in either of the range images, these properties cannot be used as constraints. The binary constraints describe the relative properties between pairs of surface segments. The properties we use are (1) connectivity, (2) the angle between the surface patches, (3) the range of distances between the two surface patches, (4) the range of the components of the vector spanning the two surface patches, and (5) the properties of the interfacing edge. Each constraint is measured and tested against a predetermined threshold. For surface

segments that have an occluding edge the neighbors information is not complete (a neighbor may be hidden) and the connectivity information may be inaccurate. Therefore, for such cases only a *weak* match is hypothesized which is subject to conformation or rejection based on further evidence.

Matching between the two hypergraphs representing the scenes is achieved by computing the match between the component hyperedges. A match between the two hyperedges is hypothesized. The two hyperedges are matched by matching the attributed graphs representing the hyperedges. An order of vertices is established at each stage of the match. The order determines the branches taken in the search tree. The order is determined by listing the hyperedges connected to the vertices that have been matched in current hypothesis. The matching procedure starts by selecting the largest hyperedge H_1 and H_1' in the two scenes. The vertex with the largest degree is selected as the first node n_1 and it is matched with the corresponding vertex in the second hyperedge. The unary and the binary constraints are checked to evaluate the match between the hyperedges. Once the hyperedge-match has been established, the second set of hyperedges are selected. The next hyperedge H_2 is the hyperedge connected to H_1 at n_1. A match for each of the hyperedges connected to H_1 at n_1 is found. The search then proceeds to find matches of hyperedges connected to H_1 at other vertices belonging to H_1. The procedure goes down the list of all the vertices in the hypergraphs in the order evaluated earlier. Once a match for a hyperedge is found that hyperedge is marked as *matched*. The marked hyperedges are not considered in the future hypotheses.



Figure 2: *The depth maps of a sequence of range images.*

Termination of the matching procedure occurs if the fraction of surface patches matched exceeds a threshold. Once a match has been determined (i.e., the search procedure has reached the leaf node of the tree), the number of positive pairings (i.e., non-wild card pairings) is computed. If this number is less than the threshold fraction then the procedure backtracks and searches other branches. At every stage the best possible match is compared with the current best match. If the best possible match is smaller than the current match, then the search along that branch is abandoned and the next branch is investigated.

RESULTS

In this section we present an example of a range image sequence and describe how the matching algorithm computes the surface correspondences. The algorithm was tried successfully on different types of range image sequences.

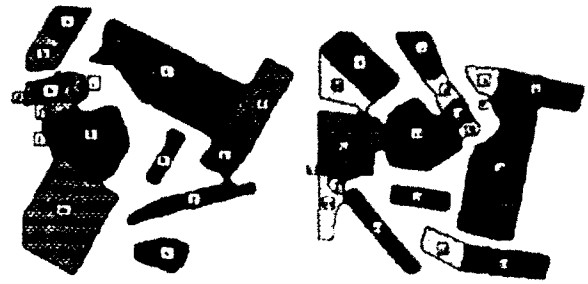


Figure 3: *The segmented range images*

Figures 2-4 illustrate the algorithm on an example. Figure 2 shows the depth maps of two frames in the sequence of range images. The scenes consist of a jumble of different kinds of objects. The camera is moved to obtain the second frame of the sequence. The segmentation algorithm of [10] was applied on the images and the results input to the matching algorithm. The segmented results are shown in figure 3. The first step of the algorithm generates the attributed graph of the scene and computes the cliques in the graph. Each clique forms a hyperedge in the generated hypergraph. The hypergraphs generated are shown in the figure 4. For each hyperedge the component vertices form an attributed graph. In the figure the arcs of the attributed graph are shown in the hyperedges. Using the properties of the edge, interfacing two surface segments, it is determined if two surfaces are connected. If there exists an occluding edge between two surfaces then the arc in the attributed graph is *weak* (shown in the figure 4 with dotted lines). A match based on a *weak* arc is a *weak* match and further evidence is required to confirm the hypothesis.

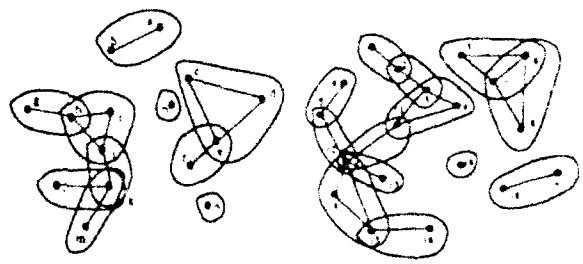


Figure 4: *The generated hypergraphs of the range images.*

The first hyperedge pair hypothesized to match is $\{h, i, j\}$ in the first scene matches $\{3, 4, 11\}$. The vertex with the highest degree h is considered as the first vertex. The unary constraints lease only one option i.e., $(h, 3)$ as the first node in the interpretation tree. However, the next two vertices i and j do not match any vertex so they are matched with the wild card $*$. Note that in the final match that is obtained the pairing $(h, 3)$ is an incorrect pairing. The algorithm backtracks and finds the correct match even though we start with an incorrect match. We present the first few steps to illustrate how the algorithm works. The second hyperedge considered for match is $\{h, g\}$ because it is connected to the first hyperedge at h. Since the current hypothesis is $(h, 3)$, the next hyperedge match considered is between $\{h, g\}$ and $\{3, 2\}$. The unary constraints are satisfied between the pair $(g, 2)$ so the binary constraints of angle, distance and the spanning vector are tested. All the constraints are satisfied so the match pair is accepted.

in the current hypothesis. The next hyperedge considered now is $\{j,k\}$ as it is connected to the first hyperedge. The match between $\{j,k\}$ and $\{11,12\}$ is tried and the pairing $(k,11)$ satisfies all the constraints, but the connectivity is not satisfied (k is not connected to h while 11 is connected to 3). At this point we use the fact that the arc between 3 and 11 is a *weak* one so it can be broken and all the constraints are satisfied.

The procedure continues till a complete match (i.e., all the vertices are accounted for) is obtained. The match size is evaluated and if a better match can be obtained, the procedure backtracks to improve the results. The final matching results are:

I	a	b	c	d	e	f	g	h	i	j	k
II	6	5	8	17	18	16	1	2	3	*	11

I	l	m	n	o	*	*	*	*	*	*	
II	*	12	19	15	4	7	9	10	13	14	

It may be observed that in the example shown there are many errors in segmentation (for eg. surfaces j, 7, 14, 13, etc.) and there are surfaces that get occluded in one of the scenes (for eg. 4 and l); notwithstanding, the algorithm performs well and the correspondences are evaluated.

CONCLUSION

Computing motion and tracking an object over a sequence of range images involves establishing correspondence between the features of the object in different images in the sequence. The question of finding correspondence in a sequence of range images is very different from finding correspondence between a model and an object description. The fundamental difference lies in the fact that the model description of the object is complete, while in case of a sequence of range images, both descriptions of the scene are incomplete. The lack of information forces one to impose only weak constraints and allow for larger tolerances.

We presented a new framework and procedure to compute the correspondences between surface segments in a sequence of range images. Fundamental to our framework is the hypergraph representation of the range images. The hierarchical representation of hypergraphs not only reduces the search space significantly, but also facilitates the encoding of the topological and geometrical information. In addition to the topological and geometrical information obtained from the scene we also use a priori knowledge of the scene obtained from the physics of the laser scanning process used to produce the range images. Each piece of information used reduces the complexity of the matching procedure by pruning the search space. The solution is robust and accounts for errors in segmentation, occlusions of surfaces, and noise in the data. By using the topological information to guide the search procedure, the average case complexity of the algorithm is reduced significantly.

Acknowledgments

We would like to thank Debi Paxton for proof reading the paper. This research was supported in part by Army Research Office, Contract *DAAL 03-91-G 0050* and Air Force Office of Sponsored Research, Contract *F49620-92-C-0027*.

References

[1] F. Arman and J.K. Aggarwal. Model-based object recognition in dense range images - a review. *To Appear in ACM Computing Surveys*, 1992.

[2] C. Berge. *Graphs and Hypergraphs*. North-Holland Publishing Company, Amsterdam, 1983.

[3] W.E.L. Grimson. The combinatorics of object recognition in cluttered environments using constrained search. *Artificial Intelligence*, 44:121-165, 1990.

[4] W.E.L. Grimson. The combinatorics of heuristic search termination for object recognition in cluttered environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):920-935, Sept. 1991.

[5] W.E.L. Grimson and D.P. Huttenlocher. On the verification of hypothesized matches in model-based recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1201-1213, Dec. 1991.

[6] W.E.L. Grimson and T. Lozano-Perez. Localizing overlapping parts by searching the interpretation tree. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):469-482, 1984.

[7] V. Kumar. Algorithms for constraint satisfaction problems: A survey. *AI Magazine*, pages 32 - 44, Spring 1992.

[8] B. Sabata and J.K. Aggarwal. Correspondence of surfaces in a sequence of range images using hypergraphs. Submitted for Publication.

[9] B. Sabata, F. Arman, and J. K. Aggarwal. Segmentation of 3-d range images using pyramid data structures. Accepted for publication in Computer Vision, Graphics, and Image Processing: Image Understanding.

[10] B. Sabata, F. Arman, and J. K. Aggarwal. Segmentation of range images using pyramidal data structures. In *Proceedings of International Conference in Computer Vision*, pages 662-666, Osaka, Japan, December 1990.

[11] A.K.C Wong. Knowledge representation for robot vision and path planning using attributed graphs and hypergraphs. In A.K.C. Wong and A. Pugh, editors, *Machine intelligence and knowledge engineering for robotic applications*, pages 113-143. Springer Verlag, Berlin Heidelberg, 1987.

[12] A.K.C. Wong, S.W. Lu, and M. Rioux. Recognition and shape synthesis of 3-d objects based on attributed hypergraphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(3):279 - 290, March 1989.

TRANSFORM CLUSTERING FOR MODEL-IMAGE FEATURE CORRESPONDENCE

Raj Talluri and J. K. Aggarwal

Computer and Vision Research Center
The University of Texas at Austin
Austin, Texas 78712, USA

ABSTRACT

In this paper we present a novel technique for establishing a robust and accurate correspondence between a 3d model and a 2d image. We present a transform clustering approach to isolate the transformation that maps the model features to the image features. It is shown that this transform clustering technique alleviates the problems with using the traditional Hough transform techniques used by previous researchers. We demonstrate the effectiveness of our approach in estimating the position and and pose of an autonomous mobile robot navigating in an outdoor urban environment. We present experimental results of testing this approach using a model of an airport scene.

INTRODUCTION

The task of establishing a reliable and accurate correspondence between an image of a scene and a stored model of it occurs in a large number of computer vision problems. Autonomous navigation of a mobile robot given *a priori* model of the environment and model-based object recognition are two examples of computer vision tasks in which the model-image correspondence needs to be addressed. In the context of autonomous navigation, the robot is provided with a preloaded world model of the environment. The world model could be in different forms, such as a Digital Elevation Map (DEM), a CAD description, or a floor map. The robot uses an onboard camera to image the environment. Once we establish a correspondence between the image and the model, the robot's position and pose can be determined. This position information can be used by the robot to successfully navigate in its environment. In the context of model-based object recognition, we are given a geometric description of the object to be recognized and an image of the scene in which the object is present. The task is to isolate the object in the scene by using the image. Model-image correspondence are particularly difficult because the image and the model are usually in different formats, different co-ordinate frames and of different dimensions.

A popular approach to solving this problem is to extract features from the image and search the model description for the corresponding set of features. The type of features required and the number of features used depends on the model description and what is assumed to be known about the scene. For example, in navigating the robot in an indoor structured environment with a given CAD model of the environment, it is common practice to use line segments as features [3]. On the other hand, in navigating the robot in an outdoor mountainous terrain given a DEM of the environment, using curves may be a logical choice [9].

Typically, in these problems the model and the camera (robot) are specified in two different co-ordinate systems. Once we extract the relevant features from the image and identify the corresponding features in the model, we can computer the transformation $\mathcal{T}$ that maps the model features into the image features. The parameters of this transformation are the required position and pose of the camera (robot) with respect to the model. Solving for the parameters of $\mathcal{T}$, once a set of model-image feature correspondences is established, is a very well studied problem [2]. Therefore, the crucial task to be accomplished is that of establishing a reliable and accurate correspondence. Noise, occlusions, errors in feature detection and inaccurate model descriptions further complicate this correspondence problem.

Transform Clustering: Previous researchers have considered the technique of matching a key model feature, such as a long edge or a set of lines in specific orientations, to establish an initial transformation [1, 6]. Subsequent assignments are then used to refine this transformation. New assignments are selected on predictions of a model feature, projected into the image using the current transformation. However, these techniques assume that it is possible to initially select a correct key model feature, which may not always be possible.

Some researchers used the generalized Hough transform and its related parameter hashing techniques to perform *transform clustering* to isolate the transformation mapping the model features into the image features [6, 5, 10]. The generalized Hough transform works by first quantizing the n-dimensional parameter space into discrete buckets or bins. The parameters are the components describing $\mathcal{T}$. From the given image, features are extracted using a feature extractor. Then all the possible model-image feature correspondences are hypothesized and, for each hypothesis, the parameter vector is computed. For each parameter vector so computed, its n components are quantized and used as indices to vote in one of the n-dimensional buckets. Searching for large clusters is then accomplished by finding the buckets with a large numbers of entries. Sometimes it is possible that one correspondence may not give explicitly all the components of the parameter vector, but may only give a range of possible values for each component. In this case, entries are

This research was supported by in part by Army Research Office contract DAAL03-91-G-0050 and in part by Air Force Office of Scientific Research (AFSC) contract F49620-89-C-0044.

made into all the buckets within range. The advantage of this approach is that clustering provides a robust criterion for selecting valid model feature assignments. The effects of missing or incorrect features due to occlusion, shadows, or low contrast, are not felt.

The problems associated with using the Hough transform approach to transform clustering are that large transform clusters may occur randomly. If these clusters are as large or larger than those due to the correct transform, the estimation procedure that relies only on the Hough transform will be erroneous. If the number of buckets is increased, then the possibility of random large clusters is alleviated but the number of computations grows rapidly. Grimson [4] summarizes these problems with the generalized Hough transform.

This paper presents a method to reduce the problems associated with the Hough transform approach to transform clustering by using a partition of the parameter space, which is not necessarily uniform. The partition is, in fact, *intelligent* and uses *a priori* model information. Due to the geometric constraints imposed by the model and the camera geometry, not all model features may be *visible* in all camera positions. Typically occlusions between the model features affect their visibility at various positions. However, since we know the 3d descriptions of the model features, these geometric constraints can be pre-computed and used to partition the parameter space to reduce the probability of the occurrence of random transform clusters.

We demonstrate the effectiveness of our approach in estimating the position and and pose of an autonomous mobile robot. The robot is assumed to be navigating in an outdoor, urban environment. The 3d description of the lines that constitute the rooftops of the buildings is given as a world model. The position and pose of the robot are estimated by establishing a correspondence between the lines extracted from the image (image features) and the lines that constitute rooftops of the buildings (model features). By exploiting the visibility constraints imposed by the 3d world model and the camera geometry, we partition the parameter space into into distinct, non-overlapping regions called *Edge Visibility Regions* (EVRs) [7]. In each of these regions, we also store the list of model features that are visible from within that region. We then hypothesize a correspondence between all pairs of model and image features and compute the range of possible transformations for each hypothesis. We vote in all the regions in the parameter space where this transformation is valid. After considering all the pairings, we select the regions in the parameter space with the large numbers of votes as the candidate EVRs for position estimation. The actual correspondence and position estimation are then performed by a constrained search process within these EVRs using a interpretation tree search paradigm.

PARTITIONING THE PARAMETER SPACE

Consider the world coordinate system $OXYZ$ and the robot coordinate system $O'X'Y'Z'$ shown in Figure 1. Generally, the transformation T that transforms $OXYZ$ into $O'X'Y'Z'$ has six degrees of freedom: three rotational and three translational. Sometimes, depending on the application, some of these degrees of freedom can be elimi-

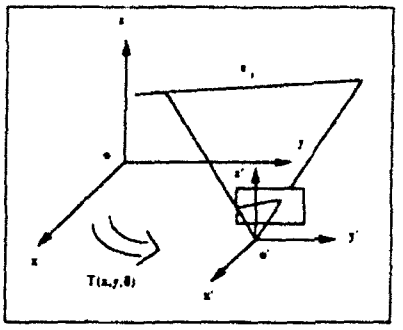


Figure 1: The world and robot co-ordinate systems

nated. Most mobile robot self-location tasks make the assumption that the robot is on the ground (OXY plane), so the Z-translation (the height of the robot above the ground) is assumed to be known or to be zero. The camera on the robot is assumed to have zero roll (rotation about X-axis), and the tilt angle of the camera, (rotation about the Y-axis) is assumed to be measurable. So, there are effectively three parameters in the transformation: two translational (X, Y) and one rotational θ (the pan angle of the camera, which is a rotation about the Z-axis). Likewise, in this paper we have only three parameters of T: X, Y and θ. The parameter space of the transformation is thus the entire OXY plane and the range of robot orientation θ is 0 through 360 degrees.

In this section, we briefly we describe a method for partitioning the OXY plane into regions called Edge Visibility Regions (EVRs) using the given world model description. For more details see [7]. Associated with each EVR is a list of the world model features *visible* in that region, called the *visibility list* (VL). No two adjacent EVRs have the same VL. Also stored for each entry in the VL of an EVR is the range of robot orientations from which the feature is visible. Thus, each EVR is a region of space which has the topological property that from its points, the same set of edges of the model are visible through a complete circular scan. The EVR representation partitions the entire parameter space of (X,Y,θ) and captures the visibility constraints between the world model features.

The algorithm that divides the OXY plane into the desired EVRs, along with their associated VLs, uses three subprocesses called *Split*, *Project*, and *Merge*. The algorithm's basic idea is to start with the entire OXY plane as one EVR with a NULL visibility list. Each of the polygon that makes up the building's rooftop in the world model is considered in turn by extending its edges, and the EVRs that are intersected are divided into two new ones. The new EVRs then replace the old one, and the VLs of the new EVRs are updated to account for the visibility of this edge by considering it to be visible in one half-plane, say the half-plane into the left of the edge, and invisible in the other. The *Split* process handles this updating. For each new rooftop considered, the mutual occlusion of the rooftop's edges with the other existing rooftops is handled by forming the *shadow region* of these edges on the other existing rooftops. The *Project* process handles the forming of these shadow regions. Finally, the *Merge* process concatenates all the adjacent EVRs with identical VLs into one EVR. After partitioning the OXY plane into EVRs,

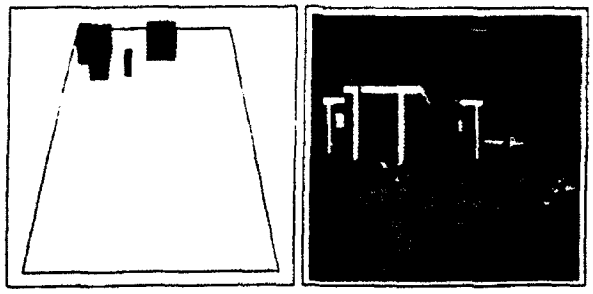


Figure 2: (a) World model (b) Robot view

the range of the robot's orientations for which each model feature in the VL of an EVR is visible, is also computed and stored. An efficient method to compute these ranges is also developed. Figure 2(a) shows the world model and Figure 5(b) shows the EVR description computed from this world model.

FEATURE EXTRACTION

In this research, we used a scale model of the Austin Executive Park Airport to test the position estimation algorithms developed. The world model thus consists of the 3d descriptions of the *rooftops* of the three buildings in this airport. Figure 2(a) shows this world model. A calibrated camera is placed in this environment and used to acquire the images of the model. These are then used as the robot's views. Figure 2(b) shows one such view. We use a Canny edge detector to extract the edges from this image. Contiguous edges are then linked using a pixel chaining algorithm. We then use a line fitting technique to form line segments from these pixel chains. These line segments are then thresholded by length to remove all the lines shorter than 20 pixels. Figure 3(a) shows these lines. We use a *rooftop extraction* technique to select the lines that correspond to the rooftops only. The technique scans each column of the line segment image from top to bottom and selects the *topmost* lines only in each column. All the lines that lie below, completely within the projection of a selected line, are then discarded. The lines isolated using this technique are then considered as the image features. Figure 3(b) shows these lines. Notice that the image feature extraction procedure is far from perfect. Some of the lines that correspond to rooftops are not extracted and, due to noise and occlusion, some of the extracted lines *do not arise from the rooftops* but *from* extraneous objects such as trees and telephone poles. The task is thus to use the transform clustering and the search technique to correctly isolate the model features and the noise features from these image features and accurately estimate the robot's position and pose in the environment.

MODIFIED HOUGH TRANSFORM

Having formed the EVR description of the environment and extracted the features from the images, we use a modified Hough transform to isolate a small set of EVRs likely to contain the robot's location. The EVRs are used as a partitioning of the parameter space (X, Y, θ) of the transformation. We find that this partitioning alleviates the problems of traditional Hough transform, namely, the random occurrence of large clusters and the resulting need for the large amounts of memory required to perform the

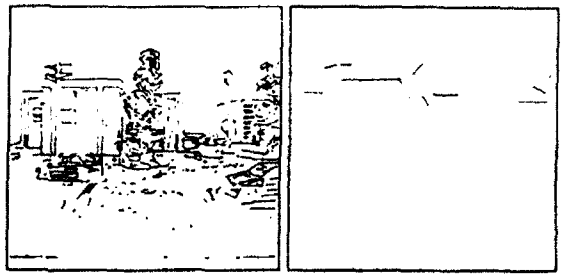


Figure 3: (a) Detected lines (b) Image features

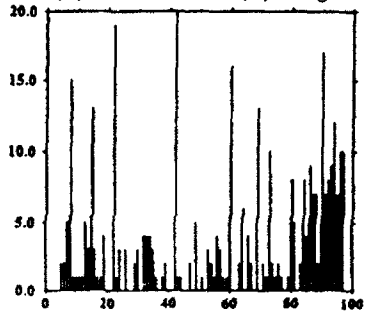


Figure 4: EVR no. vs the number of votes

fine partitioning of the parameter space to eliminate this problem. Since it is difficult to accurately extract the end points of the rooftops, we use infinite lines and not line segments as the image features. The image features are 2d lines and the model features are 3d lines. Using one 2d to 3d line correspondence, we can compute the orientation of the robot θ and get a constraint on the position of the robot of the form aX + bY + c = 0, where a,b, and c are constraints. This constraint describes a line L in the OXY plane. See [8] for details of the derivation.

We hypothesize all the possible model-image feature correspondences, and for each hypothesis compute the θ and get the constraint line L on (X, Y). We now vote in all the EVRs where: 1) the line L intersects the EVR; and 2) the θ lies within the range of possible robot orientations in the visibility list of the EVR. We finally select the EVR with a largest numbers of votes as the candidate EVRs most likely to contain the robot's location. Figure 4 shows a plot of the EVR number vs. the number of votes. Figure 5(a) shows the complete EVR description and Figure 5(b) shows the selected candidate with a large number of votes.

INTERPRETATION TREE SEARCH

Having isolated the candidate set of EVRs most likely to

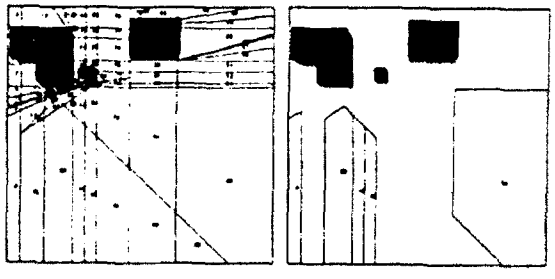


Figure 5: (a) EVR description (b) EVRs isolated by the Hough transform

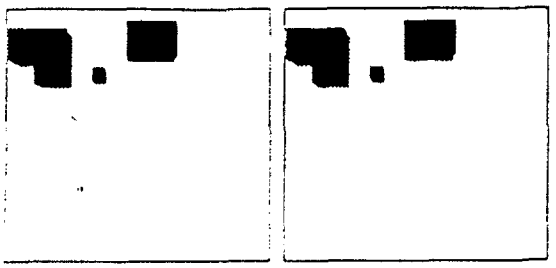


Figure 6: (a) Final EVR (b) Estimated robot location

Actual Position	Actual Pose deg	Estimated Position	Estimated Pose deg	EVR No.
(875,410)	0	874.93,411.29	1.03	42
(750,1010)	-30	752.45,1016.21	-32.02	90
(900,50)	5	901.74,49.84	4.87	8
(800,800)	-5	798 63,800.12	-4.58	84
(1320,1350)	-15	1324.27,1347.46	-16.14	89
(350,1400)	-90	352.23,1404.67	-92.27	94
(460,1400)	-75	460.15,1406.14	-78.04	91
(425,1450)	-85	428.57,1454.21	-88.13	93
(1350,900)	-15	1351.12,896.14	-14.21	83
(1475,600)	0	1480.11,605.26	1.02	69

Table 1: The Search results

contain the robot's location using the modified Hough transform, we now wish to isolate the robot's location more precisely among these EVRs. For each of the candidate EVRs we form an *interpretation tree* of all the possible model-image feature correspondences and then search this trees to isolate the correct set of correspondences. Note that these tree are very short since we only need to consider those model features that are present in each EVR's VL. Also by using the geometric constraints established by the EVR, that is, its extent in the OXY plane and the range of possible θ values, we can prune large parts of this interpretation tree.

This search process finally isolates the correct EVR containing the robot's location and a set of model-image feature correspondences. Using all of these correspondences in a least squares framework, the robot's position and pose are accurately estimated. Figure 6(a) shows the EVR isolated as containing the robot's location and Figure 6(b) shows the final estimated robot's position. We find that the estimated position and pose obtained by these techniques are quite close to their true values. Table 1 compares the estimated and the actual values obtained from the test runs using the world model shown in Figure 2(a).

CONCLUSIONS

This paper presented a novel and efficient transform clustering technique for establishing a robust and accurate correspondence between a 3d model and a 2d image. We demonstrate the effectiveness of this technique in estimating the position and pose of an autonomous mobile robot in an outdoor urban environment consisting of polyhedral buildings. It is shown that this transform clustering technique alleviates the problems associated with the traditional Hough transform techniques used by previous researchers.

Although we have demonstrated the utility of the technique for the mobile robot self-location problem, the approach can be easily extended to other computer vision tasks such as model-based object recognition. One possible approach is to precompute the *characteristic views* or *aspects* of the object to be recognized and use these to partition the parameter space. By imposing suitable and practical restrictions on the number of degrees of freedom in the transformation between the model and the imag [6], the number of aspects can be kept tractable. By selecting an appropriate set of features from the image and using a similar transform clustering approach as described in this paper, it is possible to isolate a small set of aspects of the object corresponding to the given image. Using a tree search technique it is then possible to establish a more accurate correspondence between the image features and the model features and isolate the correct aspect, and thereby recognize the object from the given set of models.

References

[1] N. Ayache and O. Faugeras. HYPER: a new aproach for the recognition and positioning of 2d objects. *IEEE Trans. on PAMI*, 8(1):44-54, 1986.

[2] R. M. Haralick et. al. Pose estimation from corresponding point data. *IEEE Tran. on Sys., man and cybernetics*, 19(6):1426-1445, November/December 1989.

[3] C. Fennema and A. R. Hanson. Experiments in autonomous navigation. In *Proc. of the tenth Int. Conf. on Pattern Recognition*, pages 24-31, Atlantic City, 1990.

[4] W. E. L. Grimson and D. P. Huttenlocker. On the sensitivity of the hough transform for object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):255-274, Mar 1990.

[5] W.E.L. Grimson and T. Lozano-Perez. Lo alizing overlapping parts by searching the interpretation tree. *IEEE Trans. on PAMI*, 9(4):469-482, 1987.

[6] Teresa M. Silberberg, David A. Harwood, and Larry S. Davis. Object recognition using oriented model points. *Computer Vision, Graphics, and Image Processing*, 35:47-71, 1986.

[7] R. Talluri and J. K. Aggarwal. Edge visibility regions - a new representation of the environment of a mobile robot. In *IAPR Workshop on Machine Vision Applications, MVA '90*, pages 375-380, Tokyo, Japan, November 1990.

[8] R. Talluri and J. K. Aggarwal. Mobile robot self location using constrained search. In *Proc. IEEE Workshop on Intelligent Robots and Systems, IROS '91*, Japan, November 1991.

[9] R. Talluri and J. K. Aggarwal. Position estimation for a mobile robot in an outdoor environment. To appear in the *IEEE Transactions on Robotics and Automation*, 1992.

[10] D. W. Thompson and J. L. Mundy. Three-dimensional model matching from an unconstrained viewpoint. In *Proc. of IEEE Intl. Conf on Robotics and Automation*, pages 208-220 Raleigh, March 1987.

Reprinted from *Computer Vision–ECCV '92, Second European Conference on Computer Vision*, Santa Margherita Ligure, Italy, May 1992 Proceedings. Springer Verlag Lecture Notes in Computer Science (588), Berlin, 1992, pp. 720-724.

Detecting 3-D Parallel Lines for Perceptual Organization*

Xavier Lebègue and J. K. Aggarwal

Computer and Vision Research Center, Dept. of Electrical and Computer Engr., ENS 520, The University of Texas at Austin, Austin, Texas 78712-1084, U.S.A.

Abstract.

This paper describes a new algorithm to simultaneously detect and classify straight lines according to their orientation in 3-D. The fundamental assumption is that the most "interesting" lines in a 3-D scene have orientations which fall into a few precisely defined categories. The algorithm we propose uses this assumption to extract the projection of straight edges from the image and to determine the most likely corresponding orientation in the 3-D scene. The extracted 2-D line segments are therefore "perceptually" grouped according to their orientation in 3-D. Instead of extracting all the line segments from the image before grouping them by orientation, we use the orientation data at the lowest image processing level, and detect segments separately for each predefined 3-D orientation. A strong emphasis is placed on real-world applications and very fast processing with conventional hardware.

1 Introduction

This paper presents a new algorithm for the detection and organization of line segments in images of complex scenes. The algorithm extracts line segments of particular 3-D orientations from intensity images. The knowledge of the orientation of edges in the 3-D scene allows the detection of important relations between the segments, such as parallelism or perpendicularity.

The role of perceptual organization [5] is to highlight non-accidental relations between features. In this paper, we extend the results of perceptual organization for 2-D scenes to the interpretation of images of 3-D scenes with any perspective distortion. For this, we assume *a priori* knowledge of prominent orientations in the 3-D scene. Unlike other approaches to space inference using vanishing points [1], we use the information about 3-D orientations at the lowest image-processing level for maximum efficiency.

The problem of line detection without first computing a free-form edge map was addressed by Burns et al. [2]. His algorithm first computes the intensity gradient orientation for all pixels in the image. Next, the neighboring pixels with similar gradient orientation are grouped into "line-support regions" by a process involving coarse orientation "buckets." Finally, a line segment is fit to the large line-support regions by a least-squares procedure. An optimized version of this algorithm was presented in [3].

The algorithm described in this paper is designed not only to extract 2-D line segments from an intensity image, but also to indicate what are the most probable orientations for the corresponding 3-D segments in the scene. Section 2 explains the geometry of

* This research was supported in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-89-C-0044, and in part by the Army Research Office under contract DAAL03-91-G-0050.

projecting segments of known 3-D orientation. Section 3 describes a very fast algorithm to extract the line segments from a single image and to simultaneously estimate their 3-D orientation. Finally, Sect. 4 provides experimental results obtained with images of indoor scenes acquired by a mobile robot.

2 Motivation and Assumptions

We chose to concentrate on objects which have parallel lines with known 3-D orientations in a world coordinate system. For example, in indoor scenes, rooms and hallways usually have a rectangular structure, and there are three prominent orientations for 3-D line segments: one vertical and two horizontal orientations perpendicular to each other. In this paper, any 3-D orientation is permitted, as long as it is given to the algorithm. Therefore, more complex environments, such as polygonal buildings with angles other than 90 degrees, are handled as well if these angles are known. It is important to note that human vision also relies on prominent 3-D orientations. Humans feel strongly disoriented when placed in a tilted environment.

Vertical lines constitute an interesting special case for two reasons: they are especially common in man-made scenes, and their 3-D orientation can easily be known in the 3-D camera coordinate system by measuring the direction of gravity. If a 2-axis inclinometer is mounted on the camera and properly calibrated, a 3-D vertical vector can be expressed in the 3-D coordinate system aligned with the 2-D image coordinate system. Inexpensive commercial inclinometers have a precision better than 0.01 degree. Humans also sense the direction of gravity by organs in their inner ears. In our experiments, we estimate the third angular degree of freedom of the camera relative to the scene from the odometer readings of our mobile robot. Provided that the odometer is constantly corrected by vision [4], the odometer does not drift without bounds.

We can infer the likely 3-D orientation of the line segments from their 2-D projections in the image plane. With a pinhole perspective projection model, lines parallel to each other in the 3-D scene will converge to a vanishing point in the 2-D projection. In particular, if the orientation of the camera relative to the scene is known, a vanishing point can be computed for each given 3-D orientation before the image is processed. All the lines that have a given orientation in 3-D *must* pass through the associated vanishing point when projected. Conversely, if a line *does not* pass through a vanishing point, it cannot have the 3-D orientation associated with that vanishing point. In practice, if a line *does* pass through a vanishing point when projected, it is likely to have the associated 3-D orientation.

To summarize, the line detection algorithm of Sect. 3 knows in each point of the image plane the orientation that a projected line segment would have if it had one of the predefined 3-D orientations. Therefore, the basic idea is to detect the 2-D segments with one of the possible orientations, and mark them with the associated 3-D orientation hypothesis.

3 Detecting Segments and Estimating their 3-D Orientation

3.1 Coordinate Systems and Transformations

The coordinate systems are W (the World coordinate system, with a vertical z-axis), R (the Robot coordinate system, in which we obtain the inclinometer and odometer readings), C (the Camera coordinate system), and P (the coordinate system used for

the perspective projection on the retina). The homogeneous coordinate transformation matrix from W to R is $T_{WR} = T_{roll} T_{pitch} T_{heading} T_{translations}$. T_{roll} and T_{pitch} are known with a good precision through the inclinometer. $T_{heading}$ is estimated by the odometer and $T_{translations}$ is not used here. T_{RC}, the coordinate transformation matrix from R to C, needs to be completely determined through eye/wheel calibration. Finally, T_{CP} is known through camera calibration.

3.2 Overview of the Algorithm

The processing can be outlined as follows:

1. Line support region extraction: compute the angle between the intensity gradient at each pixel and the expected direction of the projection of each 3-D orientation (see Sect. 3.3 for details). Use a loose threshold to allow for noise in the gradient orientation. Reject improper pixels and 3-D orientations.
2. Non-maxima suppression: keep only the local gradient maxima along the estimated perpendicular to the line.
3. Pixel linking: create chains of pixels using a partial neighborhood search in the direction of the estimated vanishing points. This creates noisy linear chains.
4. Line fitting: perform a least-squares fit of line segments to the pixel chains. Recursively break the pixel chains which cannot be closely approximated with a line segment into smaller chains.
5. Global orientation check: compute the match between each line and each 3-D orientation, like in the line support extraction step but with a much tighter threshold.

If the *a priori* heading is very uncertain, the lines will be extracted with loose thresholds, the true heading will be estimated, and the algorithm can then be run again with tight thresholds for the correct categorization.

3.3 Extracting Line Support Regions

For each pixel in the input intensity image and for each category of possible 3-D orientations, we compute the angle between the intensity gradient and the expected direction of the line in 2-D. The expected line is given by the current pixel and the vanishing point associated with the 3-D orientation. It is not necessary to compute the location of the vanishing point (which may lie at infinity).

The homogeneous transformation matrix changing world coordinates into projective coordinates is $T_{WP} = T_{CP} T_{RC} T_{WR}$. Let $[p_x, p_y, p_z, 0]^T_W$ be a non-null vector in the 3-D direction under consideration. If $[su, sv, s, 1]^T_P = T_{WP} [x, y, z, 1]^T_W$ defines the relation between a 2-D point $[u, v]^T$ and its antecedent by the perspective projection, then

$$[s'u', s'v', s', 1]^T_P = T_{WP} \left([x, y, z, 1]^T_W + [p_x, p_y, p_z, 0]^T_W \right)$$

defines another point of the estimated 2-D line. A 2-D vector $\mathbf{d}$ in the image plane pointing to the vanishing point from the current point is then collinear to $[u' - u, v' - v]^T$. Algebraic manipulations lead to $[d_u, d_v]^T = [a_x - a_z u, a_y - a_z v]^T$ where

$$[a_x, a_y, a_z, 0]^T = T_{WP} [p_x, p_y, p_z, 0]^T_W \ .$$

Note that a_x, a_y, and a_z need to be computed only once for each 3-D orientation.

The current pixel is retained for the 3-D direction under consideration if the angle between $\mathbf{d}$ and the local gradient $\mathbf{g}$ is 90 degrees plus or minus an angular threshold γ. This can be expressed by

$$\frac{\|\mathbf{d} \times \mathbf{g}\|}{\|\mathbf{d}\| \cdot \|\mathbf{g}\|} > \cos\gamma$$

or equivalently:

$$(d_x\, g_y - d_y\, g_x)^2 > (d_x^2 + d_y^2)\,(g_x^2 + g_y^2)\,\Gamma$$

with $\Gamma = (\cos\gamma)^2$ computed once for all. Using this formulation, the entire line support extraction is reduced to 8 additions and 11 multiplications per pixel and per 3-D orientation. If an even greater speedup is desired, $(g_x^2 + g_y^2)$ may be computed first and thresholded. Pixels with a very low gradient magnitude may then be rejected before having to compute $\mathbf{d}$.

4 Results

The algorithm was implemented in C on an IBM RS 6000 Model 530 workstation, and tested on hundreds of indoor images obtained by our mobile robot. The predefined 3-D orientations are the vertical and the two horizontal orientations perpendicular to each other and aligned with the axes of our building. Figures 1 and 2 show the results of line extraction for one image in a sequence. The processing time is only 2.2 seconds for each 512 by 480 image. Preliminary timing results on a HP 730 desktop workstation approach only a second of processing, from the intensity image to the list of categorized segments. The fast speed can be explained partly by the absence of multi-cycle floating-point instructions from the line *orientation equations*, when properly expressed.

The lines are not broken up easily by a noisy gradient orientation, because the orientation "buckets" are wide and centered on the noiseless gradient orientation for each 3-D orientation category. The output quality does not degrade abruptly with high image noise, provided that the thresholds for local gradient orientations are loosened. The sensitivity to different thresholds is similar to that of the Burns algorithm: a single set of parameters can be used for most images. A few misclassifications occur in some parts of the images, but are marked as ambiguities.

We have compared the real and computed 3-D orientation of 1439 detected segments from eight images in three different environments. The presence of people in some scenes, as well as noise in the radio transmission of images, did not seem to generate many misclassifications. The most frequent ambiguities occurred with horizontal segments parallel to the optical axis: 1.1 % of them were classified as possibly vertical in 3-D.

5 Conclusion

We have presented a new algorithm for detecting line segments in an image of a 3-D scene with known prominent orientations. The output of the algorithm is particularly well suited for further processing using perceptual organization techniques. In particular, angular relationships between segments in the 3-D scene, such as parallelism or perpendicularity, are easily verified. Knowledge of the 3-D orientation of segments is a considerable advantage over the traditional 2-D perceptual organization approach. The orientation thresholds of the 2-D perceptual organization systems cannot handle a significant perspective distortion (such as the third orientation category in Fig. 2). The independence from the perspective distortion brings more formal angular thresholds to

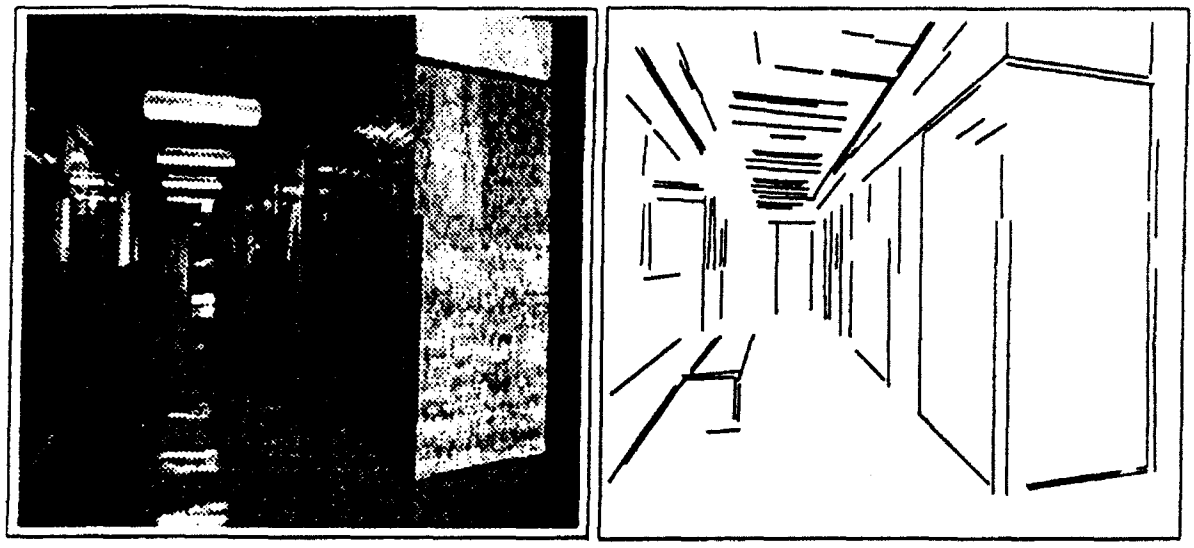


Fig. 1. (a) The input intensity image, and (b) the 2-D segments



Fig. 2. The line segments associated with each 3-D orientation

the perceptual organization process. By using the 3-D orientation at the lowest image processing level, both the quality and speed of the algorithm were improved. The ultimate benefits of this approach were demonstrated on real images in real situations.

References

1. S. T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4):435–462, November 1983.
2. J. B. Burns, A. R. Hanson, and E. M. Riseman. Extracting straight lines. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8(4):425–455, July 1986.
3. P. Kahn, L.Kitchen, and E. M. Riseman. A fast line finder for vision-guided robot navigation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(11):1098–1102, November 1990.
4. X. Lebègue and J. K. Aggarwal. Extraction and interpretation of semantically significant line segments for a mobile robot. To appear in *Proc. IEEE Int. Conf. Robotics and Automation*, Nice, France, May 1992.
5. D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.

Proceedings of the 1992 IEEE
International Conference on Robotics and Automation
Nice, France - May 1992

Extraction and Interpretation of Semantically Significant Line Segments for a Mobile Robot*

Xavier Lebègue and J. K. Aggarwal
Computer and Vision Research Center,
Dept. of Electrical and Computer Engr., ENS 520,
The University of Texas at Austin,
Austin, Texas 78712-1084, U.S.A.

Abstract

This paper describes new algorithms for detecting and interpreting linear features of a real scene as imaged by a single camera on a mobile robot. The low-level processing stages are specifically designed to increase the usefulness and the quality of the extracted features for a semantic interpretation. The detection and interpretation processes provide a 3-D orientation hypothesis for each 2-D segment. This in turn is used to estimate the robot's orientation and relative position in the environment and to delimit the free space visible in the image. Next, the orientation data is used by a motion stereo algorithm to fully estimate the 3-D structure when a sequence of images becomes available. From detection to 3-D estimation, a strong emphasis is placed on real-world applications and very fast processing with conventional hardware.

1 Introduction

This paper presents a new approach for extracting *semantically significant* line segments from monocular images, for estimating a mobile robot's orientation and relative position, and for identifying key objects associated with the free space, such as walls and doorways. The common idea is to design all the stages of image interpretation, including the lowest image processing level, such that they provide the higher stages with the most semantically useful features.

The tasks to be accomplished by the robot are usually specified in high-level semantic terms, such as "go down the hallway and go through the last door to the left." In order to execute this task, the robot must be able to identify the objects of interest (here, hallways and doors) in its perception of the environment. One approach is to reconstruct a 3-D line segment description of the environment of the robot from several intensity images (see for example [1]), and then to match groups of those segments to selected object models. For this, several 3-D segments are selected according to orientation and position criteria [9]. A 3-D hypothesis of an object is generated and matched to the selected segments. The process is repeated for other segments and other hypotheses. The high-level semantic interpretation is then used for determining the free space or finding objects of interest.

Since the high-level interpreter is looking for line segments of particular orientations in 3-D, we designed our lower-level image-processing stages to take advantage of this information. We will show practical examples of how this top-down information can benefit the feature extraction stage by reducing the amount of unwanted features, increasing the sensitivity to good features, and drastically speeding the computation.

Section 2 of this paper discusses the *a priori* knowledge of a few prominent 3-D orientations for lines in the scene. Section 3 describes a very fast algorithm to extract useful line segments from a single image. The algorithm simultaneously hypothesizes the 3-D orientation of segments. Section 4 shows how the robot's orientation in the scene is extracted from a single image, and how the results of Section 3 may be used for a first semantic interpretation of the scene without the 3-D depth information. Finally, Section 5 extends these results to a sequence of *monocular* images. A Kalman filter is used for a recursive estimation of the position of 3-D segments. Unlike purely metrical approaches for the recursive estimation, the qualitative information derived independently for each image is used to enhance the quality and usefulness of the

*This research was supported in part by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) Contract F49620-89-C-0044, and in part by the Army Research Office under contract DAAL03-91-G-0050.

results while decreasing the computation time. The high-level semantic interpretation of the resulting 3-D maps is discussed. Experimental results with an indoor mobile robot are provided and analyzed.

2 Motivation and assumptions

2.1 Knowledge of prominent 3-D orientations

Estimating the 3-D structure of a scene from a single visual image is impossible without certain assumptions about the structure and the projection geometry. We chose to concentrate on objects which have parallel lines with known 3-D orientations in a world coordinate system. For example, in indoor scenes, rooms and hallways usually have a rectangular structure, and there are three prominent orientations for 3-D line segments: one vertical and two horizontal orientations perpendicular to each other. Outdoor urban scenes also display similar characteristics. In this paper, any 3-D orientation is permitted, as long as it is given to the algorithms. Therefore, more complex environments, such as polygonal buildings with angles other than 90 degrees, are handled as well if these angles are known. The assumption of known prominent 3-D orientations is consistent with the goal of identifying objects of interest for a mobile robot, such as walls, doorways, or entire buildings.

2.2 Orientation of the camera relative to the scene

Vertical lines constitute an interesting special case for two reasons: they are especially common in man-made scenes, and their 3-D orientation can easily be known in the 3-D camera coordinate system. If a 2-axis inclinometer is mounted on the mobile robot and properly calibrated, a 3-D vertical vector can be expressed in the 3-D coordinate system aligned with the 2-D image coordinate system. Inexpensive commercial inclinometers have a precision better than 0.01 degrees. Although vibrations and calibration errors may corrupt this measurement, it can serve as a good estimate of the camera's roll and pitch. We will assume that either the robot is equipped with an inclinometer, or that it drives on an approximately planar horizontal surface.

We estimate the third angular degree of freedom of the camera from odometer readings of the robot. Provided that the odometer is constantly corrected by vision (which it will be in Section 4), the odometer does not drift without bounds. The orientation of the camera with respect to the scene follows from the orientation of the robot and the eye/wheel calibration data.

2.3 Coordinate systems and transformations

The coordinate systems , as shown in Figure 1, are W (the World coordinate system, with a vertical z-axis), R (the Robot coordinate system), C (the Camera coordinate system), and P (the coordinate system used for the perspective projection on the retina).

The homogeneous coordinate transformation matrix from W to R is:

$$T_{WR} = T_{roll}^{robot}\, T_{pitch}^{robot}\, T_{heading}^{robot} \begin{bmatrix} 1 & 0 & 0 & -x_W^{robot} \\ 0 & 1 & 0 & -y_W^{robot} \\ 0 & 0 & 1 & -z_W^{robot} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

x_W^{robot}, y_W^{robot}, and z_W^{robot} are approximated by the odometer but are not used in this section.

T_{RC}, the coordinate transformation matrix from R to C, needs to be completely determined through eye/wheel calibration. T_{CP}, the perspective projection matrix, is also determined through calibration. Therefore, we have

$$\begin{bmatrix} su \\ sv \\ s \\ 1 \end{bmatrix}_P = T_{CP} T_{RC} T_{WR} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_W$$

where u and v represent the coordinates in pixels of a point on the image plane.

2.4 Projecting segments of known 3-D orientation

In this section, we draw inferences about the likely 3-D orientation of the line segments from their 2-D projections in the image plane. With a pinhole perspective projection model, lines parallel to each other in the 3-D scene will converge to a vanishing point in the 2-D projection (refer to Figure 2). The knowledge of vanishing points provides useful information about the structure of the scene and about the camera's orientation with respect to the scene [2, 3, 10]. In particular, if the orientation of the camera relative to the scene is known, a vanishing point can be computed for each given 3-D orientation before the image

s processed. All the lines that have a given orientation in 3-D *must* pass through the associated vanishing point when projected. Conversely, if a line does not pass through a vanishing point, it cannot have the 3-D orientation associated with that vanishing point. In practice, if a line does pass through a vanishing point when projected, it is *likely* to have the associated 3-D orientation. The results of Section 3 will show how likely this is in practical situations.

To summarize, the line detection algorithm of Section 3 knows in each point of the image plane the orientation that a projected line segment would have if it had one of the predefined 3-D orientations. Therefore, the basic idea is to detect the 2-D segments with one of the possible orientations, and mark them with the associated 3-D orientation hypothesis.

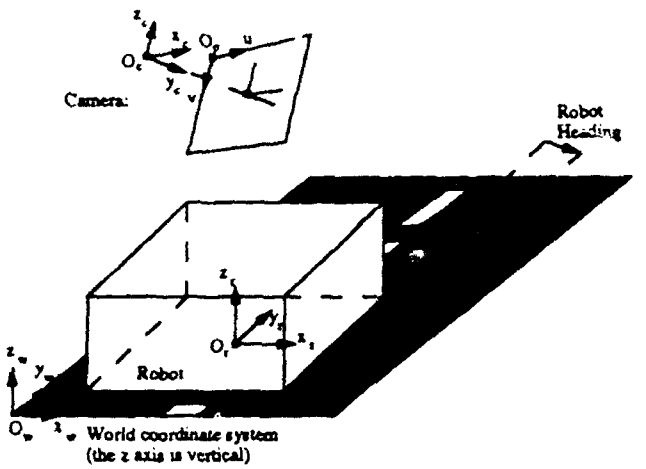


Figure 1: **The coordinate systems**

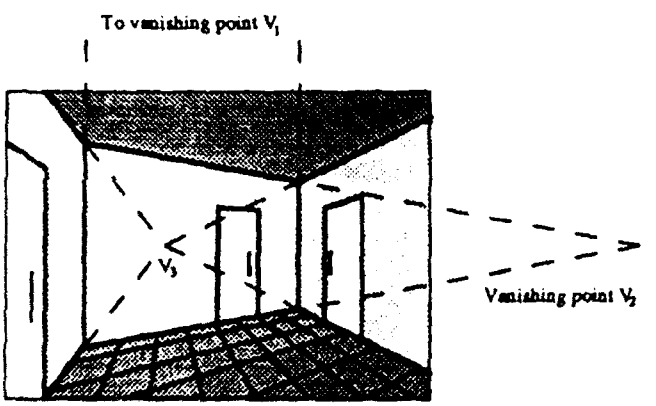


Figure 2: **The effects of the heading on segment orientations**

3 Detecting segments and estimating their 3-D orientation

3.1 Overview of the algorithm

The processing can be outlined as follows (see [7] for details):

1. Line support region extraction: compute the angle between the intensity gradient at each pixel and the expected direction of the projection of each 3-D orientation. Use a loose threshold to allow for noise in the gradient orientation. Reject improper pixels and 3-D orientations. Using the formulation of [7], the entire line support extraction is reduced to 8 additions and 11 multiplications per pixel and per 3-D orientation.

2. Non-maxima suppression: keep only the local gradient maxima along the estimated perpendicular to the line.

3. Pixel linking: create chains of pixels using a partial neighborhood search in the direction of the estimated vanishing points. This creates noisy linear chains.

4. Line fitting: perform a least-squares fit of line segments to the pixel chains. The pixel chains which cannot be closely approximated with a line segment are broken recursively into smaller chains.

5. Global orientation check: compute the match between each line and each 3-D orientation, like in the line support extraction step but with a much tighter threshold.

If the robot's *a priori* roll, pitch and heading are very uncertain, the lines will be extracted with loose thresholds, the true angles will be estimated, and the algorithm can then be run again with tight thresholds for the correct categorization.

3.2 Results

The algorithm was implemented in C on an HP 730 workstation, and tested on hundreds of indoor images obtained by our mobile robot (see [8] for a video presentation of the results). The predefined 3-D orientations are the vertical and the two horizontal orientations perpendicular to each other and aligned with the axes of our building. Figures 3 to 5 show the results of line extraction for one image in a sequence (The intensity image is deformed because the camera is equipped with a 6 mm wide-angle lens, as described

in Section 5). The processing time is only 1.2 seconds for each 512 by 480 image, from the intensity image to the list of categorized segments. The fast speed can be explained partly by the absence of multi-cycle floating-point instructions from the line orientation equations, when properly expressed. The lines are not broken up easily by a noisy gradient orientation, and the output quality does not degrade abruptly with high image noise, provided that the thresholds for local gradient orientations are loosened. A few misclassifications occur in some parts of the images, but are marked as ambiguities.



Figure 3: **The input intensity image**



Figure 4: **The line segments associated with each 3-D orientation**

Table 1 compares the real and computed 3-D orientation of 1439 detected segments from eight images in three different environments. The presence of people in some scenes, as well as noise in the radio transmission of images, did not seem to generate many misclassifications. The percentages in each column sum to more than 100 % because of the multiple classification of ambiguous segments.

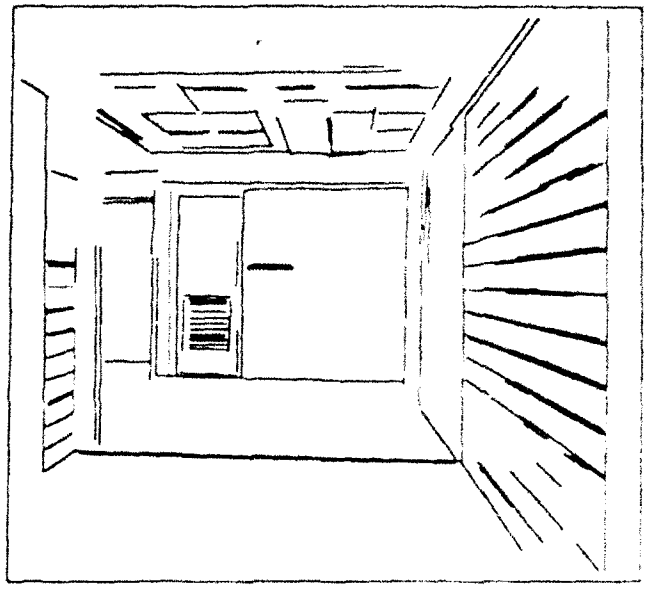


Figure 5: **All of the 2-D segments**

How detected segments are classified				
Classified as:	Real 3-D orientation			
	Vertical	Hor. 1	Hor. 2	Other
Vertical	100 %	0 %	1.1 %	27 %
Horiz. 1	0 %	100 %	0.3 %	0 %
Horiz. 2	0.4 %	0 %	99.8 %	73 %
Total	100.4 %	100 %	101.2 %	100 %
No. of seg.	217	88	1123	11

Table 1: **Hypothesizing 3-D orientations: success and error rates**

4 Using information from a single image

4.1 Computing the robot's orientation

4.1.1 Algorithm

In this section we show how the robot's roll, pitch and heading may be computed *a posteriori* from the 2-D segments extracted in Section 3. To achieve the best precision, each angle should be evaluated from the vanishing point which varies the most with that angle. We begin with the roll, which is best derived from the 3-D vertical lines if the camera is looking approximately forward. Because of our definition of the heading (a rotation around the vertical axis of W), the vanishing point associated with the vertical lines does not depend on the heading. If all the 2-D segments classified as 3-D vertical are parallel in the image, the

vanishing point lies at infinity, the pitch p is zero, and the roll r is the 2-D orientation of the segments in the image. Otherwise, we first estimate the location (u_v, v_v) of the vanishing point by a weighted least-squares method. Let $T^{RP}_{i,j}$ be the element i, j of the matrix T_{RP}. After algebraic manipulations, we find:

$$r = \arctan \frac{\begin{pmatrix} (T^{RP}_{2,3}T^{RP}_{3,2} - T^{RP}_{3,3}T^{RP}_{2,2})u_v + \\ (T^{RP}_{3,3}T^{RP}_{1,2} - T^{RP}_{1,3}T^{RP}_{3,2})v_v + \\ (T^{RP}_{1,3}T^{RP}_{2,2} - T^{RP}_{2,3}T^{RP}_{1,2}) \end{pmatrix}}{\begin{pmatrix} (T^{RP}_{2,1}T^{RP}_{3,2} - T^{RP}_{3,1}T^{RP}_{2,2})u_v + \\ (T^{RP}_{3,1}T^{RP}_{1,2} - T^{RP}_{1,1}T^{RP}_{3,2})v_v + \\ (T^{RP}_{1,1}T^{RP}_{2,2} - T^{RP}_{2,1}T^{RP}_{1,2}) \end{pmatrix}}$$

The pitch of the robot may also be computed from the vanishing point of vertical lines, but for a better precision, we use another vanishing point more dependent on pitch and heading. In our implementation, we use the vanishing point which lies closest to the center of the image. We obtain the pitch and the heading with a method identical to the computation of roll described above.

4.1.2 Results

The quality of the results depends on the error on the *vanishing point extraction* and the precision of calibration. To illustrate the performance of heading computation from individual images, we plotted the path of the robot around a building floor. In Figure 6, the heading extracted by the vision algorithm was combined to the translation given by odometry to estimate the path of the robot. At the beginning and end of the experiment, the robot was approximately aligned with the center of a hallway. The robot stopped about five meters before the departure point. The total traveled distance was 125 meters, and 97 images were used to compute the headings. The algorithm seemed unaffected by people moving in the field of view and by a significant noise due to the transmission of the images by radio. The processing time for *a posteriori* heading estimation is negligible compared to the time needed for line extraction.

4.2 Image interpretation

Interpreting a single segment image is necessary when no 3-D segment information is available. This occurs for the first frames of a motion stereo sequence, or when the baseline of a stereo setup is too small compared to the depth of the scene. Ideally, interpreting a single segment image should not be too different from interpreting a 3-D segment map to maintain consistency when the latter representation becomes available. Identifying objects by searching for segments of a particular 3-D orientation is a good step to achieve this consistency.

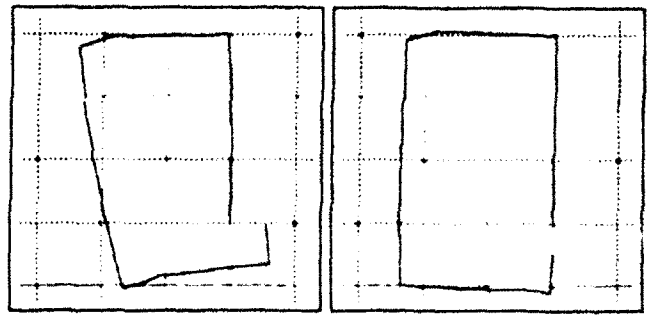


Figure 6: The path of the robot (a) as estimated by odometry, and (b) after correction by vision

The interpretation of the individual segments in a single 2-D image requires *a priori* knowledge about the type of scene. It can be achieved by heuristic methods. The knowledge of the orientation of 3-D segments is particularly well suited to finding the floor and the walls in indoor images. Extracting the floor is essential for map-making, for determining the free space, and for computing the position of the robot relative to the visible environment. In most cases, the outline of the floor can be described entirely by segments of predefined horizontal 3-D orientation. A heuristic method can extract an approximate outline of the floor from a single segment image. In real applications, the floor may not be perfectly extracted, but this first approximation can be corrected after a sequence of images becomes available, by estimating recursively the 3-D position of the segments.

5 Estimating the position of 3-D segments

5.1 Introduction

The recursive estimation of 3-D segments from a sequence of images is the next logical step toward scene understanding. It aims at estimating the depth information that is absent from 2-D images. Existing approaches include [1, 4, 5], to name only a few. Viéville and Faugeras [11] concentrate on estimating motion and structure from *monocular* segment images and present results on the reconstruction of a calibration pattern. Jezouin and Ayache [6] review different

segment-tracking techniques using monocular images and present results on 3-D reconstruction from synthetic aerial images.

Applications in autonomous mobile robots imply the ability to deal with real indoor scenes while using limited computation resources. As in the other sections of this paper, we aim at providing useful features for a semantic interpretation of the environment of the robot. We only consider the segments which have been extracted by the algorithm of Section 3. We will show how this choice benefits the robustness and speed of the recursive estimation process.

5.2 Preliminary observations

A 3-D segment should be observed from very different angles to be accurately reconstructed. Two approaches can be used to achieve this with a monocular camera. The first one, active vision, pans and tilts the camera to keep some interesting features in the field of view. The second approach, which we chose, consists of fitting the camera with a wide-angle lens. A 3-D segment is usually first detected when far ahead of the robot, but as the robot drives closer, the 2-D projection moves to one side of the focus of expansion. A short focal length will enable the 3-D segment to remain in the field of view long enough for an accurate reconstruction. The drawback to using a wide-angle lens is geometric distortion. The first processing step is therefore to correct this distortion.

The 3-D segments which are parallel to the direction of motion cannot be properly reconstructed without considering their endpoints. The experience of all researchers using real images indicates that endpoints are unreliable, because 2-D segments may be broken or extended (due to specular reflection, for example). Therefore, we will not use endpoints to estimate the parameters of infinite 3-D lines. There is no justification for wasting computation to estimate features that will not help the interpretation of the scene. In practice, we will not attempt to reconstruct the 3-D segments parallel to the direction of motion. Those segments are sent to the semantic interpreter with an estimate of orientation but not of position, unless the robot changes its course. This contrasts with other approaches that do send an estimate of the position and orientation of those segments, both with a huge uncertainty.

Figure 7 presents a block diagram of the recursive estimation of 3-D segments. As in Section 4, we combine translations from odometry to rotations from vanishing points to estimate the motion of the robot. Because the rotations are obtained from each 2-D image with an excellent accuracy, we do not introduce any uncertainty on their parameters. Translations, however, rely partly on odometry. Therefore, we model the location of the robot in the world coordinate system by a 3-D vector t_k and its covariance matrix $E(t_k t_k^T)$ under the Gaussian noise assumption.

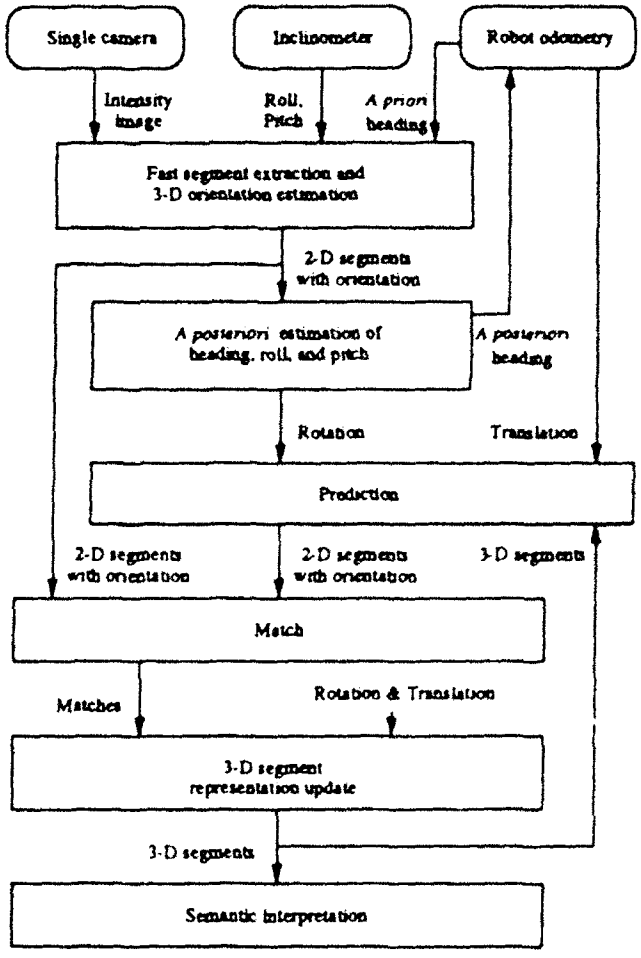


Figure 7: Recursive estimation of 3-D segments

5.3 Representation of lines

Since the algorithm of Section 3 already provides the most likely 3-D line orientation, we only estimate two positional parameters for each 3-D line. Let us define a coordinate system L_n for each predefined 3-D orientation n as follows: L_n is a rotation of W so that the 3-D orientation n is parallel to the z axis of L_n. The $T_{L_n W}$ matrices are all known by definition. The state vector of a tracked 3-D line is then made of the first two components of the coordinates in L_n of any point of the line.

Because we are not using stereovision, a problem arises when a 2-D segment is acquired for the first time. Our approach is to model it by a state vector with arbitrary depth and a very large uncertainty in depth. After repeated observations from different angles, the uncertainty ellipsoid associated with the state vector will shrink around the best estimate of the position.

The measurement $y_{k,i}$ of a 2-D line in frame k given its associated 3-D orientation is defined as follows: if the line is "rather horizontal," it can be described in the image plane by $v = \alpha u + y_{k,i}$ with some parameter α. If the line is "rather vertical," $y_{k,i}$ is such that $u = \alpha v + y_{k,i}$. Deciding whether a line is rather horizontal or vertical is based on the estimated 3-D orientation, so that singularities in the representation are avoided yet consistency across frames is maintained. The explicit state and measurement equations for the Kalman filter are not derived here for lack of space.

5.4 Implementation and results

The implementation benefits from hypothesizing the 3-D orientation of lines before estimating their 3-D position in the following ways:

- Matching is attempted only between segments of the same orientation in 3-D. In practice, other implementations of 3-D segment recursive estimators use arbitrary *2-D* orientation "buckets" to achieve the same purpose. By matching segments of given 3-D orientations, we have a more formal approach to the problem of limiting possible matches.

- No computation is wasted on segments which cannot be reconstructed because they are parallel in 3-D to the direction of motion.

- The problem of singularities in the representation (see [6] for example) is avoided: although there is still a need for different line parameterizations, the appropriate one needs to be determined only once for each segment being tracked.

- Traditional approaches have difficulties dealing with moving objects and a moving camera. Large moving objects will corrupt the 3-D map of the environment and yield incorrect motion estimates if the egomotion is estimated from the image. By tuning the line extractor to certain orientations of interest, it is unlikely that an unexpected moving object could durably display misleading features. Therefore, the identification of the visible objects of interest will not be significantly hampered by large moving objects. This is extremely important in real-world situations.

Several approaches are possible for tracking segments across images. If a conventional token-tracker is used [5], matching is necessary only between segments with the same 3-D orientation. The results we are presenting here have been obtained with manual token-tracking. We are currently working on a new type of token-tracker that relies on topological and semantic information to eliminate incorrect matches.

Figure 8 shows the vertical segments in a top view of a corridor, as estimated after 1, 4, 7, and 10 images in a sequence. The robot moved nine meters during the sequence, or about a quarter of the length of the corridor (the motion is from left to right in the figure). The segments visible for the first time are given an arbitrary depth and a very large uncertainty in depth. The width of the corridor at the location of the robot was estimated with an error varying from 0.4% to 4.7%.

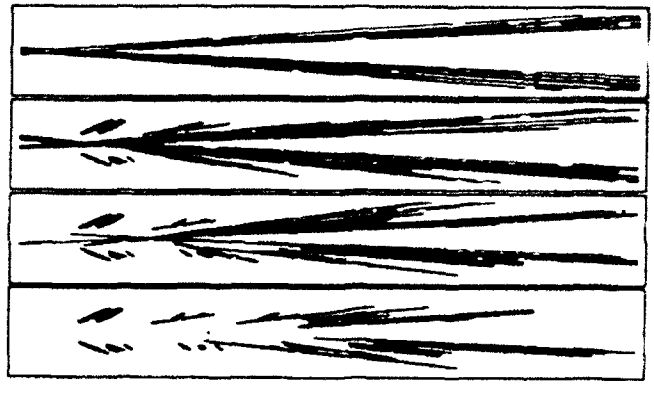


Figure 8: **Top view of a corridor after 1, 4, 7, and 10 images (uncertainty of vertical segments)**

5.5 High-level semantic interpretation

The semantic interpretation of the 3-D segment maps should be consistent with the interpretation of single images since the 3-D information is not present or accurate for some segments. The search for segments associated with objects of interest depends primarily on 3-D segment orientation and 2-D image topology. The position of 3-D segments is used in a later stage to verify object hypotheses and to estimate their range. The reflection of segments in the floor, often a problem with 2-D interpretation, is handled by clipping the 3-D segments below the floor plane.

Vertical segments in the top view				
Door no.	Δx (in mm)	Δy (in mm)	Error in Δx	Angle of the frame
1	917	-54	-9.7%	-3.4 deg
2	1083	4	6.6%	0.2 deg
3	1013	-18	-0.3%	-1.0 deg
4	996	25	-2.0%	1.4 deg
5	769	-62	-24.3%	-4.6 deg
6	1114	16	9.6%	0.8 deg
7	421	-72	-58.6%	9.7 deg
8	292	99	-71.3%	18.7 deg

Table 2: **Comparing estimated door frame edges to their model**

The semantic interpretation of particular objects is application-dependent. To illustrate how the recursive estimation of oriented segments can be used by an interpreter, we will describe the detection of the door frames in the corridor of Figure 8. The door frames are aligned with the axes of the building, therefore segments can be selected based on their 3-D orientation and 2-D topology. To verify a door frame hypothesis, the two selected vertical segments are checked for their distance and the 3-D location of their endpoints. The segments should also belong to a 3-D plane formed by two predefined 3-D orientations. Since the positions of the 3-D segments are uncertain, the Mahalanobis distance is used to adapt the thresholds.

As an illustration, Table 2 analyzes the vertical edges of the door frames shown in the last top view of Figure 8. Ideally, the two vertical segments should be 1016 mm apart and the angle of the door frame in the scene should be zero. The last two doors are about 9 meters ahead of the robot and the 3-D position of their segments is still very uncertain. The number of images in which the third and fourth door frames were visible explains the high precision of their reconstruction.

6 Conclusion

We have presented a new approach to the interpretation of monocular line images for the navigation of a mobile robot. We described algorithms for interpreting single images and for recursively updating interpretations with monocular sequences of images. The quality and usefulness of the results of each processing stage justified the practical constraint placed on the problem: the *a priori* knowledge of a few prominent 3-D segment orientations. This constraint is reasonable since even humans get disoriented when placed in a tilted environment, despite having our own "inclinometers" in our inner ears. By reasoning in terms of what features the high-level interpretation stage will actually use, we were able to design lower-level algorithms for better and faster processing. The ultimate benefits of this approach were demonstrated on real images in real situations, using conventional computing hardware, and with image acquisitions every few seconds.

References

[1] N. Ayache. *Artificial Vision for Mobile Robots*. The MIT Press, 1991.

[2] S. T. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4):435–462, November 1983.

[3] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–139, March 1990.

[4] J. Crowley and P. Stelmaszyk. Measurement and integration of 3-D structures by tracking edge lines. In *Proc. First European Conf. on Computer Vision*, pages 269–280, Antibes, France, April 1990. Springer-Verlag.

[5] R. Deriche and O. Faugeras. Tracking line segments. In *Proc. First European Conf. on Computer Vision*, pages 259–268, Antibes, France, April 1990. Springer-Verlag.

[6] J. L. Jezouin and N. Ayache. Computing 3D structure from a monocular sequence of images with known motion. Technical report, Matra-LTIS, BP 235, 78052 Guyancourt, France, 1991.

[7] X. Lebègue and J. K. Aggarwal. Detecting 3-D parallel lines for perceptual organization. In *Proc. Second European Conf. on Computer Vision*, Santa Margherita Ligure, Italy, May 1992. Springer-Verlag.

[8] X. Lebègue and J. K. Aggarwal. Semantically significant line segments for a mobile robot... ROBOTEX. In *Video Proc. IEEE Int. Conf. Robotics and Automation*, Nice, France, May 1992.

[9] M. Thonnat. Semantic interpretation of 3-D stereo data: finding the main structures. In *Proc. Eighth Int. Conf. on Pattern Recognition*, pages 1051–1054, Paris, France, October 1986.

[10] T. Tsubouchi and S. Yuta. Map assisted vision system of mobile robots for reckoning in a building environment. In *Proc. IEEE Int. Conf. Robotics and Automation*, pages 1978–1984, Raleigh, March 1987.

[11] T. Viéville and O. Faugeras. Feed-forward recovery of motion and structure from a sequence of 2D-line matches. In *Proc. Third Int. Conf. on Computer Vision*, pages 517–520, Osaka, Japan, December 1990.

A Digital Technique to Estimate Second-Order Distortion Using Higher Order Coherence Spectra

Yong Soo Cho, Sung Bae Kim, *Member, IEEE*, Elmer L. Hixson, *Life Senior Member, IEEE*, Edward J. Powers, *Fellow, IEEE*

***Abstract*—A digital spectral method to evaluate second-order distortion of a nonlinear system, which can be represented by Volterra kernels up to second order and which is subjected to a random noise input, is discussed. The importance of departures from the commonly assumed Gaussian excitation is investigated. The Hinich test, which makes use of the statistical properties of bispectrum, is shown to be an appropriate test for orthogonality in the system identification of a second-order Volterra system. Tests for Gaussianity of two important sources, which are commonly used for Gaussian inputs in nonlinear system identification are presented: 1) commercial software routines for simulation experiments and 2) noise generators for practical experiments. The deleterious effects and misleading results of assuming a Gaussian input when, in fact, it is not, are demonstrated. The random input method to evaluate (in terms of the second-order Volterra kernels and distortion factors) the second-order distortion of a nonlinear system is compared with the sine-wave input method using both simulation and experimental data. The proposed approach is utilized to model, identify, and quantify the linear response and second-order distortion products of a loudspeaker in the low frequency band by using higher order coherence spectra.**

I. Introduction

DISTORTIONS due to electromechanical effects in the reproduction component are becoming more important as the state of the art in audio engineering continues to progress. The harmonic and intermodulation distortions, which are most significant in high-fidelity systems, usually result from nonlinear transfer characteristics in some component. The measurement of harmonic and intermodulation distortions, which are the major topic of this paper, can be conducted with tone generators and filters. For example, special instruments that measure harmonic distortion and intermodulation distortion in amplifiers and other audio components have been developed [1]. A harmonic distortion analyzer includes a tone generator plus a high-pass filter, which suppresses the fundamental frequency but allows the overtones to be measured. Harmonic distortion (HD) is expressed as a percentage [2]:

$$\% \text{ HD} = \frac{\text{sum of power in harmonics 2 and up}}{\text{total power}} \times 100\%.$$

To measure intermodulation distortion, a signal having two frequencies f_1 and f_2 is supplied to the component under test. Then the output from the component is fed into a high-pass filter, giving a waveform which consists of the high frequency carrier and its low frequency modulation. The percentage of intermodulation distortion (IMD) is given by

$$\% \text{ IMD} = \frac{\text{amplitude of modulation}}{\text{amplitude of carrier}} \times 100\%.$$

In contrast to the relatively common sine-wave input approach to measure distortion, this paper describes a digital higher order spectral technique to evaluate distortion for a nonlinear time-invariant component which can be represented by Volterra kernels and which is subjected to broad-band random inputs. The advantages of the random input approach are: a) one does not need to repeat the same experiments by changing frequencies of oscillators and filters to measure nonlinear distortions over the band of frequencies of interest, which is a tedious and time-consuming job in the sinusoidal input approach, b) band-limited noise is a closer approximation to actual input signals, e.g., music and speech, and c) we can quantify distortion over the entire frequency band using higher order coherence spectra.

The Volterra series approach [3], where the nonlinear response is expressed by a sum of linear, quadratic, and higher order functionals, has recently attracted a considerable amount of interest in the systems engineering area since many physical systems exhibiting inherent nonlinear characteristics can be studied and modeled by this approach [4], [5]. When one attempts to use the Volterra series to describe a nonlinear system, the system modeling problem reduces to the measurement of the Volterra kernels. A relatively common assumption underlying this approach involves the fact that the input is assumed to be

Manuscript received July 20, 1990; revised February 28, 1991. This work was supported by the DOD Joint Service Electronics Program under Contract AFOSR F49620-89-C-0044, and in part by the Department of Navy Contracts N00167-88-K-0049 and N00014-88-K-0638.

The authors are with the Department of Electrical and Computer Engineering and Electronics Research Center, the University of Texas at Austin, Austin, TX 78712-1084.

IEEE Log Number 9106575

1053-587X/92$03.00 © 1992 IEEE

a random waveform which possesses Gaussian statistics, an assumption which allows a substantial simplification of the relevant mathematics [6], [7]. The assumption of Gaussian excitation is also utilized in many technological areas, for example, the nonlinear response of offshore structures and ships to random seas [8]. In acoustics, experiments have often been performed under the assumption that the random noise generators used for an input signal have Gaussian statistics, see, for example, [9] and [2]. However, we will show using Hinich's Gaussianity test [10] that commonly used random inputs, such as that provided by a noise generator, cannot automatically be regarded as sufficiently Gaussian for nonlinear identification studies, even though the one-dimensional probability density function of the input is approximately Gaussian with a small skewness. We illustrate this point by utilizing the Hinich procedure to test for Gaussianity of two important sources, which are commonly used as Gaussian inputs in nonlinear system identification: 1) commercial software routines for simulation experiments, and 2) noise generators for practical experiments.

Recently Kim and Powers [11] presented, without assuming particular statistics for the input (i.e., the approach is valid for both Gaussian and non-Gaussian inputs), a practical digital method of estimating linear and quadratic transfer functions (the first- and second-order Volterra kernels in the frequency domain) of a nonliner time-invariant system, which may be described by a Volterra series up to second order. Furthermore, the degree of linearity and nonlinearity (quadratic in this case) as a function of frequency was described by generalizing the concept of coherence spectra. In this paper, we quantify the second-order nonlinear distortion of audio components in terms of coherence spectra at each frequency [12] as well as second-order Volterra kernels at sum and difference frequencies. In order to compare the random input method with the classical two-tone input method, we estimate the quadratic transfer function, which accounts for the second-order harmonic and intermodulation distortion, by both methods and compare them. We also demonstrate the deleterious effects and the misleading results of assuming a Gaussian input when, in fact, it is not sufficiently Gaussian [13]. Finally, the approach is utilized to model, identify, and quantify the linear response and second-order distortion products of a loudspeaker at low frequency.

After the problem definition and preliminaries are given in Section II, the importance of taking into account departures from Gaussianity in the excitation is discussed in Section III. Then, we describe the approach to quantifying nonlinear distortion with digital higher-order coherence spectra in Section IV. Bichromatic (two-tone) versus random input estimates for second-order distortion are described in Section V. In Section VI, we present the results of testing a noise generator output for Gaussianity and present experimental results of applying this approach of this paper to quantify second-order distortion of a loudspeaker. A summary is included in Section VII.

II. Problem Definition and Preliminaries

Consider a single-input/single-output nonlinear system which can be described by the Volterra series up to order two; let $y(t)$ be the output and $x(t)$ the input; the system is represented by

$$y(t) = \int_{-\infty}^{\infty} h_1(t-\tau)x(\tau)\,d\tau + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(t-\tau_1, t-\tau_2)x(\tau_1)x(\tau_2)\,d\tau_1\,d\tau_2 \tag{1}$$

where $h_1(\tau)$ and $h_2(\tau_1, \tau_2)$ represent the first-order and second-order Volterra kernels, respectively. In the transform domain, (1) may be written

$$Y(f) = H_1(f)X(f) + \int_{-\infty}^{\infty} H_2(f_1, f_2)X(f_1)X(f_2) \cdot \delta(f - f_1 - f_2)\,df_1\,df_2 \tag{2}$$

where $X(f)$ and $Y(f)$ are the Fourier transforms of $x(t)$ and $y(t)$, respectively. $\delta(f)$ is the Dirac delta function. The linear transfer function $H_1(f)$, which can be expressed in polar form as $|H_1(f)|\exp[j\phi_1(f)]$, is the first-order frequency domain Volterra kernel. The quadratic transfer function $H_2(f_1, f_2) = |H_2(f_1, f_2)|\exp[j\phi_2(f_1, f_2)]$ in polar form, is the two-dimensional Fourier transform of $h_2(t_1, t_2)$. The quadratic transfer function can always be written in a symmetric form, $H_2(f_i, f_j) = H_2(f_j, f_i)$, without loss of generality [3]. Since, in general, $H_2(f_1, f_2)$ will not be equal to $H_2(f_1, -f_2)$, different values of distortion at different frequencies are directly reflected in this second-order Volterra kernel.

The second-order factor of intermodulation distortion (D_{IM2}) at frequency $f_1 \pm f_2$ is defined, in decibels, as

$$D_{IM2} = 20\log[\text{IM2}] = 20\log\left[\left|\frac{V_0(f_1 \pm f_2)}{V_0(f_1)}\right|\right] \quad [\text{dB}] \tag{3}$$

where $V_0(f_1 \pm f_2)$ is the output voltage at the frequency $f_1 \pm f_2$ due to distortion and $V_0(f_1)$ is the output voltage at frequency f_1 due to the linear nature of the system under test. This term may be expressed in terms of the frequency domain Volterra kernels in the form [5]

$$D_{IM2} = 20\log\left[\left|\frac{H_2(f_1, \pm f_2)}{H_1(f_1)}\right| V_I\right] \quad [\text{dB}] \tag{4}$$

where V_I is the amplitude of both input voltages ($v_i(t) = V_I\cos 2\pi f_1 t + V_I\cos 2\pi f_2 t$). Likewise, the second-order harmonic distortion factor D_{H2}, which is defined as

$$D_{H2} = 20\log[H2] = 20\log\left[\left|\frac{V_0(2f)}{V_0(f)}\right|\right] \quad [\text{dB}] \tag{5}$$

may be expressed in terms of the Volterra kernels as

$$D_{H2} = 20\log\left[\left|\frac{1}{2}\right|\frac{H_2(f, f)}{H_1(f)}\right| V_I\right] \quad [\text{dB}] \tag{6}$$

where V_I is the amplitude of the input voltage ($v_i(t) = V_I \cos 2\pi ft$). Thus, the second-order intermodulation distortion factor D_{IM2} and harmonic distortion factor D_{H2} of a component under test can be obtained easily at various frequencies, if the linear transfer function and quadratic transfer function of the component can be estimated simultaneously at those frequencies.

Here we consider a general technique for analyzing the harmonic and intermodulation distortion of a nonlinear system with a random noise input, instead of sinusoidal inputs. As indicated in earlier paragraphs, the approach rests upon determination of the Volterra kernels. Thus, in this paper, we focus on the determination of the kernels and the quantification of the goodness of the resulting Volterra model through higher order coherence spectra. In general, one can express (2) using the discrete form as

$$Y(f_m) = H_1(f_m)X(f_m) + \sum_{i+j=m} H_2(f_i, f_j)X(f_i)X(f_j) + \epsilon(f_m) \quad (7)$$

where $X(f_m)$ and $Y(f_m)$ are the N-point discrete Fourier transforms of the input and the output, respectively. $\epsilon(f_m)$ is the difference between the model output $Y(f_m)_{\text{model}}$ (the first two terms on the RHS of (7)) and the true output $Y(f_m)$ at frequency $m\Delta f$ ($\Delta f = 1/NT$, T: sampling period). The symbol $f_{N/2}$ denotes the Nyquist frequency associated with the sampling of the time domain signal. Since (7) is linear with respect to the transfer functions, one can obtain an optimum estimation of the transfer functions by minimizing the mean-square error $E[|\epsilon(f_m)|^2]$. To simplify notation, $X(f_m)$ will be denoted as $X(m)$, $Y(f_m)$ as $Y(m)$, etc. To find the optimum solution, we consider

$$\frac{\partial E[|\epsilon(m)|^2]}{\partial H_1(m)} = 0$$

$$\frac{\partial E[|\epsilon(m)|^2]}{\partial H_2(i,j)} = 0, \qquad i + j = m \quad (8)$$

which yields the following two equations:

$$E[X^*(m)Y(m)] = H_1(m)E[|X(m)|^2] + \sum_{i+j=m} H_2(i,j)E[X^*(m)X(i)X(j)] \quad (9)$$

$$E[X^*(k)X^*(l)Y(m)] = H_1(m)E[X^*(k)X^*(l)X(m)] + \sum_{i+j=m} H_2(i,j)E[X^*(k)X^*(l)X(i)X(j)] \quad (10)$$

where * denotes the complex conjugation. In (9) and (10), $E[X^*(m)Y(m)]$ is the crosspower spectrum between input and output $S_{yx}(m)$, $E[|X(m)|^2]$ is the autopower spectrum of the input $S_{xx}(m)$, $E[X^*(m)X(i)X(j)]$ is the complex conjugate of the auto bispectrum of the input $B_{xxx}(i, j)$, and $E[X^*(k)X^*(l)Y(m)]$ is the cross bispectrum (where $k + l = m$) between input and output $B_{yxx}(k, l)$. Note that the fourth-order spectral moment appearing on the right-hand side of (10) becomes zero when $k + l \neq i + j = m$ because of the properties of higher order spectral moments and thus (10) is considered only for $k + l = i + j = m$.

III. Gaussianity in Quadratic System Identification

A relatively common assumption for the estimation of the linear and quadratic transfer functions in (9) and (10) is that the input possesses zero-mean Gaussian statistics, which allows terms containing the third-order moment (i.e., the auto bispectrum) of the input to vanish. Consequently, the estimations of the transfer functions for a zero-mean Gaussian excitation are independent and there is no coupling between them because, under Gaussian excitation, the first two terms on the right-hand side of the Volterra equation (7) are orthogonal. Thus, the linear and quadratic transfer functions for a zero-mean Gaussian input are obtained separately as follows [6], [7]:

$$H_1(m) = \frac{S_{yx}(m)}{S_{xx}(m)} \quad (11)$$

$$H_2(i,j) = \frac{B_{yxx}(i,j)}{2S_{xx}(i)S_{xx}(j)}, \qquad i + j \neq 0 \quad (12)$$

where the autopower spectrum is used in the denominator of (12) since fourth-order spectral moment appearing on the right-hand side of (10) can be expressed in terms of second-order spectral moments when the input is Gaussian. Thus, we recover the well-known result for the linear transfer function, namely, it is equal to the quotient of the crosspower spectrum divided by the autopower spectrum. The expression for the quadratic transfer function, although not as well known, is somewhat analogous and is given by the quotient of the cross bispectrum divided by the autopower spectra at frequency i and j. The factor of two appearing in the denominator of (12) arises from the symmetry property, $H_2(i, j) = H_2(j, i)$.

The assumption of Gaussian excitation made in (11) and (12), which allows a substantial simplification of the relevant mathematics, should be accepted with caution, even though the marginal distribution function of the excitation is approximately Gaussian. The Gaussianity of an input signal can be tested several ways. One of these is a moment test of the univariate marginals of the data as described by Pearson [14]; sample coefficients of skewness and kurtosis are used to reject or accept the Gaussianity hypothesis at a certain confidence level. The Gaussianity of the signal can be also checked by goodness-of-fit tests, such as the Kolmogorov-Smirnov test and chi-square test [15]. All of these tests focus on the univariate marginal distribution. The distribution of the marginals is compared to a Gaussian univariate distribution. However, as pointed out by Brockett *et al.* [16], the univariate marginal distribution of the time series may satisfy Gaussianity tests even if the joint distribution for several time points is non-Gaussian. For this reason, we use the so-

called Hinich test [10] to examine Gaussianity of the signal. In Hinich's approach, a sample bispectrum is used to construct a statistic to test whether the bispectrum of the signal is nonzero. A rejection of the null hypothesis implies a rejection of the hypothesis that the signal is Gaussian. More specifically, a sample bispectrum of an input signal $\{x(0), x(1), \cdots, x(N-1)\}$ is first defined by

$$F(i, j) = N^{-1} X(i) X(j) X^*(i + j). \tag{13}$$

To obtain a consistent estimate of the bispectrum, the sample bispectrum is smoothed over a square of M^2 points centered at $((2m-1)M/2, (2n-1)M/2)$ in the two-dimensional frequency plane. The estimator is

$$\hat{B}_{xxx}(m, n) = M^{-2} \sum_{i=(m-1)M}^{mM-1} \sum_{j=(n-1)M}^{nM-1} F(i, j). \tag{14}$$

Since the asymptotic distribution of each estimator in (14) is complex normal, the distribution of

$$\zeta_{m,n} = \frac{\hat{B}_{xxx}(m, n)}{[N/M^2]^{1/2} [\hat{S}_{xx}(m) \hat{S}_{xx}(n) \hat{S}_{xx}(m + n)]^{1/2}} \tag{15}$$

is complex normal with unit variance. Here, $\hat{S}_{xx}$ is the estimator of the power spectrum of $x(t)$. Consequently, $|\zeta_{m,n}|^2$ is approximately chi-square with two degrees of freedom. Since $\hat{B}_{xxx}(m, n)$ is zero for all m and n if the process is Gaussian, the statistic

$$T = 2 \sum_m \sum_n |\zeta_{m,n}|^2 \tag{16}$$

is asymptotically distributed $\chi^2(2P)$ under the null hypothesis of Gaussianity. Here, P is the degree of freedom defined in a principle domain [10]. For large N and thus large P, $(2P)^{-1}\chi^2(2P)$ is approximately normal. Finally, Gaussianity can be determined by shifting and scaling the statistic to standardize and compare it to the probability of exceeding its value under a Gaussian zero-mean unit-variance distribution.

This sample bispectrum test is especially meaningful when we estimate linear and quadratic transfer functions of a second-order Volterra series under the Gaussian input assumption. Since the process of estimating the linear and quadratic transfer function for a zero-mean Gaussian input is based on the assumption that the terms containing the third-order moment of the input signal in (9) and (10) vanish, the direct test for Gaussianity of the input would be to examine whether the autobispectrum of the input vanishes or not. Note that moment coefficient of skewness, $E[x^3(n)]/E[x^2(n)]^{3/2}$, in the classical Gaussianity test examines Gaussianity of $x(n)$ only for zero lags in the third-order moment. If the autobispectrum of the input vanishes, the linear and quadratic transfer functions are obtained separately from (9) and (10), respectively (i.e., orthogonal model). However, this test does not check Gaussianity of the fourth-order spectral moment of the input signal, which is assumed to be expressed in terms of second-order spectral moments when the input signal is Gaussian. Thus, the Hinich test may be regarded as a necessary but not sufficient test for Gaussianity in quadratic nonlinear system identification. Gaussianity tests based on the trispectrum, which is the next higher order spectrum after bispectrum, are currently being investigated [17]. For cubic nonlinear system identification Gaussianity of the input signal should be tested up to the sixth-order moment.

IV. Nonlinear Coherence Analysis

For non-Gaussian inputs, one has to solve (9) and (10) simultaneously with respect to $H_1(m)$ and $H_2(i, j)$. Recently, Kim and Powers [11] developed the estimation technique of $H_1(m)$ and $H_2(i, j)$ for the general input case by expressing (7) in vector notation and solving the equation under the minimum mean square criterion as follows:

$$Y(m)_{\text{model}} = \boldsymbol{H}(m) \boldsymbol{X}(m) \tag{17}$$

where the boldface letters denote a vector quantity, and

$$\boldsymbol{H}(m) = [H_1(m), \boldsymbol{H}_2(m)] \tag{18}$$

$$\boldsymbol{H}_2(m) = \begin{cases} \left[2H_2\left(\frac{m+1}{2}, \frac{m-1}{2}\right), \right. \\ 2H_2\left(\frac{m+3}{2}, \frac{m-3}{2}\right), \cdots, \\ \left. 2H_2\left(\frac{N}{2}, m - \frac{N}{2}\right)\right], \\ \quad \text{for odd } m \\ \left[H_2\left(\frac{m}{2}, \frac{m}{2}\right), 2H_2\left(\frac{m}{2}+1, \frac{m}{2}-1\right), \cdots, \right. \\ \left. 2H_2\left(\frac{N}{2}, m - \frac{N}{2}\right)\right], \\ \quad \text{for even } m \end{cases} \tag{19}$$

where N is the number of sampled data per realization. Furthermore,

$$\boldsymbol{X}^T(m) = [X_1(m), \boldsymbol{X}_2^T(m)] \tag{20}$$

$$\boldsymbol{X}_2^T(m) = \begin{cases} \left[X\left(\frac{m+1}{2}\right) X\left(\frac{m-1}{2}\right), \right. \\ X\left(\frac{m+3}{2}\right) X\left(\frac{m-3}{2}\right), \cdots, \\ \left. X\left(\frac{N}{2}\right) X\left(m - \frac{N}{2}\right)\right], \\ \quad \text{for odd } m, \\ \left[X\left(\frac{m}{2}\right) X\left(\frac{m}{2}\right), \right. \\ X\left(\frac{m}{2}+1\right) X\left(\frac{m}{2}-1\right), \cdots, \\ \left. X\left(\frac{N}{2}\right) X\left(m - \frac{N}{2}\right)\right], \\ \quad \text{for even } m \end{cases} \tag{21}$$

where ^{T} denotes the transpose of a matrix. Then the optimum solution for $\boldsymbol{H}(m)$ in (17), under the minimum mean-square-error criterion, is given by

$$\boldsymbol{H}(m) = E[\boldsymbol{X}^*(m)\boldsymbol{X}^T(m)]^{-1}E[\boldsymbol{X}^*(m)Y(m)]. \quad (22)$$

The linear and quadratic transfer functions $H_1(m)$ and $H_2(i, j)$ estimated by (22) can be used to determine the second-order harmonic distortion factor D_{H2} at frequency $m\Delta f$ by (6) and the second-order factor of intermodulation distortion D_{IM2} at frequency $(i + j)\Delta f$ by (4). In the sinusoidal input approach, the harmonic and intermodulation distortions over the band of frequencies of interest are measured by repeating the same experiment at various frequencies and by changing frequencies of oscillators and filters, which is a tedious and time-consuming job. With broad-band random input the distortions can be estimated simultaneously over the whole frequency band of interest as long as the input has a broad (nonzero) spectrum over the band of frequencies of interest and the data set is large enough to replace the expectation operator in (22) by an ensemble average over a sufficient number of sample spectra.

Another important advantage of the random input approach is that the nonlinear distortion can be quantified in terms of coherence spectra, which is the principal focus of this paper. In a single-channel linear system, it is well known that the coherence spectrum $\gamma^2(f)$, which is a *normalized crosspower spectrum*, is defined as

$$\gamma^2(f) = \frac{|E[|X^*(f)Y(f)]|^2}{E[|X(f)|^2]E[|Y(f)|^2]} = \frac{|S_{yx}(f)|^2}{S_{xx}(f)S_{yy}(f)}. \quad (23)$$

Since $|S_{yx}(f)|^2 \le S_{xx}(f)S_{yy}(f)$, the coherence spectrum is bounded, i.e., $0 \le \gamma^2(f) \le 1$. Statistically speaking, the coherence spectrum is a measure of the degree of linear correlation (or linear relationship) between output and input as a function of frequency. The ordinary concept of the coherence function is that $X(f)$ and $Y(f)$ are perfectly correlated at the frequency f when $\gamma^2(f) = 1$, and they are completely uncorrelated when $\gamma^2(f) = 0$. However, this interpretation is valid only for a linear system. The term $\gamma^2(f)$ defined in (23) represents only the linear relationship between the input $X(f)$ at frequency f and the output $Y(f)$ at the same given frequency. Bendat and Piersol [19] have pointed out that a reduction of coherence may be due to a) extraneous noise in the measurements, b) a nonlinear relationship between input and output, and/or c) multiple relations to other quantities. Thus, when more than one frequency component of the input couple and interact in a nonlinear system to produce components with new frequencies in the output, the coherence spectrum defined in (23) does not represent the proper relationship between the input and output.

A useful interpretation of the ordinary coherence spectrum, which serves as a basis to generalize the concept, is that it is the ratio of the model output "power" (in the sense of mean-squared value per unit frequency) to the true output power. For example, one can derive a discrete version of (23) from the above interpretation as follows:

$$\gamma^2(m) = \frac{E[|Y(m)|^2]_{\text{model}}}{E[|Y(m)|^2]_{\text{true}}} = \frac{|H_1(m)|^2E[|X(m)|^2]}{E[|Y(m)|^2]} = \frac{|S_{yx}(m)|^2}{S_{xx}(m)S_{yy}(m)} \quad (24)$$

where the expression for $H_1(m)$ in (11) is utilized. Thus, one can interpret $\gamma^2(m)$ as the fraction of output power which can be accounted for by the linear model and $1 - \gamma^2(m)$ is the fraction of output power which cannot be accounted for by the linear model.

The coherence spectrum can be generalized using the above definition, for the second-order Volterra system of (7) in the frequency domain [11], [20]. By using the estimates of $H_1(m)$ and $H_2(i, j)$ computed by (22), the model output power spectrum $S_{yy}(m)_{\text{model}}$ can be expressed as follows:

$$\begin{aligned} S_{yy}(m)_{\text{model}} &= E[|Y(m)|^2]_{\text{model}} \\ &= E[|Y_L(m)|^2] + E[|Y_Q(m)|^2] \\ &\quad + 2\,\text{Re}\,E[Y_L(m)Y_Q(m)^*] \\ &= S_L(m) + S_Q(m) + S_{LQ}(m) \end{aligned} \quad (25)$$

where $Y_L(m)$ and $Y_Q(m)$ denote the first two terms on the RHS of (7), respectively, and where

$$S_L(m) = |H_1(m)|^2E[|X(m)|^2] \quad (26)$$

$$S_Q(m) = E\left[\left|\sum_i H_2(i, m - i)X(i)X(m - i)\right|^2\right] \quad (27)$$

$$S_{LQ}(m) = 2\,\text{Re}\left\{H_1(m)\sum_i H_2^*(i, m - i)E[X^*(i) \cdot X^*(m - i)X(m)]\right\}. \quad (28)$$

$S_L(m)$ is the linear contribution to the model output power at frequency $m\Delta f$, $S_Q(m)$ is the quadratic nonlinear contribution to the model output power, and $S_{LQ}(m)$ is the "interference part" of the model output power. For a perfect Gaussian input, $S_{LQ}(m)$ is zero because the autobispectrum $E[X^*(i)X^*(m - i)X(m)]$ is zero and there is no coupling between the linear and quadratic outputs (i.e., the model is orthogonal).

Dividing both sides of (25) by the actual output power spectrum $S_{yy}(m)$, we can obtain the generalized coherence spectrum, which characterizes the "goodness of fit" of the model, as follows:

$$\begin{aligned} \gamma^2(m) &= \frac{E[|Y(m)|^2]_{\text{model}}}{E[|Y(m)|^2]_{\text{true}}} = \frac{E[|\hat{Y}(m)|^2]}{E[|Y(m)|^2]} \\ &= \gamma_L^2(m) + \gamma_Q^2(m) + \gamma_{LQ}^2(m) \end{aligned} \quad (29)$$

where

$$\gamma_L^2(m) = \frac{S_L(m)}{S_{yy}(m)} \quad (30)$$

$$\gamma_Q^2(m) = \frac{S_Q(m)}{S_{yy}(m)} \tag{31}$$

$$\gamma_{LQ}^2(m) = \frac{S_{LQ}(m)}{S_{yy}(m)} \tag{32}$$

where the terms $\gamma_L^2(m)$, $\gamma_Q^2(m)$, and $\gamma_{LQ}^2(m)$ are the linear coherence spectrum, quadratic coherence spectrum, and coupled coherence function of the second-order Volterra model, respectively. $\gamma_L^2(m)$ and $\gamma_Q^2(m)$ are the fractional contributions to the output power at frequency $m\Delta f$ due to linear transfer function $H_1(m)$ and quadratic transfer function $H_2(i, j)$, respectively. As discussed above, $\gamma_{LQ}^2(m)$ is nonzero for non-Gaussian inputs, and denotes the fraction of the model output power at frequency $m\Delta f$ due to the "interference" of the output of the linear and quadratic transfer functions. The terms $\gamma_L^2(m)$ and $\gamma_Q^2(m)$ are positive, but not bounded by unity since $\gamma_{LQ}^2(m)$ can be positive or negative depending on the phase relationship between the linear response and quadratic response. However, the total coherence $\gamma^2(m)$ at each m is bounded by zero and unity. Possible reasons for coherence loss, $1 - \gamma^2(m)$, are a) the neglect of higher order (third or higher) terms in the model, b) additive noise in the measurements, and c) extraneous inputs.

For the Gaussian input case $S_{LQ}(m)$ and, hence, $\gamma_{LQ}^2(m)$ are zero since the linear and quadratic outputs $Y_L(m)$ and $Y_Q(m)$ are orthogonal. In this case one can obtain $\gamma_L^2(m)$ and $\gamma_Q^2(m)$ using (11), (12), (26), (27), (30), and (31) as follows:

$$\gamma_L^2(m) = \frac{|S_{xx}(m)|^2}{S_{xx}(m)S_{yy}(m)} \tag{33}$$

$$\gamma_Q^2(m) = \frac{1}{2}\sum_i \frac{|B_{xxx}(i, m-i)|^2}{S_{xx}(i)S_{xx}(m-i)S_{yy}(m)}. \tag{34}$$

Notice that $\gamma_L^2(m)$ in (33) is identical with (24) for a linear model, and $\gamma_Q^2(m)$ in (34) is identical with the coherence spectrum derived from a purely quadratic model [7]. That is, the generalized system coherence spectra defined by (30)–(32) with (26)–(28) reduces to the ordinary coherence spectra given by (33) and (34) when the input excitation is zero-mean Gaussian.

V. Bichromatic Versus Random Process Estimates

In the previous section, the random input approach to estimate second-order harmonic and intermodulation distortion of the component under test is described. The measurement of nonlinear distortions, however, has traditionally been conducted by supplying sinusoidal signals and measuring the extent to which harmonic or combination tones are generated in the output of the component under test. In order to calculate the response of the quadratically nonlinear system subject to two sinusoidal inputs with frequency f_1 and f_2 ($f_1 > f_2$), let the input be

$$x(t) = \sum_{i=1}^{2} |V_i| \cos [2\pi f_i t + \theta_i] \tag{35}$$

which can also be expressed as

$$x(t) = \tfrac{1}{2}\sum_{i=-2}^{2} V_i \exp [j2\pi f_i t] \tag{36}$$

where

$$V_{i-1} = V_i^*, \qquad V_{i0} = 0, \qquad f_{-i} = -f_i.$$

The linear response to the inputs specified by (35) is given by [18]

$$y_1(t) = \sum_{i=1}^{2} |V_i|\,|H_1(f_i)| \cos [2\pi f_i t + \theta_i + \phi_1(f_i)]. \tag{37}$$

The quadratic response to the input specified by (36) is given by

$$y_2(t) = \tfrac{1}{2}\sum_{i_1=-2}^{2}\sum_{i_2=-2}^{2} V_{i_1}V_{i_2}H_2(f_{i_1}, f_{i_2}) \cdot \exp [j2\pi(f_{i_1} + f_{i_2})t]. \tag{38}$$

Since $H_2(-f_1, -f_2) = H_2^*(f_1, f_2)$ and $H_2(f_1, f_2) = H_2(f_2, f_1)$ in real systems, (38) can be expressed as

$$\begin{aligned} y_2(t) = {} & \sum_{i=1}^{2} |V_i|^2 |H_2(f_i, f_i)| \\ & \cdot \cos [2\pi(2f_i)t + 2\phi_2(f_i, f_i)] \\ & + 2|V_{i_1}|\,|V_{i_2}|\,|H_2(f_1, f_2)| \\ & \cdot \cos [2\pi(f_1 + f_2)t + \theta_1 + \theta_2 + \phi_2(f_1, f_2)] \\ & + 2|V_{i_1}|\,|V_{i_2}|\,|H_2(f_1, -f_2)| \\ & \cdot \cos [2\pi(f_1 - f_2)t + \theta_1 - \theta_2 + \phi_2(f_1, -f_2)] \\ & + \sum_{i=1}^{2} |V_i|^2 H_2(f_i, -f_i). \end{aligned} \tag{39}$$

The first term in (39) denotes second-order harmonic distortions, the second term denotes an intermodulation distortion at the sum frequency, the third term denotes an intermodulation distortion at the difference frequency, and the last term denotes the DC component. Note that the second and third terms accounting for intermodulation distortion contain the factor of two while the first term accounting for harmonic distortion does not. Note also that the output at the difference frequency, $f_1 - f_2$, contains not only the intermodulation distortion at that frequency but also the second harmonic distortion due to frequency f_2 if $f_1 - f_2 = 2f_2$. These harmonic and intermodulation distortions (or $H_2(f_1, -f_2)$ and $H_2(f_2, f_2)$) can be identified by supplying only one sinusoidal signal first to measure harmonic distortion (or $H_2(f_2, f_2)$) and by subtracting the harmonic distortion from the total distortion subject to two sinusoidal signals to obtain intermodulation distortion. One can also see that the output at the difference frequency $f_1 - f_2$ contains the intermodulation distortion at that frequency as well as linear

response at frequency f_2 if $f_1 - f_2 = f_2$. The intermodulation distortion (or $H_2(f_1, -f_2)$) can be obtained by supplying only one sinusoidal signal first to measure linear response $y_1(t)$ and by subtracting $y_1(t)$ from the total response corresponding to the two sinusoidal signals.

In order to compare the two-tone input method with random input method, a known system is used in a simulation experiment as follows [11]:

$$y(t) = -0.7x(t) - 0.9x(t-2) + 0.8x^2(t) + 0.5x^2(t-2). \quad (40)$$

The quadratic transfer function of the above system is given by

$$H_2(f_i, f_j) = 0.8 + 0.5 \exp[-j4\pi(f_i + f_j)]. \quad (41)$$

In Fig. 1 we present an amplitude and a phase of $H_2(f_i, f_j)$ estimated by the two-tone input method, discussed in this section; i.e., by changing frequencies of two sinusoidal signals and measuring the reponse at the sum and difference frequencies. Note that it is enough to plot only the upper and lower triangular region of $H_2(f_i, f_j)$ in the f_1 and f_2 plane because of various symmetry properties. Values in the upper triangular region, where $f_1 \geq 0, f_2 \geq 0, f_1 \geq f_2$, represent distortions at sum frequencies, while values at the lower triangular region, where $f_1 \geq 0, f_2 \leq 0, f_1 \geq |f_2|$, represent distortions at difference frequencies. The values located on the 45° line in the upper triangular region correspond to harmonic distortion while the others correspond to intermodulation distortions. The valley between those two regions exists since distortions due to a sinusoidal signal interacting with DC component are not calculated and assigned a value of zero.

Fig. 2 shows the amplitude and phase of $H_2(f_i, f_j)$ estimated by the random input method described in Section IV. For this simulation the output data are generated by applying a zero-mean exponential input (IMSL GGEXN) to the known system (40). The input and output data records were divided up into 2000 segments of 64 samples each. Each data segment is Fourier transformed via the FFT to calculate various sample spectra in (22) and averaged to obtain the spectral estimates. As expected, the quadratic transfer function, accounting for second-order distortion, estimated by the two-tone input method shows good agreement with the one estimated by the random input method except for small fluctuations. Once the transfer functions are estimated, distortion factors can be calculated by substituting the estimated transfer functions into (4) and (6). For example, D_{H2} at frequency $5\Delta f$ ($\Delta f = 1/64$ Hz) is -11.3 (-11.1) dB, D_{IM2} at $8\Delta f - 7\Delta f$ is 1.1 (1.7) dB, and D_{IM2} at $8\Delta f + 7\Delta f$ is -10.9 (-10.5) dB (the values in parenthesis are estimated by the random input method and are preceded by the values estimated by the two-tone input method). Traditionally, distortion factors are measured by (3) and (5) with sinusoidal signals, which give the same results obtained by (4) and (6) with sinusoidal signals.

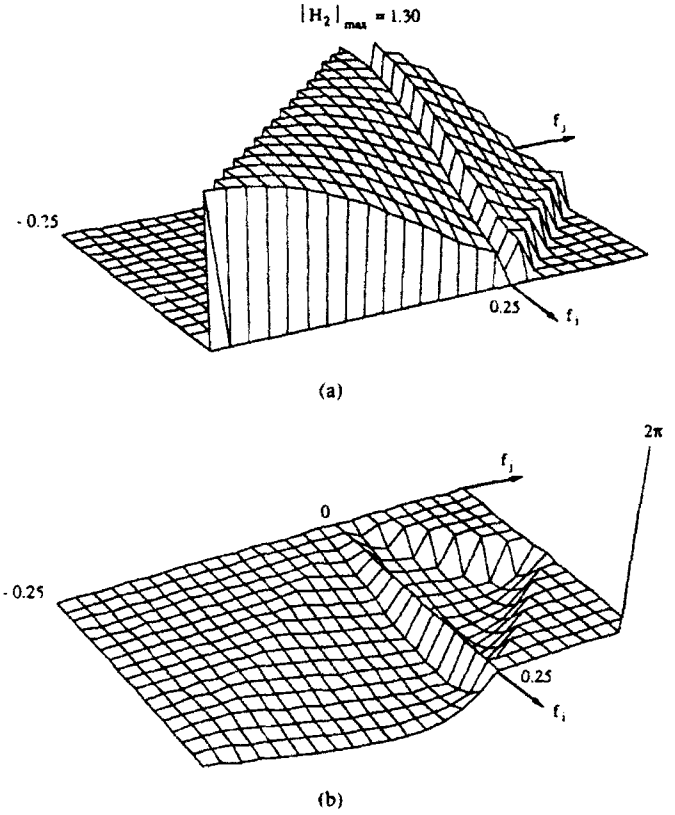


Fig. 1. Amplitude and phase of quadratic transfer function $H_2(i, j)$ (see (41)) estimated by two-tone input method: (a) estimated amplitude, (b) estimated phase.

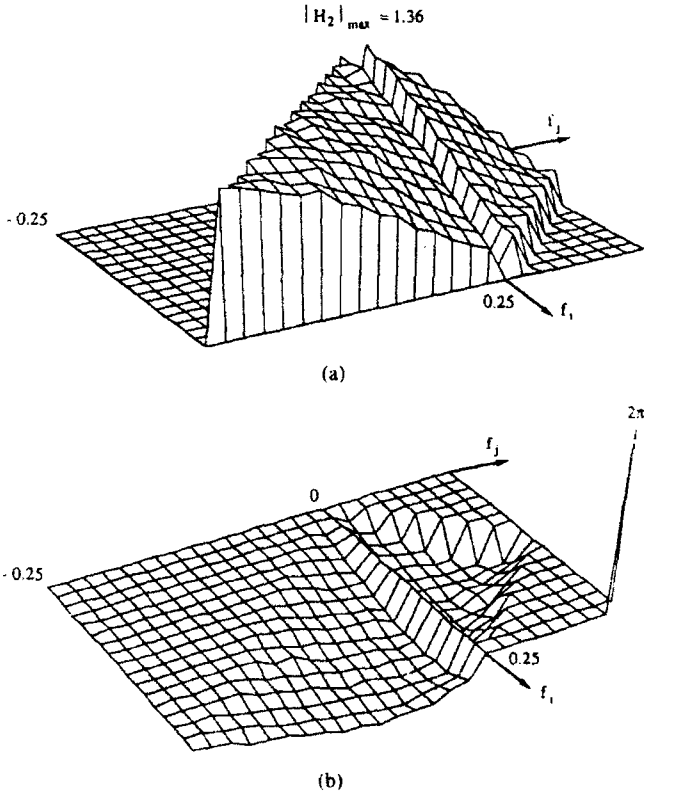


Fig. 2. Amplitude and phase quadratic transfer function $H_2(i, j)$ estimated by random input method: (a) estimated amplitude, (b) estimated phase.

VI. Testing for Gaussianity and Experimental Results

A. Experimental Setup

Loudspeakers tend to be the weakest component in most high-fidelity sound-reproducing systems. The nonlinear distortion of the loudspeaker is typically a maximum at the lower frequencies. The principal cause for nonlinear distortion at the lower frequencies is nonuniform Bl product versus the excursion of the voice coil. Also, nonlinearities in the compliance of the suspension and surround are not negligible. The prominent nonlinearities are: a) the Bl product, b) the self-inductance of the voice coil, and c) the nonlinear compliance of the suspension and surround [21]. a) and b) may result from a nonuniform flux distribution.

The canonical equation for an antireciprocal transducer at low frequencies can be given by the following when air loading is neglected:

$$Bli = m\ddot{x} + R_m\dot{x} + kx \tag{42}$$

$$V_g = R_e I + \frac{d(L_e i)}{dt} + Bl\dot{x} \tag{43}$$

where Bl is the force factor, i is the voice-coil current, x is the voice-coil excursion, and m, R_m, and k are the moving mass, the mechanical damping, and the stiffness of the mass-spring system. V_g is the generator voltage, R_e and L_e are the electrical resistance and inductance of the voice coil.

Kaizer [21] modeled the loudspeaker in the low-frequency band by a Volterra series expansion after approximating Bl, k, and L_e in (42) and (43) by a truncated power series as follows:

$$Bl = Bl_0 + b_1 x + b_2 x^2 \tag{44}$$

$$k = k_0 + k_1 x + k_2 x^2 \tag{45}$$

$$L_e = L_{e0} + l_1 x + l_2 x^2. \tag{46}$$

The amount of distortion will depend on many design parameters, such as magnet size, voice-coil design, suspension, etc. Also, it depends on the input power level. As pointed out by Greiner and Sims [22], distortion can increase dramatically, for input power above a few watts. In general, minimizing cone travel by using large-diameter high-efficiency drivers tends to reduce distortion.

To demonstrate the feasibility and practicality of this paper's proposed approach, the second-order Volterra series has been utilized to model, identify, and quantify the linear response and the second-order distortion product of loudspeakers in the low-frequency region. The arrangement of Fig. 3 was used to measure the distortion performance of the compression driver typical of those used to drive acoustic horns. The noise source used puts out a flat noise spectrum band limited to 500 Hz. A low-pass filter with a cutoff frequency of 420 Hz was inserted before the amplifier which was adjusted to drive the compression driver at 10 W. The compression driver was placed in an anechoic duct with a probe (condenser microphone cartridge). The microphone power supply with a probe and amplifier used in this experiment is assumed linear over the band of interest (10–500 Hz) from the specifications provided by the manufacturer. A digital signal processor (DSP 56000) took 98 304 data samples from the input and output signal at a 1-kHz sampling rate. These were divided up into 384 segments of 128 samples each. In order to compute various sample spectra appearing in (22), each data segment is Fourier transformed via the FFT, and then the 384 sample spectra were averaged to get the spectral estimates. The frequency resolution Δf is, therefore, 500 Hz ÷ 64 samples = 7.8125 Hz.

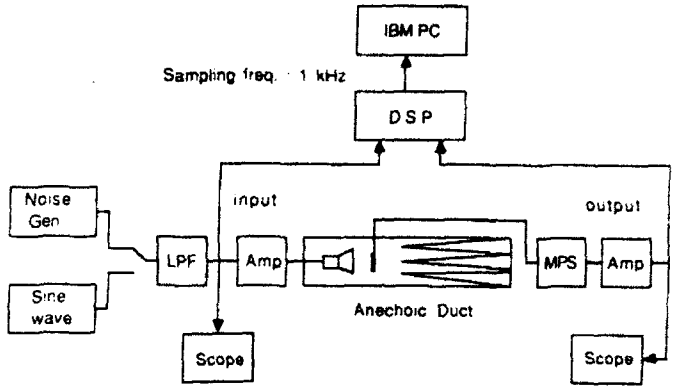


Fig. 3. Block diagram of distortion analysis setup. Digital signal processor: DSP56000, Motorola; Low-pass filter: Krohn-hite mode 150AR (cutoff frequency = 420 Hz); Probe condenser microphone cartridge: type 4133 Band K; Microphone power supply: type 2804, Band K.

B. Testing for Gaussianity

In this section, we test the noise generator for Gaussianity. First, we examine the Gaussian hypothesis by performing two types of tests. The first of these is the moment test of univariate marginals as described by Pearson [14]. On the basis of 5000 samples from the noise generator waveform, the following values are found:

$$\text{skewness } \sqrt{\beta_1} = -0.0123; \qquad \text{kurtosis } \beta_2 = 3.11934.$$

According to the table provided by Pearson the hypothesis should be rejected at 0.05 level of significance if $\sqrt{\beta_1}$ lies outside the range -0.057 to 0.057, and at 0.01 level of significance if $\sqrt{\beta_1}$ lies outside the range -0.081 to 0.081. The observed value lies inside the limit. Also, the hypothesis should be rejected at 0.05 level of significance if β_2 lies outside the range 2.89 to 3.12, and at 0.01 level of significance if β_2 lies outside the range 2.85 to 3.17. The observed value lies inside the limit. Thus, the Gaussian hypothesis is *not* rejected on the basis of this test.

The histogram of the noise generator estimated from the same data set (5000 samples) is shown in Fig. 4. One may assume that the signal from the noise generator is nearly Gaussian from the reasonably bell-shaped histogram with a skewness of -0.0123. Thus, when one deals with a linear system or univariate marginal distribution of a signal, the noise generator output can often be used in experiments as a Gaussian source. However, as was

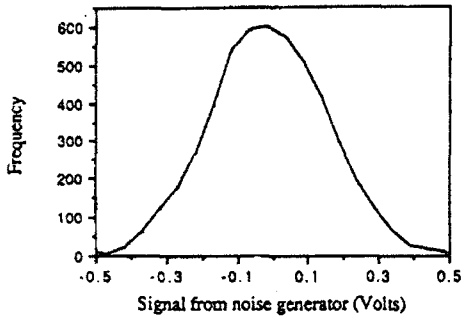


Fig. 4. Histogram of the signal from the noise generator shown in Fig. 3 (variance = 0.026, skewness = −0.0123, kurtosis = 3.119).

pointed out by Brockett *et al.* [16], the univariate marginal distribution of the time series may satisfy Gaussianity tests even if the joint distribution of several time points is non-Gaussian. Thus, if one wants to analyze the nonlinearity of the system by applying the noise generator to a nonlinear system as a Gaussian input signal, the Gaussianity statistics should be carefully investigated beforehand, since the univariate marginal distribution of the input signal may pass classical statistical tests for Gaussianity without the signal being sufficiently Gaussian for nonlinear system identification purposes. If the input signal is actually non-Gaussian, the Gaussian assumption may lead to incorrect estimates of the Volterra kernels, since the higher order spectra used to estimate the kernels are sensitive to departures from Gaussianity.

Since we want to estimate the linear and quadratic transfer function of the loudspeaker by applying a noise generator as an input source, the Hinich test is used to examine the Gaussianity of the input signal. To become familiar with Hinich's Gaussianity statistic value, first we test the Gaussianity of two known signals: the first is a zero-mean Gaussian generated by IMSL GGNPM, and the second is a zero-mean exponential input generated by IMSL GGEXN. The Gaussianity statistic value (standardized value of the statistic T in (16)) for the Gaussian signal is −1.36, which is not large enough to reject Gaussianity [16]. The Gaussianity statistic value for the exponential signal is 75.55, which is large enough to reject the Gaussianity hypothesis, as we expected.

Next, the Gaussianity of the signal from the noise generator is tested. The resulting statistic value by the Hinich test is 102.10 (49 152 samples). All Gaussianity statistic values for the noise generator data records generated under different conditions (different sample points, different power, different cutoff frequency) were large enough to reject the Gaussianity hypothesis. Thus, one can see that the signal from the noise generator cannot be used as a Gaussian source for quadratic system analysis. It can, however, be used to identify the quadratic system if we utilize the "general input" approach described in Section IV.

C. Experimental Results

Since the input signal from the noise generator to the nonlinear loudspeaker system is not sufficiently Gaussian,

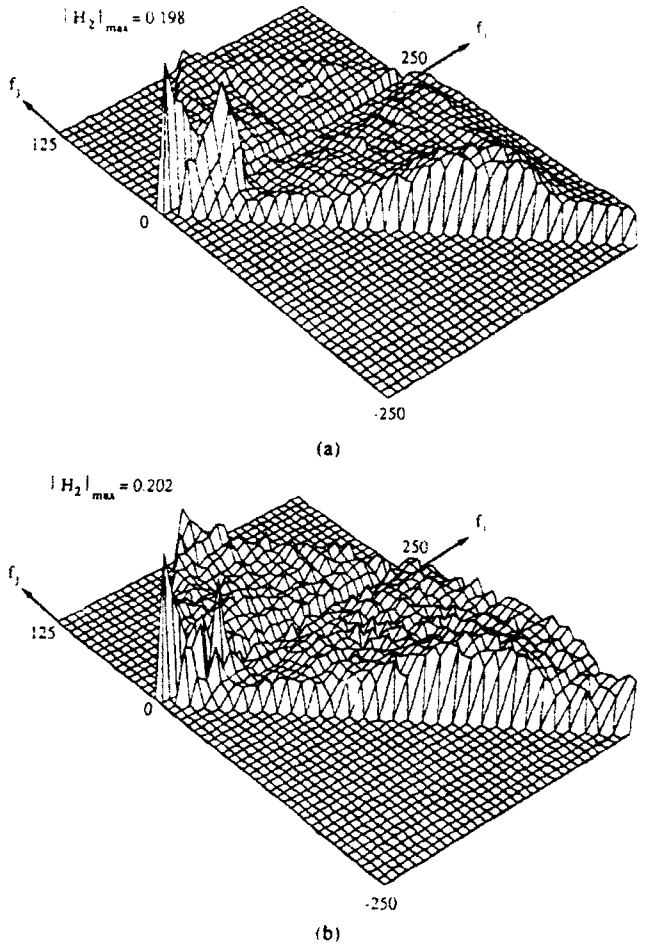


Fig. 5. Amplitude of the loudspeaker quadratic transfer function $H_2(i, j)$ experimentally estimated using: (a) two-tone input method, (b) random input method.

one has to solve (9) and (10) simultaneously for $H_1(m)$ and $H_2(i, j)$. The linear and quadratic transfer functions are estimated according the solution given by (22). The estimated $H_2(i, j)$ (smoothed over adjacent frequencies), accounting for second-order distortion, is shown in Fig. 5(b). To compare the proposed technique with the conventional analog two-tone method, $H_2(i, j)$ is measured by the method described in Section V. Two oscillators, both of them having the same amplitude, are applied to the amplifier. The output of the amplifier, which generates a sinusoidal signal at 10 W, is connected to the component under test to calculate the distortions at sum frequencies and difference frequencies by (39). $H_2(i, j)$, in Fig. 5(a), measured by the two-tone input method, shows good agreement with the one obtained using the random input approach except for relatively small fluctuations. Note the large values along the diagonal line in the difference interaction zone, which suggests that difference frequency generation dominates the distortion process.

Distortion factors of the loudspeaker can be calculated by substituting the estimated transfer functions into (4) and (6). Fig. 6 shows the results for D_{IM2} at $f_1 + f_2$ for a fixed frequency $f_2 (= 1\Delta f)$ as a function of frequency f_1. Reasonable agreement between the two-tone input method

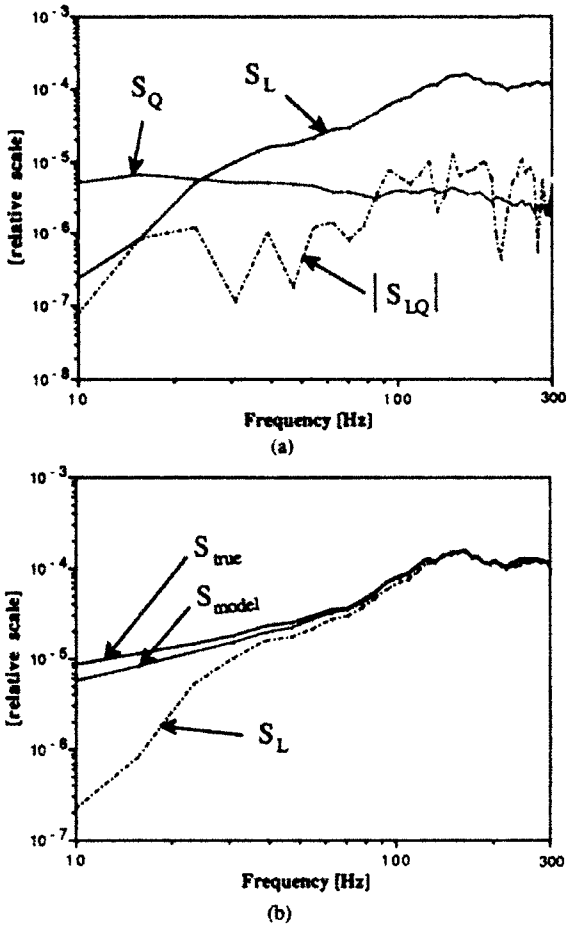


Fig. 8. Actual and model output power spectrum of loudspeaker with noise generator input: (a) linear part of model output power spectrum S_L, quadratic part of model output power spectrum S_Q, and interference part (absolute value) of output power model spectrum S_{LQ}, (b) measured output power spectrum S_{true}, total model output power spectrum $S_{model} = S_L + S_Q + S_{LQ}$, and linear part of model output power spectrum S_L.

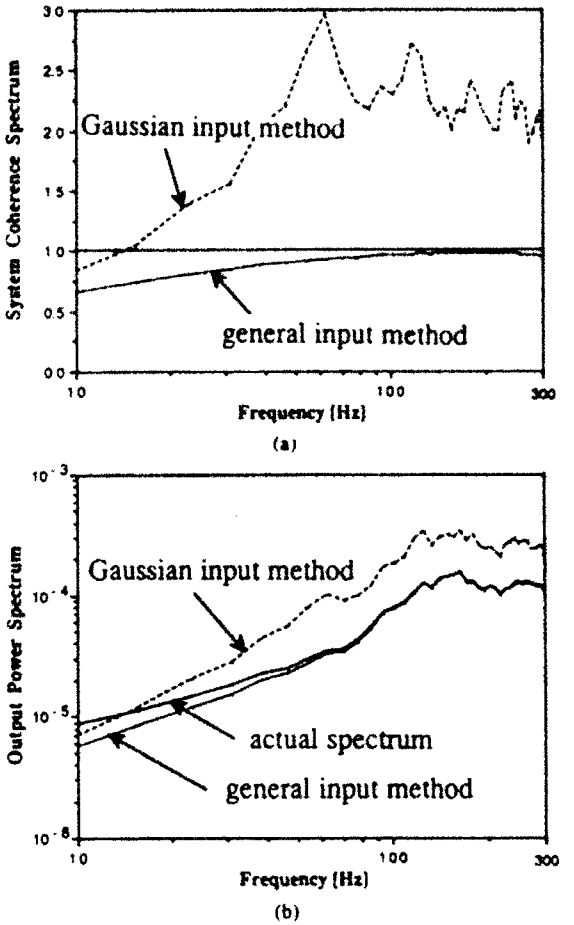


Fig. 9. System coherence spectra and output power spectra of loudspeaker: (a) system coherence spectra for loudspeaker Volterra models calculated using general-input method and assumed-Gaussian input method. (b) output power spectrum of loudspeaker: actual (observed) spectrum and output power spectra calculated from Volterra models determined using the general-input method and assumed-Gaussian input method.

Gaussian excitation case. The system coherence calculated for the assumed Gaussian input method is plotted in Fig. 9(a) and we note that except at low frequencies it is considerably greater than unity. In fact, this plot indicates that the Volterra model calculated using (11) and (12) overestimates the response of the loudspeaker by up to a factor of 3. On the other hand, as shown in Fig. 9(a), the Volterra model calculated on the basis of (22) and valid for general inputs (non-Gaussian as well as Gaussian) yields a system coherence which is close but somewhat less than unity over the entire frequency band. In Fig. 9(b) we plot the actual power spectrum of the loudspeaker response when driven by the noise source, and the responses "predicted" by the Volterra models estimated using the assumed Gaussian input method and the general input method. Again we see that the general input method "predicts" a response which is in much closer agreement with that which is actually observed. On the other hand, the assumed Gaussian input method overpredicts the response over the vast majority of the frequency spectrum. In summary, Fig. 9 graphically illustrates, by way of example, the deleterious effects of assuming Gaussian excitation when, in fact, it is not.

VII. Summary

In this paper, a digital higher order spectral method to evaluate the second-order distortion of a nonlinear system, which can be represented by Volterra kernels up to second order and which is subjected to a random input, was described. The common assumption that the input possesses Gaussian statistics was investigated. The preliminary results indicate that one must carefully check the degree of Gaussianity of random signal generators used in experimental nonlinear system identification studies. Specifically, we have shown that it is not sufficient to demonstrate that the marginal distribution is approximately Gaussian. In view of the fundamental role the bispectrum plays in the system identification of second-order system, the Hinich test, which makes use of the statistical properties of the bispectrum, appears to be particularly appropriate.

After checking the Gaussianity of the noise generator and demonstrating that it is not sufficiently Gaussian, the general input method was utilized to estimate the Volterra linear and quadratic transfer function of a nonlinear system. Then, the quadratic transfer function accounting for second-order distortion was estimated by applying (bi-

chromatic) sinusoidal signals to the system to compare the random input method with the classical two-tone input method. Relatively good agreement was obtained for both simulation and experiment. Finally, the low-frequency loudspeaker distortion is quantified in terms of nonlinear coherence spectra and the relative nonlinear contribution to the output power.

In the future, the random-input method of this paper will be extended to handle cubic distortion. Furthermore, the technique might ultimately be used to construct a nonlinear inversion circuit for the purposes of compensating distortion. Because of the generality of the Volterra approach, the approach of this paper can be utilized to quantify distortion in other components of an audio system. Moreover, it can be applied to measure distortion introduced by any type of system, including all types of communication systems.

REFERENCES

[1] A. M. Rudkin, *Electronic Test Equipment*. London: Granada, 1981.

[2] F. F. Mazda, *Electronic Instruments and Measurement Techniques*. Cambridge: Cambridge University Press, 1987.

[3] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear System*. New York: Wiley, 1980.

[4] S. Narayanan and H. C. Poon, "An analysis of distortion in bipolar transistors using integral charge control model and Volterra series," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 341-351, 1973.

[5] A. Borys, "On intermodulation and harmonic distortion in single amplifier active filters," *J. Audio Eng. Soc.*, vol. 28, no. 10, pp. 706-712, Oct. 1980.

[6] L. J. Tick, "The estimation of transfer functions of quadratic systems," *Technometrics*, vol. 3, pp. 563-567, Nov. 1961.

[7] J. Y. Hong, Y. C. Kim, and E. J. Powers, "On modeling the nonlinear relationship between fluctuations with nonlinear transfer functions," *Proc. IEEE*, vol. 68, pp. 1026-1027, Aug. 1980.

[8] D. W. Choi, R. W. Miksad, E. J. Powers, and F. J. Fischer, "Application of digital cross-bispectral analysis techniques to model the nonlinear response of moored vessel system in random seas," *J. Sound Vibr.*, vol. 99, no. 3, pp. 309-326, 1985.

[9] R. E. Maurer and S. Narayanan, "Noise loading analysis of a third-order nonlinear system with memory," *IEEE Trans. Commun. Technol.*, vol. COM-16, no. 5, pp. 701-712, Oct. 1969.

[10] M. Hinich, "Testing for Gaussianity and linearity of a stationary time series," *J. Time Ser. Anal.*, vol. 3, no. 3, pp. 169-176, 1982.

[11] K. I. Kim and E. J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 11, pp. 1758-1769, Nov. 1988.

[12] Y. S. Cho, S. B. Kim, E. L. Hixson, and E. J. Powers, "Nonlinear distortion analysis using digital higher-order coherence spectra," in *Proc. Int. Conf. Acoust., Speech, Signal Processing '90*, vol. 2, 1990, pp. 1165-1168.

[13] Y. S. Cho, W. T. Oh, S. B. Kim, and E. J. Powers, "Testing for Gaussianity in nonlinear system identification," in *Proc. IEEE 1990 Int. Symp. Circuits Syst.*, vol. 2, 1990, pp. 1450-1453.

[14] E. S. Pearson, "A further development of tests for normality," *Biometrica*, vol. 22, pp. 139-149, 1930.

[15] W. J. Conover, *Practical Nonparametric Statistics*. New York: Wiley, 1971.

[16] P. L. Brockett, M. Hinich, and G. R. Wilson, "Nonlinear and non-Gaussian ocean noise," *J. Acoust. Soc. Amer.*, vol. 82, no. 4, pp. 1386-1394, Oct. 1987.

[17] J. W. DalleMolle and M. J. Hinich, "The trispectrum," in *Proc. Workshop Higher-Order Spectral Anal.* (Vail, CO), June 1989, pp. 68-72, 28-30.

[18] D. D. Weiner and J. E. Spina, *Sinusoidal Analysis and Modeling of Weakly Nonlinear Circuits*. New York: Van Nostrand Reinhold, 1980.

[19] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*. New York: Wiley-Interscience, 1971.

[20] Y. C. Kim, W. F. Wong, E. J. Powers, and J. R. Roth, "Extension of the coherence function to quadratic models," *Proc. IEEE*, vol. 67, pp. 428-429, 1979.

[21] A. J. M. Kaizer, "Modeling of the nonlinear response of an electrodynamic loudspeaker by a Volterra series expansion," *J. Audio Eng. Soc.*, vol. 35, no. 6, pp. 421-433, Jun. 1987.

[22] R. A. Greiner and T. M. Sims, Jr., "Loudspeaker distortion reduction," *J. Audio Eng. Soc.*, vol. 32, no. 12, pp. 956-963, Dec. 1984.



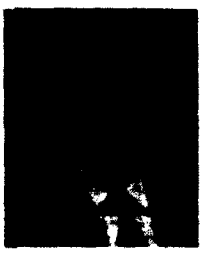
Yong Soo Cho received the B.S. degree in electronics engineering from Chung-Ang University, Seoul, Korea, and the M.S. degree in electronics engineering from Yonsei University, Seoul, Korea, in 1984 and 1987, respectively. He is currently working toward the Ph.D. degree in electrical engineering at the University of Texas at Austin.

During 1984 he was a Research Engineer at Goldstar Electrical Company, Osan, Korea. Since 1988 he has held a research position in the Electronics Research Center, the University of Texas at Austin, where he has engaged in research on time series analysis of random data. His research interests are mainly in signal processing including nonlinear system analysis using digital higher order spectra, nonlinear digital filtering, and estimation of nonstationary process.



Sung Bae Kim (S'87-M'87) received the B.S. and M.S. degress in nuclear engineering from the Seoul National University, Seoul, Korea, in 1975 and 1977, respectively, and the Ph.D. degree in electrical and computer engineering in 1986 from the University of Texas at Austin.

From 1975 to 1980 he was a Research Scientist in the Korea Atomic Energy Research Institute, Seoul, Korea. During this period, he was awarded an IAEA fellowship, and he conducted research on the pattern recognition of two-phase flows in the Instrumentation and Controls Division at the Oak Ridge National Laboratory, Oak Ridge, TN, from 1977 to 1978. In 1987, he joined the Electronics Research Center at the University of Texas at Austin. His research interests are concerned with utilizing nonlinear digital signal processing techniques in system modeling and analysis of nonlinear phenomena appearing in various physical media.



Elmer L. Hixson (S'48-A'50-M'55-SM'57-LS'90) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Texas in 1947, 1948, and 1960, respectively.

From 1948 to 1954 he was employed by the U.S. Navy Electronics Laboratory, San Diego, CA, where he worked in radio wave propagation and underwater acoustics. He then joined the faculty of the University of Texas where he is presently Professor of Electrical Engineering. From 1954 to 1964 he worked part time for the Applied Research Laboratory (ARL, UT). His professional interests are in electroacoustics, acoustic array processing, and noise control.



Edward J. Powers (S'57-M'58-SM'80-F'83) received the B.S. degree from Tufts University, Medford, MA, the M.S. degree from the Massachusetts Institute of Technology, Cambridge, and the Ph.D. degree from Stanford University, Stanford, CA, in 1957, 1959, and 1965, respectively, all in electrical engineering.

From 1959 to 1965 he was employed by Lockheed Missiles and Space Company in both Sunnyvale and Palo Alto, CA. At Lockheed his work was concerned with plasma physics and various electromagnetic aspects of reentry vehicles. In 1965 he joined the University of Texas at Austin, where he is currently the Texas Atomic Energy Research Foundation Professor, and Director of the Electronics Research Center. His primary professional interests lie in the innovative application of digital signal processing concepts, based on higher order statistics, to various problems involving nonlinear wave and turbulence phenomena.

A Digital Method of Modeling Two-Input Quadratic Systems with General Random Inputs

C. K. An, E. J. Powers, and C. P. Ritz

***Abstract*—A digital method of estimating linear and quadratic transfer functions of a two-input/multiple-output quadratically nonlinear system with general (i.e., non-Gaussian as well as Gaussian) random inputs is described. The approach is based on a frequency domain second-order Volterra functional series representation for a two-input/multiple-output system.**

I. Introduction

It is the objective of this correspondence to present a digital method of estimating, from the raw input and output time series data, the linear and quadratic transfer functions necessary to characterize a quadratically nonlinear system with two inputs and multiple outputs. The transfer functions are expressed in terms of various polyspectra (power spectra, bispectra, and fourth-order spectra) which, in turn, are digitally estimated from the raw input and output time series data. Since the approach is valid for general (e.g., non-Gaussian) inputs, Gaussian inputs are included as a special case. Our approach is based on a frequency domain second-order Volterra functional series representation [1] for a two-input/multiple-output system. The nonlinear transfer functions, i.e., the Fourier transforms of the Volterra kernels, measure the strength of the nonlinear coupling between various frequency components present in the input and output signals, and thus enable one to quantify the rate at which energy is exchanged between various modes of the nonlinear system.

Since the values of the higher order spectral moments used in determining the Volterra transfer functions are very sensitive to departures from Gaussianity, it is important to take such departures into account. Such departures may arise because of the past nonlinear history of the signal (e.g., as occurs in transition to turbulence [2]), or if the source of the exciting signal is not sufficiently Gaussian [3].

Since the approach closely follows that originally developed for single-input quadratically nonlinear systems, the reader is referred to [4] for specific details. Also see [5].

II. Estimation of Transfer Functions

In this correspondence we assume that the unknown nonlinear system is of second order, thus higher order terms may be neglected. Since our approach is based on a frequency domain second-order Volterra functional series representation (with the dc term subtracted), we will start from the following input-output relationship in the discrete frequency domain:

$$\begin{aligned}\hat{Y}(m) = {} & L_1(m)X_1(m) + L_2(m)X_2(m) \\ & + \sum_{i=1}^{M/4} 2Q_{11}(i, m-i)X_1(i)X_1(m-i) \\ & + \sum_{i=1}^{M/4} 2Q_{22}(i, m-i)X_2(i)X_2(m-i) \\ & + \sum_{i=1}^{M/4} Q_{12}(i, m-i)X_1(i)X_2(m-i) \\ & + \sum_{i=s}^{M/4} Q_{21}(i, m-i)X_2(i)X_1(m-i) \qquad (1)\end{aligned}$$

where $s = (m+1)/2$ and m is odd. For m even, similar expressions may be obtained. In Fig. 1 is shown the two-dimensional frequency plane over which the various quadratically nonlinear transfer functions are defined. Because of symmetry considerations it is necessary to determine the quadratic transfer functions in regions S_1 and D_1 only. Note that in order to avoid aliasing in the output of a quadratically nonlinear system we have assumed the input is band limited to $f_{sq}/4$, where f_{sq} denotes the sampling frequency of both the input and output. In (1) it is assumed that the number of data points in one segment is M, thus the upper limit in the summations is $M/4$, which, of course, corresponds to $f_{sq}/4$. Note, that a schematic interpretation of (1) is provided in Fig. 2.

We may rewrite (1) using vector notation as follows:

$$\hat{Y}(m) = x^t(m)h(m) \qquad (2)$$

where t denotes transpose, and where the elements of $x(m)$ and $h(m)$ are similar, although generalized for the two-input case, to those of [4].

Next, we solve for $h(m)$ in such a way as to minimize the mean-square error between the model output $\hat{Y}(m)$ and the actual observed output $Y(m)$. The solution is determined by solving the following matrix equation for $h(m)$ [4]:

$$E\{x^*(m)Y(m)\} = E\{x^*(m)x^t(m)\}h(m) \qquad (3)$$

Manuscript received August 28, 1989; revised August 31, 1990. This work was supported by the DoD Joint Services Electronics Program through the Air Force Office of Scientific Research (AFSC) under Contract F49620-89-C-0044 and in part by the Department of the Navy under Contract N00014-88-K-0638, and the Texas Advanced Technology Program.

The authors are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712-1084.

IEEE Log Number 9101859.

1053-587X/91/1000-2320$01.00 © 1991 IEEE

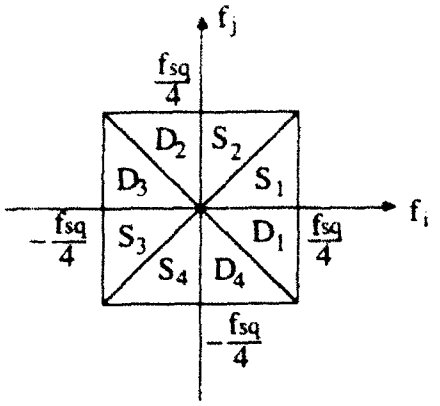


Fig. 1. Two-dimensional frequency domain in which quadratically nonlinear transfer functions are defined.

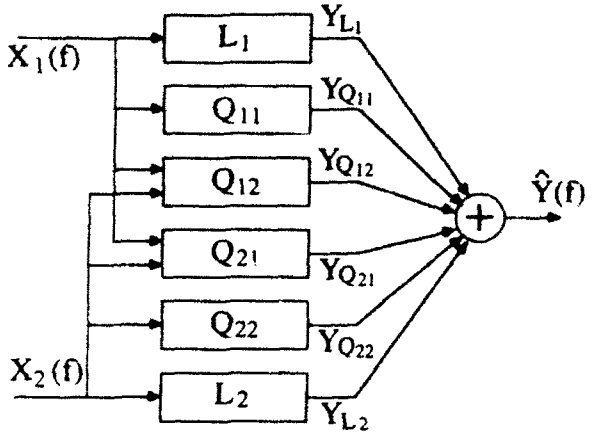


Fig. 2. Quadratically nonlinear system model with two inputs.

where $x^*(m)$ is the complex conjugate of $x(m)$. Equation (3) is a linear matrix equation in terms of the transfer function vector $h(m)$, consequently the solution for $h(m)$ is given by

$$h(m) = [E\{x^*(m)x^t(m)\}]^{-1}E\{x^*(m)Y(m)\} \tag{4}$$

if $E\{x^*(m)x^t(m)\}$ in (4) is not singular.

The LHS of (3) consists of cross-power spectra and cross bispectra between the two inputs and the output; and the matrix $E\{x^*(m)x^t(m)\}$ in (3) consists of autopower spectra, cross-power spectra, autobispectra, cross-bispectra, autofourth-order spectra and cross-fourth-order spectra of the two inputs. Since the bispectra and fourth-order spectra are multidimensional functions of frequency, the estimation is computationally very expensive. When the inputs are Gaussian, considerable simplification results in that the autobispectra are zero, and the fourth-order spectral moments can be expressed in terms of sums of products of second-order spectra (i.e., power spectra). For instance, when $M = 64$ and $m = 2$, for each realization only 162 matrix elements must be calculated under the zero-mean Gaussian assumption. However, the number of elements we must estimate for the non-Gaussian case is 2145 for each realization. Since 2016 elements of the 2145 elements are fourth-order spectra, almost all of the computational effort is spent to estimate the fourth-order spectra in the matrix $E\{x^*(m)x^t(m)\}$ in (3) when the inputs are not zero-mean Gaussian.

III. Analysis of Simulation Data

A. Known Quadratic System

The general input method described in the previous section has been applied to simulation data in order to demonstrate the feasibility and correctness of the approach. The discrete-time system used in the simulation is given as follows

$$\begin{aligned} y(n) = {} & -0.64x_1(n) + x_1(n-2) - 0.64x_2(n) + x_2(n-2) \\ & + 0.9x_1^2(n) + 0.5x_1^2(n-1) + 0.9x_2^2(n) + 0.5x_2^2(n-2) \\ & + 1.2x_1(n)x_2(n) + 0.6x_1(n)x_2(n-1) \\ & + 0.3x_1(n-1)x_2(n) + 1.5x_1(n-1)x_2(n-1). \end{aligned} \tag{5}$$

Therefore the linear and quadratic transfer functions are given by

$$L_1(f) = -0.64 + e^{-i4\pi f} \tag{6}$$

$$L_2(f) = -0.64 - e^{-i4\pi f} \tag{7}$$

$$Q_{11}(f_i, f_j) = 0.9 + 0.5e^{-i2\pi(f_i+f_j)} \tag{8}$$

$$Q_{22}(f_i, f_j) = 0.9 + 0.5e^{-i4\pi(f_i+f_j)} \tag{9}$$

$$Q_{12}(f_i, f_j) = 1.2 + 0.3e^{-i2\pi f_i} + 0.6e^{-i2\pi f_j} + 1.5e^{-i2\pi(f_i+f_j)} \tag{10}$$

$$Q_{21}(f_i, f_j) = 1.2 + 0.6e^{-i2\pi f_i} + 0.3e^{-i2\pi f_j} + 1.5e^{-i2\pi(f_i+f_j)} \tag{11}$$

For a quantitative measure of the quality of the estimates, the percent root mean square errors (%RMSE) of the estimated linear and quadratic transfer functions are defined as follows:

$$\%\text{RMSE}_l = \sqrt{\frac{1}{N_l}\sum_{m}\frac{|L(m) - \hat{L}(m)|^2}{|L(m)|^2}} \times 100\,[\%] \tag{12}$$

$$\%\text{RMSE}_q = \sqrt{\frac{1}{N_q}\sum_{i}\sum_{j}\frac{|Q(i,j) - \hat{Q}(i,j)|^2}{|Q(i,j)|^2}} \times 100\,[\%] \tag{13}$$

where $\hat{}$ signifies an estimated quantity, the absence of a hat denotes a "true" value and N_l is the number of frequency points over which the linear transfer function is calculated and N_q is the number of frequency points in the regions S_1 and D_1. Note that in each case the error is averaged over all relevant frequencies.

The approach developed in the previous section has been tested by utilizing exponentially distributed white input signals $x_1(n)$ and $x_2(n)$, which are partially correlated with each other. These signals are low-pass filtered with a cutoff frequency of a quarter ($f = 0.25$) of the sampling frequency of the input signals. Generated via (5) are 320 000 sample point data records for both the input and output signals. They are divided into 5000 segments of 64 data points each. The various higher order spectral moments needed in (4) are computed by ensemble averaging over the appropriate sample spectra, which are, in turn, estimated from discrete Fourier transforms of each segment. First, to illustrate deleterious effects of assuming Gaussian excitation, when it is in fact not, we estimate all the linear and quadratic transfer functions assuming Gaussian excitation and compute the %RMSE. Next we estimate the transfer functions, using the more general method of this paper.

When we use the assumed Gaussian input method, the average %RMSE of the two linear transfer functions is 40.3% and that of the four quadratic transfer functions is 43.4%. If we use the general input method for the same exponential input signals, the average %RMSE of the linear transfer functions drops to 3.4% and that of the quadratic transfer functions to 4.6%. Thus, the average %RMSE's of the transfer functions calculated assuming Gaussian inputs, when in fact they are not Gaussian, are approximately 10 times greater than those of the transfer functions estimated by the general input method.

With vectorization, the computation time of the general input method is about 34 seconds while computing on a CRAY X-MP/

24 supercomputer. This fact suggests that the general input method developed in this correspondence is highly vectorizable.

B. Output Power and Coherence Spectra

To further test the validity of the estimated transfer functions, the model output power spectrum is compared with the actual output power spectrum, the latter being computed from $y(n)$ in (5). For this purpose, new exponentially distributed input data, which were not used to estimate the model transfer functions, are applied in order to construct the model output. The actual and model output power spectra are shown in Fig. 3(a). The system coherence spectrum $\gamma_{system}(m)$ is defined by

$$\gamma_{system}(m) = \frac{\hat{P}_y(m)}{P_y(m)} \tag{14}$$

where $\hat{P}_y(m)$ is the model output power spectrum and $P_y(m)$ is the true output power spectrum for frequency m. Since the system coherence spectrum is the fraction of the actually observed output power spectrum which can be accounted for by the model, values near unity indicate a good model and, hence, good transfer function estimates. As shown in Fig. 3(b), the system coherence spectrum up to $f = 0.25$ is near unity, suggesting the estimated linear and quadratic transfer functions are good models of the two-input quadratically nonlinear system over this frequency range.

In order to quantify the relative "importance" of the various linear and nonlinear transfer functions appearing in the model, we now proceed to express the overall system coherence in terms of its component partial coherence spectra. Quantitatively, we can represent the model output $\hat{Y}(m)$ at frequency m as follows:

$$\hat{Y}(m) = \hat{Y}_{L_1} + \hat{Y}_{L_2} + \hat{Y}_{Q_{11}} + \hat{Y}_{Q_{22}} + \hat{Y}_{Q_{12}} \tag{15}$$

where $\hat{Y}_{L_1}$, $\hat{Y}_{L_2}$, $\hat{Y}_{Q_{11}}$, and $\hat{Y}_{Q_{22}}$ represent the model outputs of L_1, L_2, Q_{11}, and Q_{22} in Fig. 2, respectively. The $\hat{Y}_{Q_{12}}$ term takes into account the outputs of both Q_{12} and Q_{21}, since Q_{12} and Q_{21} actually represent a single quadratic transfer function for which the inputs are X_1 and X_2.

The model output power spectrum $\hat{P}_y(m)$ is obtained by multiplying (15) by $\hat{Y}^*(m)$, then taking an expected value. Next we divide the equation for $\hat{P}_y(m)$ by the actual (i.e., "true") output power spectrum, $P_y(m) = E[Y^*(m)Y(m)]$, and we observe that the LHS becomes the system coherency spectrum $\gamma_{system}(m)$ and the terms on the RHS represent the fractional contributions (i.e., partial coherence spectra) of each transfer function to the model output power spectrum at frequency m, i.e.,

$$\gamma_{system}(m) = \gamma_{1,1} + \gamma_{2,2} + \gamma_{11,11} + \gamma_{22,22} + \gamma_{12,12} \tag{16}$$

$$+ \gamma_{1,2} \tag{17}$$

$$+ \gamma_{1,11} + \gamma_{1,22} + \gamma_{1,12} + \gamma_{2,11} + \gamma_{2,22} + \gamma_{2,12} \tag{18}$$

$$+ \gamma_{11,22} + \gamma_{11,12} + \gamma_{22,12}. \tag{19}$$

For example, $\gamma_{1,1}$ represents in fraction form, the contribution to the total output power at frequency m of the linear L_1 transfer function alone. Similarly, $\gamma_{22,22}$ represents the fractional contribution to observed output power at a frequency m of the quadratic Q_{22} transfer function alone, and so on.

Note that contributions of those terms given in (16) are intrinsically positive. On the other hand, the contributions to the model output power represented by the terms in (17)–(19) result from the interaction of the complex outputs of different transfer function models. If the input is exactly Gaussian all six terms in (18) will be zero, since they involve odd order spectral moments. It is particularly important to note that the terms in (17)–(19) may be negative, since they correspond to the real part of a complex quantity.

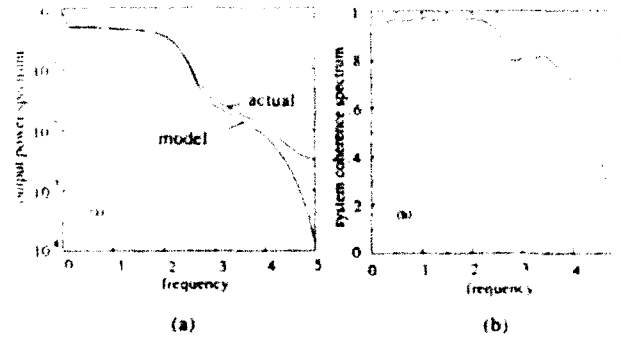


Fig. 3. (a) Comparison of actual and model output power spectra of non-zero-mean exponentially distributed inputs by general input method. (b) system coherence spectrum.

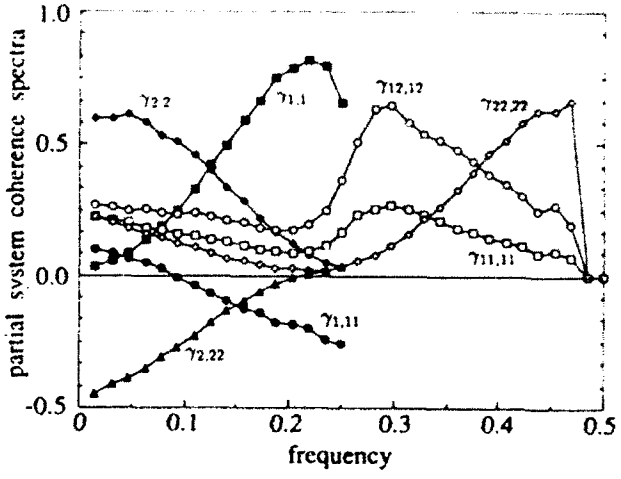


Fig. 4. Partial system coherence spectra estimated by using general input method for zero-mean exponentially distributed inputs.

In Fig. 4 we have plotted the dominant seven of the 15 partial coherence spectra for the simulation experiment. Note that those terms dependent upon a linear transfer function exist only in the region $0 < f < 0.25$, because of the antialiasing filters applied to the inputs. Observe, in accordance with our previous discussions, that all the terms appearing in (16) are positive, while $\gamma_{1,11}$ and $\gamma_{2,22}$ are both positive and negative. Because of the nature of this simulation example, all partial coherence spectra shown in Fig. 4 are of comparable maximum amplitude, but peak up at different frequencies. Although of no physical significance in this simulation example, such plots are capable of potentially providing significant insight into the nonlinear transfer of energy from various spectral bands in the two inputs to other spectral bands in the output of a quadratically nonlinear system.

IV. Conclusion

There are a number of practical applications where it is important to have the capability to model dual-input nonlinear systems, some examples of which follow: ground vibration tests of aircraft where two random exciting forces are applied in order to inject sufficient energy into all modes of the system; and experimental studies of transition to turbulence in fluids and plasmas, where the two inputs consist of two different velocity components, and plasma density and potential fluctuations, respectively. The approach described in

this correspondence is currently being utilized to investigate such applications.

Acknowledgement

The computing resources for this work were provided by the University of Texas System Center for High Performance Computing.

References

[1] S. A. Billings, "Identification of nonlinear systems—a survey," *Proc. Inst. Elec. Eng.*, vol. 127, pt. D, no. 6, pp. 272-285, 1980.

[2] Ch. P. Ritz, E. J. Powers, R. W. Miksad, and R. S. Solis, "Nonlinear spectral dynamics of a transitioning flow," *Phys. Fluids*, vol. 32, no. 12, pp. 3577-3588, Dec. 1988.

[3] Y. S. Cho, W. T. Oh, S. B. Kim, and E. J. Powers, "Testing for Gaussianity in nonlinear system identification," in *Proc. 1990 IEEE Int. Symp. Circuits Syst.* (New Orleans, LA), May 1-3, 1990, pp. 1450-1453.

[4] K. I. Kim and E. J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 12, pp. 1758-1769, Nov. 1988.

[5] C. K. An, E. J. Powers, and C. P. Ritz, "Frequency domain modeling of dual-input/multiple-output quadratic systems with general random inputs," in *Proc. IEEE Int. Symp. Circuits Syst.* (Espoo, Finland), June 7-9, 1988, pp. 2209-2211.

Bispectral magnitude and phase recovery using a wide bandwidth acousto-optic processor

Stacy K. Kniffen, Michael F. Becker, and Edward J. Powers

A hybrid optical–digital processor has been developed that computes both the magnitude and phase of the bispectrum for wide bandwidth (10 MHz to 1 GHz) rf signals. The overall optical architecture is that of a modified Mach–Zehnder interferometer that contains three acousto-optic modulators and appropriate transforming lenses. The intensity distribution in the output plane of the interferometer contains an interference term that represents the real part of the bispectrum multiplied by a spatial carrier (the interference fringes). To isolate the bispectrum information, the output image is digitized and digitally filtered. The imaginary part of the bispectrum is obtained by Hilbert transforming the real part, and then computing the bispectrum magnitude and phase. The processor is tested with four different combinations of rf test signals. Each signal has a bandwidth of either 6 or 12 MHz. Test results that illustrate the performance of the processor in the recovery of magnitude and phase information for the bispectrum of quadratically related signals are presented.

I. Introduction

The bispectrum is an example of a class of functions called higher-order spectra or polyspectra. Such functions are useful in the analysis of nonlinear phenomena and nonlinear systems. Estimates of the polyspectra (which include the power spectrum as a special case) are defined mathematically as expected values of successively higher-order products of Fourier transforms. For the case of the bispectrum, this is a triple product[1]:

$$B_{xx}(f_1, f_2) = \lim_{T \to \infty} T^{-1} E[X^*(f_1) X^*(f_2) X(f_1 + f_2)], \quad (1)$$

where $X(f)$ is the Fourier transform of a random process $x(t)$ over a finite time interval T, and $E[\,]$ is the expected value of a random process. Note that this definition assumes that the random process is stationary to the necessary order. Nonlinearities result in the redistribution of a signal's energy among various spectral components. The polyspectra provide a means of quantifying the resulting correlation among the spectral components. The bispectrum is particularly useful for the analysis of quadratically nonlinear systems.

Although digital implementations of the polyspectra have found numerous applications (several of which are described in Section II), the calculation of these quantities can be quite computationally intensive. In the computation of the discrete bispectrum, the formation of the triple products of discrete Fourier transforms (DFT's) requires of the order of $N\beta$ complex multiplies per second for real-time operation (where N is the size of the DFT and β is the signal bandwidth). This does not even account for the initial computation of the DFT's. Thus the applications to date have been limited to higher-order spectral analysis of fairly narrow-band signals and/or non-real-time processing. To extend the techniques of higher-order spectral analysis to wider bandwidths (10 MHz and above) in real time, one rapidly outruns the capabilities of existing general-purpose digital computers and is forced to look for alternative methods of implementation. Other constraints (size, weight, power, and cooling), in addition to processing speed, must often be considered in selecting such a method. In this spirit, optical signal-processing techniques appear to hold promise as alternative or complementary methods of implementation of higher-order spectra for wide bandwidth signals.

Optical signal-processing techniques have proved to be quite powerful in performance as well as in flexibility. One of the key advantages that these methods offer is the ability to perform real-time spatial Fourier transforms with lenses. The ability to convert wide bandwidth time-varying electrical sig-

The authors are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, Texas 78712.

Received 2 January 1991
0003-6935/92/081015-15$05.00/0.
© 1992 Optical Society of America.

nals into space-varying light-amplitude distributions by using acousto-optic (AO) modulators is another important attribute of optical processing systems. The most basic signal-processing function implemented with acousto-optics is real-time power-spectrum analysis. A sampling of the other applications include: optical correlators, convolvers, interference rejection filters, and optical fast Fourier transform processors.[2,3] Two-dimensional optical processors have been developed to compute such quantities as the ambiguity function,[4] the cyclostationary function, and the Wigner distribution.[5] A noninterferometric triple-product processor[6] has also been reported.

An optical realization of the bispectrum was developed at the University of Texas by Kauderer *et al.*[7] following the preliminary studies by Florence and Song.[8] The overall optical configuration was a variation of a Mach–Zehnder interferometer (which included several lenses for transforming and scaling) and three AO modulators (as spatial light modulators). The propagation of the diffracted light through the system of lenses and the superposition of the two interferometer paths was shown to generate an intensity distribution in the output plane involving the same product of Fourier transforms as the bispectrum. The experimental part of Kauderer's work provided a proof-of-principle demonstration of the optical bispectrum processor. The tests performed emphasized the measurement of the magnitude of the bispectrum (actually the bicoherence, a normalized form of the bispectrum) for sinusoidal input signals by observing the interference pattern over a localized region of the bifrequency plane.

The objective of the research presented in this paper was to develop an optical processor capable of computing both the magnitude and the phase of the bispectrum for wide bandwidth input waveforms. For such waveforms, it was necessary to compute the value of the bispectrum over an extended region of the output (bifrequency) plane rather than simply for a single point. Some modifications to the existing optical architecture were necessary to satisfy this requirement. A substantial amount of effort was directed at the digital processing of the output of the optical system to recover the magnitude and phase of the bispectrum. The interference term in the output intensity distribution consisted of the real part of the bispectrum imposed on a spatial carrier. This term was isolated from the more slowly varying noninterference terms by digital filtering. A discrete Hilbert transform approach was used to compute the imaginary part of the bispectrum from the real part. The magnitude and phase were then computed in the normal fashion. The processor was tested with four different rf signal sets, which produced four different magnitude–phase profiles for the bispectrum. The experimental results reported here are in good agreement with expectations.

In the next section, some basic definitions, properties, and applications of the bispectrum are discussed. Section III describes the optical implementation of the bispectrum including the optical layout, the rf test signal generation process, and the initial output image processing. The application of the discrete Hilbert transform to the problem of recovering bispectrum magnitude and phase is the topic of Section IV. Test results that demonstrate the capability of the processor to recover these quantities for wide bandwidth input waveforms are presented in Section V. The last section provides a summary.

II. Definitions and Properties of the Bispectrum

Before discussing the recent developments in optical bispectrum processing, we summarize some key definitions and properties of the bispectrum. It is well known that the power spectral density function for a wide-sense stationary random process can be defined as the Fourier transform of the correlation function. The higher-order spectra also have associated higher-order correlations, which are related to the spectra through the Fourier transform. Our primary interest is in the second-order correlation, which is also known as the bicorrelation,[1]

$$R_{gff}(\tau_1, \tau_2) = E[g(t)f(t-\tau_1)f(t-\tau_2)], \quad (2)$$

where $f(t)$ and $g(t)$ are zero-mean, real-valued stationary random processes. Note that what we have written is actually the cross bicorrelation. The cross bispectrum can then be expressed as the Fourier transform of R_{gff},

$$B_{gff}(f_1, f_2) = \int\int R_{gff}(\tau_1, \tau_2)\exp[-2\pi j(f_1\tau_1 + f_2\tau_2)]\,d\tau_1\,d\tau_2, \quad (3)$$

or directly in terms of the Fourier transforms of $g(t)$ and $f(t)$,

$$B_{gff}(f_1, f_2) = \lim_{T\to\infty} T^{-1}E[G(f_1+f_2)F^*(f_1)F^*(f_2)]. \quad (4)$$

The bispectrum is a complex-valued function of two frequency variables, and it measures the correlation between the spectral components at frequencies f_1 and f_2 and their sum frequency $f_1 + f_2$ caused, for example, by a quadratic (square-law) nonlinearity. Two conditions must be met for the bispectrum to be nonzero. Obviously, the signal of interest must have nonzero spectral components at each of the three frequencies f_1, f_2, and $f_3 = f_1 + f_2$. This condition is referred to as the frequency sum rule. The second condition requires that the spectral components be statistically dependent. In essence, this requirement can be satisfied by demanding that their respective phases obey a phase sum rule given by $\phi_3 = \phi_1 + \phi_2 +$ constant. Another way of stating this is that the spectral components must be phase coherent for the bispectrum to be nonzero.

It is possible that a given time series $x(t)$ may contain a nonzero spectral component at $f_3 = f_1 + f_2$ that is independently generated and hence not correlated with the components at f_1 and f_2. This situation does not contribute to the bispectrum at f_1, f_2 because

the phase sum rule will be violated. The difference between the results obtained for the bispectrum for quadratically related (phase-coherent) and independently generated (incoherent) components provides a powerful means of distinguishing the two situations. One can also imagine a situation in which there is a mixture of coherent and incoherent energy. For such a situation, the (cross) bicoherence function can provide useful insight:

$$b_{eff}^2(f_1, f_2) = \frac{|B_{eff}(f_1, f_2)|^2}{E\{|F(f_1)F(f_2)|^2\}E\{|G(f_1 + f_2)|^2\}} \quad (5)$$

The bicoherence ranges from zero (completely incoherent) to one (completely coherent).

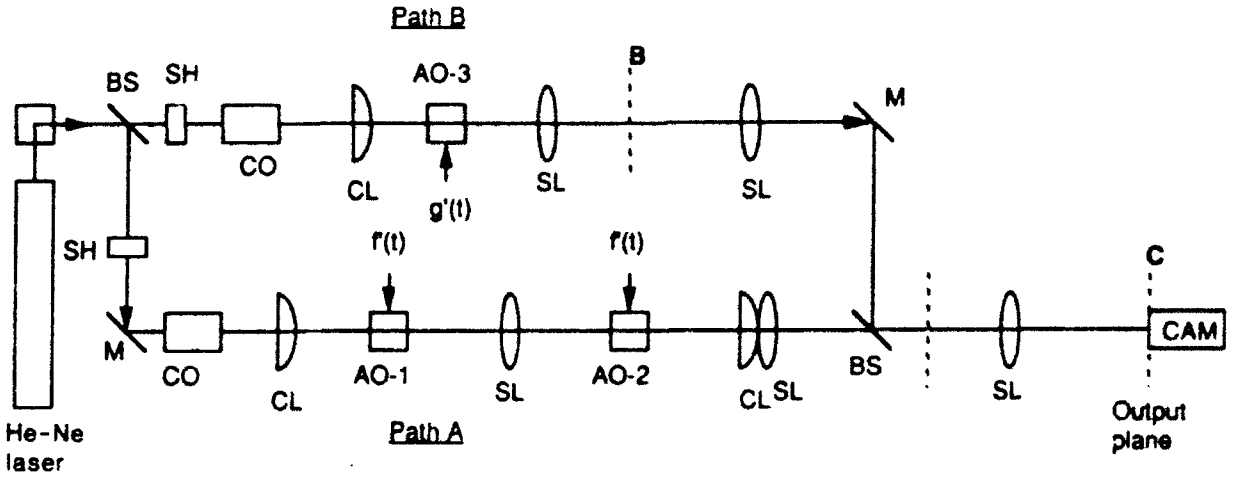
The bispectrum has been used extensively at the University of Texas at Austin for the treatment of such problems as plasma fluctuations,[9] transition to turbulence in subsonic airflow,[10] the nonlinear response of moored vessels or tension leg platforms to random sea excitation,[11] and nonlinear system modeling in the frequency domain.[12] Others have applied these methods to the enhancement of atmosphere-degraded infrared astronomical observations,[13] to the study of nonlinear relationships in electroencephalograms,[14] and for emitter characterization in holographic sonar.[15,16] Some of these applications make use of the fact that the bispectrum is zero for zero-mean additive Gaussian processes in order to enhance a non-Gaussian signal of interest in the presence of Gaussian noise. None of these applications is particularly stringent in terms of its processing requirements. Thus a digital implementation on general-purpose hardware has been the method of choice.

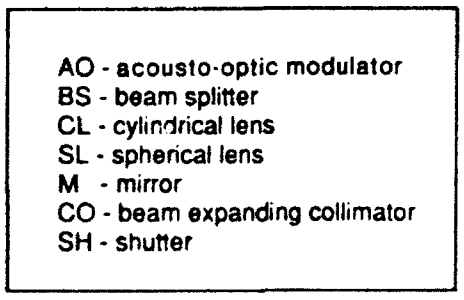
III. Optical Implementation of the Bispectrum

As stated in the introductory section, the primary objective of this research is to compute the magnitude and phase of the bispectrum over an extended region of the bifrequency plane for wide bandwidth signals. This section describes the implementation of the optical bispectrum processor in terms of the basic optical architecture, the circuit for rf test signal generation, test case selection, and image acquisition and initial processing.

A Optical Layout

Figure 1 shows the current optical architecture used to implement the bispectrum. The two modulators in path A have a 40-MHz center frequency and a 20-MHz bandwidth, while the third AO cell has a





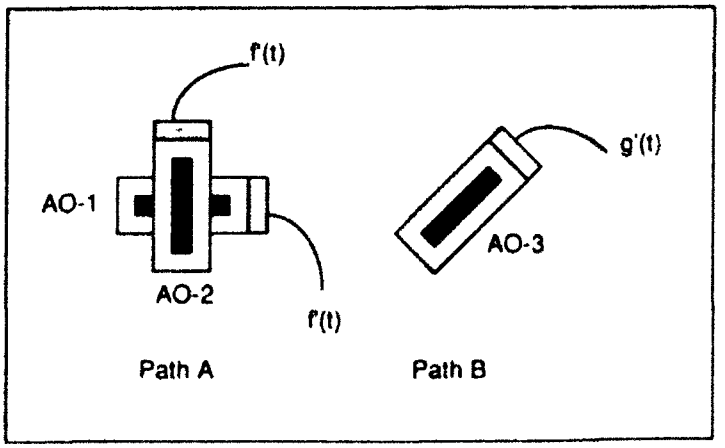


Fig. 1 Optical layout for the bispectrum processor. Inset: Orientation of the AO cells.

70-MHz center frequency and a 40 MHz bandwidth. Baseband signals $f(t)$ and $g(t)$ are upshifted into the passbands of the modulators. The upshifted signals are denoted by $f'(t)$ and $g'(t)$. Each AO cell has a maximum time aperture of 5 μs, which results in a minimum frequency resolution of 200 kHz. It should be noted that the processor depicted in the figure is capable of computing the cross bispectrum between the signals $f'(t)$ and $g'(t)$.

In each of the interferometer paths, the light from the He–Ne laser is expanded and recollimated. The spherical and cylindrical lenses are positioned with respect to the modulators and oriented so that the light comes to a focus as a line inside the cells. The inset of Fig. 1 shows the orientation of each of the three modulators as viewed from the output end of the interferometer. In Path A, the first modulator is oriented horizontally. The first diffracted order of the horizontal cell illuminates the second modulator, which is oriented vertically.

Although the basic optical architecture (a modified Mach–Zehnder interferometer) remains the same as that used by Kauderer *et al.*,[7] the present effort to generalize the technique to allow optical bispectrum processing of wide bandwidth waveforms requires a modification in the Path B portion of the layout. Specifically, the third AO modulator is oriented at a 45-deg angle, and its output is scaled by $\sqrt{2}$ with respect to Path A to provide the correct overlap between transforms over an extended region of the bifrequency plane. This is where the present layout differs from that used by Kauderer. The new configuration for the third modulator and the image scaling ensure that the frequency sum rule is satisfied and that efficient use is made of the available light.

In our discussion of the bispectrum, it is more natural to talk in terms of temporal frequency (hertz or megahertz) rather than spatial coordinates. Thus, we will adopt the symbols f_1 and f_2 as representing the two temporal frequency variables (in megahertz) that have been mapped onto a spatial coordinate system (u, v) according to $f_1 = uv_a/\lambda f_l$ and $f_2 = vv_a/\lambda f_l$, where v_a is the acoustic velocity in the AO cells, λ is the optical wavelength in free space, and f_l is the focal length of the transforming lens. Considering only the first diffracted orders from the modulators and assuming a plane-wave, Gaussian amplitude profile illumination, we can write the output amplitude distribution from Path A as[2,17]

$$F'(f_1, t)F'(f_2, t) = \int \exp(-2\pi j f_1 x / v_a) \times \exp(-x^2/2\sigma^2) f'(x - v_a t)\,dx \int \exp(-2\pi j f_2 y / v_a) \times \exp(-y^2/2\sigma^2) f'(y - v_a t)\,dy. \quad (6)$$

and from Path B as

$$G'(f_1 + f_2, t) = \int \exp[2\pi j (f_1 + f_2)x / v_a] \exp(-x^2/2\sigma^2) g'(x - v_a t)\,dx. \quad (7)$$

where we have suppressed the optical plane-wave factors [i.e., $\exp(j\omega_0 t - \mathbf{k} \cdot \mathbf{r})$] and constant factors outside the integrals. Note that in Eq. (7) the variation of the Gaussian beam in the direction perpendicular to the transform direction is assumed to be slow enough over the region of interest ($f_1 \approx f_2$) that this factor can be approximated by unity. If we include the v_a^{-1} factor in the argument of the transforms, we can express the transforms in spatial frequency units. We have suppressed this factor for now in order to make the connection with the temporal frequencies clearer.

Prior to recombining the beams from Paths A and B, we need to reintroduce the plane-wave factors that go with them. The intensity measured in the output plane is given by the square of the sum of the amplitudes. The charge-coupled device (CCD) camera placed in the output plane integrates this intensity distribution over a fairly long time interval, resulting in a measured output intensity given by

$$I_{\text{out}} = |F'(f_1)F'(f_2)|^2 + |G'(f_1 + f_2)|^2 + 2\,\text{Re}[G'(f_1 + f_2)F'^*(f_1)F'^*(f_2)\exp(j\phi(f_1, f_2))]. \quad (8)$$

where the previous time dependence has been canceled directly or averaged out. The complex exponential in the last term represents the spatial dependence of the interference fringes. The phase factor is a function of the curvature of the interfering optical wave fronts and the angle between the propagation vectors. For present purposes, this is assumed to be a linear function of the spatial coordinates, giving rise to straight fringes of uniform spacing (i.e., a constant-frequency spatial carrier). Note that the last term in the expression has the same form as the cross bispectrum, while the first two terms represent the intensity of the light from each path in the absence of the other. Figure 2 indicates the resulting intensity distribution in the interference plane for typical input signals and the projections of their transforms on the appropriate axes. The presence of a nonzero bispec-

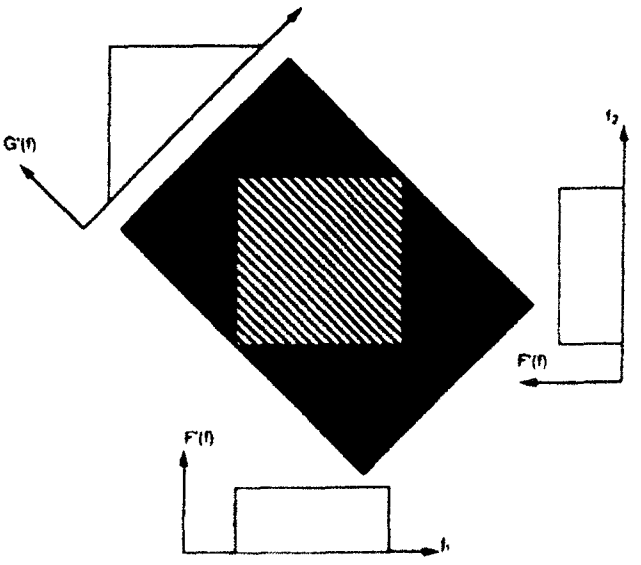


Fig. 2 Output intensity distribution and projections of the input spectra

trum will result in the appearance of interference fringes in the output image. The fringe contrast is governed by the relative amount of power present in the coherent (quadratically related) and incoherent (independently generated) signal components that were used to drive the AO cells.

B. Rf Input Circuitry

The task of generating the rf input signals to test the AO bispectrum processor effectively requires a high degree of flexibility and control over the waveform parameters as well as a relatively wide bandwidth. The bispectrum measures the correlation between different spectral components satisfying the frequency sum rule. In testing, it is desirable to be able to specify this correlation as we see fit, that is, to introduce signals related by an arbitrary quadratic transfer function. Furthermore, the signals of interest should be of the order of 10 MHz in bandwidth (for our particular system). It is also helpful if two different signals can be produced simultaneously.

The above requirements were met by the Hewlett-Packard HP8770A arbitrary waveform synthesizer (AWS). The HP8770A is the source for all the signals used to drive the AO modulators (see Fig. 3). An arbitrary digital sequence representing samples of the desired waveform is downloaded from a microcomputer to the AWS by a GPIB interface (a general-purpose interface bus, conforming to the IEEE-488 standard). In turn, the AWS converts the sequence (12-bit, 125- MHz sample rate) to an analog signal with a bandwidth of up to 50 MHz. The maximum output is 2 V peak to peak, and the output impedance is 50 Ω. In addition to the standard signal output, the AWS produces a 10-MHz sinusoidal reference signal that is phase locked with its 125-MHz clock.

In our experiments, the Path A and Path B drive signals were generated simultaneously by a single AWS by means of frequency multiplexing. The Path A drive signal $f'(t)$ was generated directly [i.e., the baseband signal $f(t)$ was upshifted into the 30–50-MHz passband of the AO cells in the signal definition process]. This was then added to the baseband signal $g(t)$, producing the AWS output signal

$$x(t) = f'(t) + g(t) \quad (9)$$

This signal was split and sent to both rf drive circuits. The Path A drive voltages were produced by bandpass filtering (30–50-MHz passband) $x(t)$ and amplifying the resulting signal. Producing the Path B drive signal was somewhat more involved since $g(t)$ must be placed in the passband of the third AO cell. The AWS output signal was mixed with an 80 MHz local oscillator, then bandpass filtered to isolate the signal of interest before rf amplification. It is important to note that the 80-MHz carrier used must be phase coherent with the AWS output signal in order to maintain coherence between the signals $f'(t)$ and $g'(t)$. To achieve this, the 10-MHz sine-wave reference signal, which is provided by the AWS, was doubled three times, resulting in an 80 MHz signal with the proper phase relationship.

Figure 4 shows the spectra of the Path A and Path B drive voltages and the original AWS output signal. The shaded regions represent the spectra of the signals, and the modulator passbands are indicated by dotted lines. The unshaded regions represent missing upper sidebands. In the case of $F'(f)$, the sideband was suppressed digitally in the definition of the signal. For $G'(f)$, the 60–80-MHz bandpass filter suppressed the upper sideband, which was produced as a result of the rf mixing process. We think of signals $F'(f)$ and $G'(f)$ as representing the lower sidebands of a pair of baseband signals placed on 40- and 80-MHz carriers, respectively. The spectral configuration of F' and G' was selected to achieve the phase-coherence requirement as well as to optimize the position of the spectra in the modulator passbands, to help with dynamic range considerations, and to satisfy the frequency sum rule.

C. Test Signal Definition

At this stage of our research, it is necessary to balance the desire to process more complicated test signals

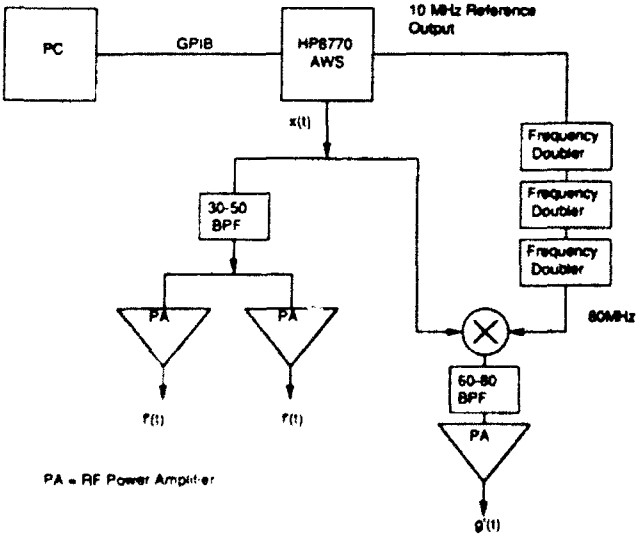


Fig. 3 Rf signal generation circuit.

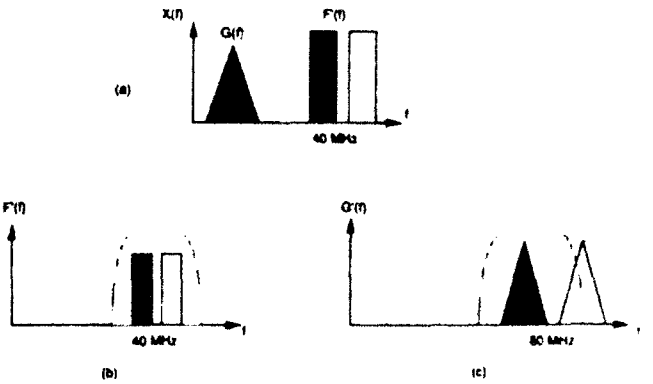


Fig. 4 Spectra of the AWS output signal (a) and the AO cell input signals (b) and (c).

with the practical need to have signals that will produce an understandable optical output when their bispectrum is computed. In order to satisfy these constraints and test the wide bandwidth processing capability of the optical implementation, the single-tone tests used by Kauderer *et al.*[7] have been replaced with tests based on sets of uniform amplitude, closely spaced tones, whose frequencies and phases are to be specified. This approach is attractive from several perspectives. In particular, the mathematical analysis of the problem and the digital sequence generation remain relatively simple.

Each of the test cases presented here was based on a single choice of $f(t)$ representing the input to some quadratically nonlinear system. The output of the system, $g(t)$, was related to $f(t)$ according to

$$g(t) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h_2(t-t_1, t-t_2) f(t_1) f(t_2) dt_1 dt_2, \qquad (10)$$

where h_2 was the desired quadratic impulse response. In the frequency domain, this became

$$G(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H_2(f_1, f_2) F(f_1) F(f_2) \delta(f - f_1 - f_2) df_1 df_2, \qquad (11)$$

where $H_2(f_1, f_2)$ was the quadratic transfer function.[12,18,19]

The signal $f(t)$ consisted of a sum of 30 uniform amplitude, constant-phase sinusoids:

$$f(t) = \mathrm{Re}[\hat{f}(t)] = \mathrm{Re}\left\{\sum_{n=1}^{30} \exp(2\pi j f_n t)\right\}, \qquad (12)$$

where $f_{n+1} - f_n$ is 200 kHz. This gave $f(t)$ an overall bandwidth of 6 MHz. Note that this choice of $f(t)$ could also be described as a series of sinc pulses 5 μs apart [see Fig. 5(a)]. Thus the problem could be restated as one of investigating the response of a quadratically nonlinear system to an impulsive input probe signal—an approach commonly used in many physical situations.

In the frequency domain, the complex signal $\hat{f}(t)$ became

$$\hat{F}(f) = \sum_{n=1}^{30} \delta(f - f_n) \qquad (13)$$

Since the tones had constant amplitudes and phases and were spaced at the minimum resolvable spacing for the AO cells, the optical output $|F(f_1)F(f_2)|^2$ was a uniform square region (6 MHz × 6 MHz) in the bifrequency plane.

With the signal $f(t)$ specified, the remainder of the signal synthesis process involved the selection of the quadratic transfer function $H_2(f_1, f_2)$, used to produce $g(t)$ from $f(t)$. It is useful to begin in the

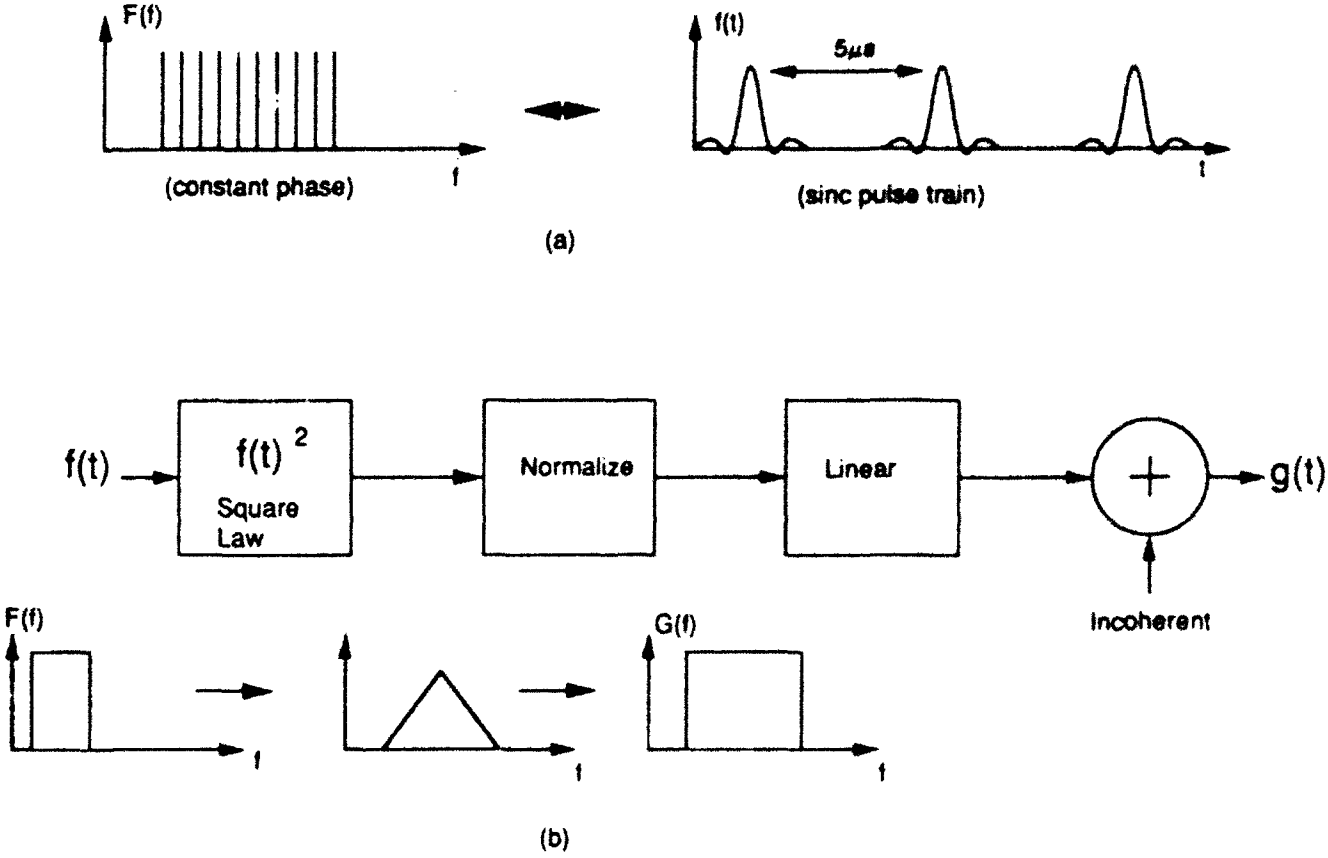


Fig. 5. (a) Definition of $f(t)$, (b) relationship between $f(t)$ and $g(t)$

frequency domain with Eq. (11) and substitute in $\hat{F}$ from Eq. (13):

$$\hat{G}(f) = \sum_{m=1}^{N} \sum_{n=1}^{N} H_2(f_m, f_n)\delta(f - f_m - f_n) \tag{14}$$

In the time domain this becomes

$$g(t) = \mathrm{Re}[\tilde{g}(t)] = \mathrm{Re}\left\{\sum_{m=1}^{N} \sum_{n=1}^{N} H_2(f_m, f_n)\exp[2\pi j(f_m + f_n)t]\right\} \tag{15}$$

Note that $H_2(f_1, f_2)$ is evaluated only at certain discrete frequencies.

The actual quadratic transfer functions (QTF's) used can be modeled with a square-law operation followed by a linear system. This is depicted in Fig. 5(b). The specific QTF's chosen do not necessarily represent real physical systems. Mathematically, they can be written as a product of two factors:

$$H(f_m, f_n) = H_{lin}(f_m + f_n)H_0(f_m, f_n). \tag{16}$$

The factor H_0 will be taken to be unity for all frequency pairs f_1 and f_2. Considered by itself, this corresponds to a simple square-law relation: $g(t) = f(t)^2$. This results in a triangular spectrum representing convolution in the frequency domain, $G(f) = F(f) * F(f)$, as Fig. 5(b) indicates.

For the term H_{lin}, four different linear transfer functions (LTF's) are used. Figures 6(a)–6(d) depict the magnitude and phase profiles for $F(f)$ and $G(f)$ associated with each LTF. The first LTF is unity and produces the triangular spectrum (a) just described. The second LTF normalizes the magnitude of the output of the square-law device, returning it to a rectangular shape (b). It has the form

$$H_{lin}(f_m + f_n) = \frac{1}{N - |N - (m + n) + 1|} \tag{17}$$

The normalization step is intended to facilitate the rf diagnostics, simplify the optical output, and lessen the image acquisition dynamic range requirements. The third and fourth LTF's also have magnitude responses given by H_{lin}. They differ from the second LTF in their phase response. The first and second LTF's have constant-phase profiles while the third and fourth have quadratic (c) and step-phase profiles (d), respectively.

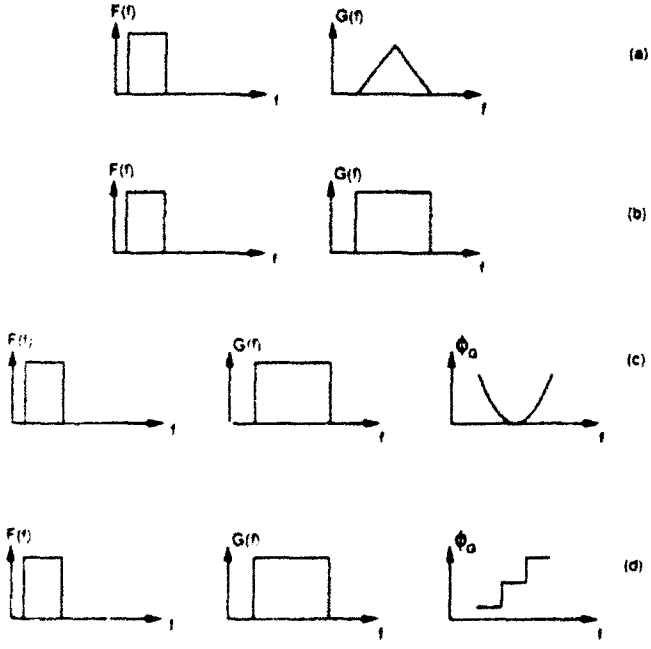


Fig. 6. Spectral profiles for test signals related by various QTF's: (a) square law, (b) normalized magnitude, (c) normalized magnitude and quadratic phase, and (d) normalized magnitude and step phase.

D. Image Acquisition and Initial Processing

The images formed in the output plane were acquired by using a 512 × 512 pixel CCD camera and microcomputer-based video frame grabber. The relatively slow frame rate (30 frames/s) was consistent with our assumption of a long integration time. The images were digitized to 8-bit resolution. In the tests presented here, the 6 MHz × 6 MHz region of interest in the bifrequency plane spanned roughly a 128 × 128 pixel subframe. Although the bispectrum is a function of position and the fringes have a spatial dependence, the images acquired in this series varied slowly enough that the spatial sampling rate was not a problem. We make use of this property in the discussion below.

Before continuing the discussion of the recovery of bispectral magnitude and phase, it is appropriate to make the connection between the information present in the output plane and the equivalent digital implementation of the bispectrum. The AO cells used in our experiments had 5-μs time apertures. Thus the effective time–bandwidth product for the 6-MHz subband was 30. Equating this to the number of DFT bins for a digital implementation means that the effective processing rate for the formation of triple products of DFT's was of the order of (6 MHz × 30 = 180 million complex multiplies/s). This was significant but not overwhelming. The potential of the optical realization of the bispectrum is more dramatically illustrated by considering a processor employing the most powerful commercially available AO modulators with bandwidths of the order of 1 GHz and time–bandwidth products of the order of 1000. Effective processing rates in the trillion complex multiplies/s range seem possible.

The magnification of the output image was such that 128 pixels roughly corresponded to a 6-MHz span. For subsequent processing, a more accurate spatial-to-frequency calibration must be obtained. It was also necessary to compensate for inversion of the image and the nonunity aspect ratio of the CCD array (it is wider than it is high). A calibration test involving a set of four tones separated by 2 MHz was included as a part of the standard data acquisition process. The signal was used to drive the two Path A modulators. The output image took the form of a 4 × 4 grid of intensity peaks that spanned the 6 MHz × 6

MHz region of interest. The intensity maxima were located and associated with the appropriate pair of temporal frequencies. Two points on the extreme corners were used to determine the slope and intercept for the pixel-to-frequency conversion for each dimension. Such a measurement was made each time anything affecting the calibration changed, including image magnification and fringe orientation.

A 256 × 256 pixel subframe (roughly 12 × 12 MHz) was extracted from the original image frame and was scaled by using the calibration information described in the previous paragraph. In the process, the original subframe was mapped into a new one properly oriented and scaled. Four adjacent pixels of the original subframe were averaged together to form a single output pixel. Thus the new subframe had a dimension of 128 × 128 pixels corresponding to 12 × 12 MHz in the bifrequency plane. This last step can be viewed as a low-pass filter and decimate operation that serves to smooth the output as well as to reduce the image processing requirements.

E. Interference Term Isolation

The interference image subframes were subsequently processed to isolate the interference term (which contains the bispectrum information) and to recover magnitude and phase information. The intensity distribution observed was as predicted in Eq. (8) except that the effect of additive background noise must be included. The expression then becomes

$$|F'F'|^2 + |G'|^2 + 2\,\mathrm{Re}[F'^*F'^*Ge^{i\phi}] + n_i(f_1, f_2). \quad (18)$$

where $n_i(f_1, f_2)$ is the background term. In order to isolate the bispectrum information contained in the interference fringes, the noninterference and background noise contributions to the total intensity must be subtracted or otherwise filtered. After dividing the result by two, we are left with the desired term representing the real part of the cross bispectrum on a spatial carrier:

$$\hat{B}_{ig}(f_1, f_2) = \mathrm{Re}[F'^*(f_1)F'^*(f_2)G'(f_1 + f_2)e^{i\phi}]. \quad (19)$$

The approach to isolating the interference term is depicted in Fig. 7(a). It differs from the method used by Kauderer *et al.*, which relied on the sequential acquisition of interference, noninterference, and background frames. The interference term containing the bispectrum information is isolated by subtracting the other frames. There are two principal disadvantages to this method: (1) the need to acquire multiple frames and (2) the frame-to-frame independence of the noise that effectively increases the noise variance of the final frame. The new method uses the spectral characteristics of the images in the present test set to accomplish the isolation step based solely on the interference frame. Even after the decimation step that accompanies the scaling of the images, the interference images are still oversampled by more than two in each dimension. The first step in the new isolation technique is to average another four pixels

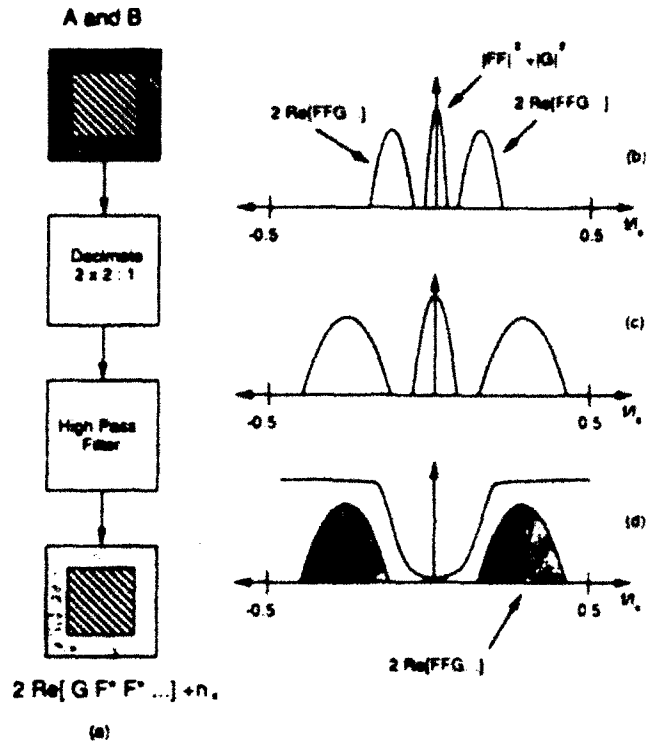


Fig. 7. (a) Flow chart of the interference term isolation process, (b) cross section of simulated interference image before the decimation step, (c) after decimation step, and (d) after high-pass filter step.

(two by two) into a single pixel and thus to produce a new 64 × 64 pixel image. This represents a second low-pass filter and decimation step. The effect of decimation on the interference image is illustrated in Figs. 7(b) and 7(c), which show simulated one-dimensional slices of the transform of the image before and after the decimation step. Initially, the spectrum of the image is compressed near the origin. The decimation operation expands the transform in terms of the normalized frequency.

There is another feature that is illustrated in Figs. 7b and 7c. The interference and noninterference terms are shown as separate in the frequency domain. This is because the interference fringes act as a spatial carrier for the bispectrum information. Depending on the spatial frequency of the fringes and the bandwidth of the bispectrum, the interference term may or may not be separate from the noninterference terms. The condition for sufficient separation is that the spatial carrier be greater than the sum of half of the bandwidth of the noninterference terms plus half of the bandwidth of the interference term. Fortunately, our data set includes images that satisfy this condition. This brings us to the second step of the image isolation process. The interference frame is high-pass filtered to remove the relatively slowly varying noninterference terms and leave the bispectrum information (on a spatial carrier). The resulting intensity distribution is given by

$$\hat{B}_{ig}(f_1, f_2) = \mathrm{Re}[F'^*(f_1)F'^*(f_2)G'(f_1 + f_2)e^{i\phi}] + \hat{n}_i. \quad (20)$$

where $\bar{n}$, is the modified background noise from the interference frame. The high-pass filter is implemented using a 19-tap, finite impulse-response filter that is convolved with each row and each column of the interference image. The filter was designed by a simple frequency sampling method.[20] Its magnitude response is shown in Fig. 7(d), superimposed on the cross section of the simulated image spectrum. It by no means represents the optimal choice. However, it is adequate for present purposes, as the results that follow show.

IV. Magnitude and Phase Recovery

In Section I it was stated that the bispectrum is complex valued and that a chief objective of this research was to recover magnitude and phase information from the optical realization of the bispectrum for wide bandwidth waveforms. The interference term that has been isolated represents only the real part of the bispectrum. We have devised a method to compute the imaginary part of the bispectrum from the real part by performing a two-dimensional discrete Hilbert transform. This section deals with the theory and implementation of the method.

A. Hilbert Transform

The Hilbert transform relates the real and imaginary parts of a complex-valued quantity. In analog signal theory, the Hilbert transform is encountered in conjunction with the theory of analytic signals.[21] It is interesting and perhaps not too surprising that the concept of the analytic signal was attributed to Gabor,[22] who also was responsible for early work in holography.[17]

In electromagnetic theory, the Hilbert transform concept is seen in terms of the Kramers–Kronig relations between the real and imaginary parts of the dielectric constant.[23] These follow as a result of the causal relation between the electric displacement and the electric field. In linear system theory, the Hilbert transform relates the real and imaginary parts of the linear transfer function under the constraint that the impulse response is causal.[18]

A discussion of the analytic signal will aid in our understanding of the Hilbert transform. We can derive a complex-valued signal $\tilde{x}(t)$ from a real-valued bandpass signal $x(t)$ as follows:

$$\tilde{x}(t) = x(t) + j\hat{x}(t), \tag{21}$$

where the imaginary part is the Hilbert transform of the real part:

$$\hat{x}(t) = -\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(t')}{t-t'}\,dt'. \tag{22}$$

Note that the Hilbert transform takes a function of a given variable into another function of that same variable. In the frequency domain, the Hilbert transform corresponds to multiplication of the spectrum of x by $-j\,\mathrm{sgn}(f)$. That is,

$$\hat{X}(f) = -j\,\mathrm{sgn}(f)X(f) \tag{23}$$

The spectrum of the analytic signal is given by

$$\tilde{X}(f) = 2X(f) \quad \text{for} \quad f < 0,$$
$$\tilde{X}(f) = 0 \quad \text{for} \quad f < 0. \tag{24}$$

Thus in order to produce a complex-valued signal from a real-valued signal, one merely sets the negative part of the spectrum to zero, multiplies the positive side by two, and performs the inverse transform:

$$\tilde{x}(t) = \int 2u(f)X(f)\exp(2\pi jtf)\,df \tag{25}$$

where $u(f)$ is a unit step function in the frequency domain.

B. The Bispectrum and the Hilbert Transform

This suggests an approach to the problem of bispectrum magnitude and phase recovery. The question becomes: can the real part of the bispectrum be Hilbert transformed to produce the required imaginary part? To understand this more clearly, we must invoke the Fourier transform relation between the bispectrum and the bicorrelation. Writing the real and imaginary parts explicitly we have

$$\mathrm{Re}[B(f_1,f_2)] = \iint [R_R\cos[2\pi(f_1t_1+f_2t_2)] + R_I\sin[2\pi(f_1t_1+f_2t_2)]]\,dt_1\,dt_2, \tag{26}$$

$$\mathrm{Im}[B(f_1,f_2)] = \iint [R_I\cos[2\pi(f_1t_1+f_2t_2)] - R_R\sin[2\pi(f_1t_1+f_2t_2)]]\,dt_1\,dt_2, \tag{27}$$

where $R(t_1, t_2) = R_R(t_1, t_2) + jR_I(t_1, t_2)$ is the bicorrelation that may be complex valued. The Hilbert transform of $\mathrm{Re}[B(f_1, f_2)]$ is given by

$$\frac{1}{\pi}\int\frac{1}{f'-f_1}\mathrm{Re}[B(f',f_2)]\,df'. \tag{28}$$

Note that this is a Hilbert transform in the frequency variable f_1. The last expression can be rewritten as

$$\iint\left[\frac{R_R(A+A^*)}{2} + \frac{R_I(A-A^*)}{2_j}\right]dt_1\,dt_2, \tag{29}$$

where

$$A = \frac{1}{\pi}\int\frac{\exp[2\pi j(f't_1+f_2t_2)]}{f'-f_1}\,df'. \tag{30}$$

This integral can be evaluated by using contour integration. The contour is taken along the real axis and around an infinite semicircle in the upper half complex f' plane. For the integral to exist, t_1 must be greater than zero. This is the causality requirement associated with the Hilbert transform. When we include the contribution at the pole $f' = f_1$ on the real

axis, we find

$$A = j \exp[2\pi j(f_1 t_1 + f_2 t_2)]u(t_1), \tag{31}$$

where $u(t_1)$ is a unit step function in the time domain. Substituting this back into the expression for the Hilbert transform of the real part of the bispectrum, we obtain

$$\iint [-R_R \sin[2\pi(f_1 t_1 + f_2 t_2)] + R_I \cos[2\pi(f_1 t_1 + f_2 t_2)]]u(t_1)\,dt_1\,dt_2. \tag{32}$$

This equals the imaginary part of the bispectrum under the constraint that $R(t_1, t_2) = 0$ for $t_1 < 0$. A similar Hilbert transform relation exists between the real and imaginary parts of the bispectrum if the bicorrelation is zero for $t_2 < 0$. The implications of these constraints on the bicorrelation can be seen by writing it explicitly in terms of the signals $f(t)$ and $g(t)$:

$$R_{gf}(t_1, t_2) = E[g(t) f(t - t_1) f(t - t_2)]. \tag{33}$$

Treating $f(t)$ as the input to some system and $g(t)$ as the output, the constraints imply that the present outputs cannot be correlated with future inputs. This is a sensible requirement that agrees with our concept of causality. Since physically realizable signals must be causal, this does not appear to be a significant restriction.

C. Approach to Magnitude and Phase Computation

Several authors describe various Hilbert transform-related approaches to phase recovery from general interference patterns. Each of these methods (and ours as well) requires that the spectra of the noninterference and interference fringe terms be separable in the Fourier transform of the interference image. Further, in each approach, a new image is produced in which one of the two sides of the spectrum of the fringes is isolated from the other spectral components.

Takeda *et al.*[24] introduce a fast Fourier transform-based method of general fringe pattern analysis in which the one-sided fringe spectrum is simultaneously isolated and translated to the origin in the frequency domain of the image and then inverse transformed. They illustrate the method for a one-dimensional fringe pattern. Macy[25] compares the Fourier transform approach to an approximate version, called the sinusoidal fitting method, and evaluates the results for two-dimensional phase variations. In a process called quadrature multiplicative moiré, Womack[26] multiplies the interference pattern by a cosine reference image and low-pass filters the result to isolate the difference frequency term. He repeats the process with a sine reference image. Kreis[27] computes the discrete Fourier transform of an interference image, zeros out half of the spectrum, and inverse transforms this to produce a complex-valued image still on a spatial carrier.

A flow chart of our method to compute the magnitude and phase of the bispectrum by the Hilbert transform is shown in Fig. 8. The method involves the convolution of the interference image with the finite impulse response sequence for the Hilbert transform. The Hilbert transformer is a 15-tap finite impulse response filter with alternating zero coefficients.[23] The magnitude and phase responses for the transformer are shown in Figs. 9(a) and 9(b), respectively. Note that the magnitude response has a bandpass character and the phase response exhibits the required $-j\,\mathrm{sgn}(f)$ profile.

We designate the isolated interference term that represents the real part of the bispectrum as $\dot{B}_r(f_1, f_2)$, where it is to be understood that the frequency variables take on discrete values corresponding to discrete image pixel positions. We designate the imaginary part of the bispectrum as $\dot{B}_i$, and we write the Hilbert transform as a convolution in which each column of the image (real part of the bispectrum) is convolved with the transform coefficients:

$$\dot{B}_i(f_1, f_2) = \sum_{n=1}^{15} h(f1 - n + 8)\dot{B}_r(n, f_2) \tag{34}$$

Note that we have also included the effect of the seven sample filter delay. Strictly speaking, the new complex image still differs from the bispectrum because of the phase factor that is responsible for the fringes. The spatial frequency of the fringes is determined by Fourier transforming the isolated interference frame for a reference test, computing the magnitude of the

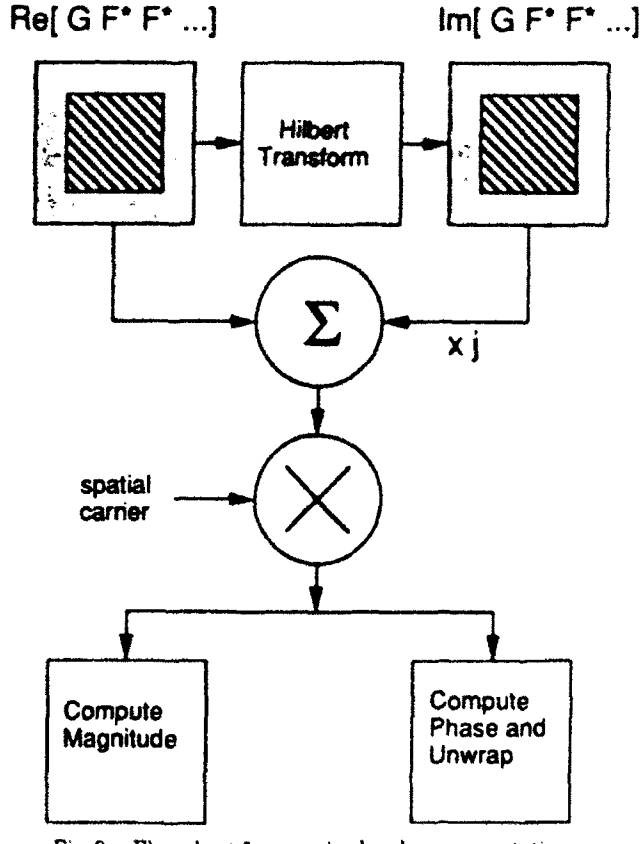


Fig. 8. Flow chart for magnitude–phase computation.

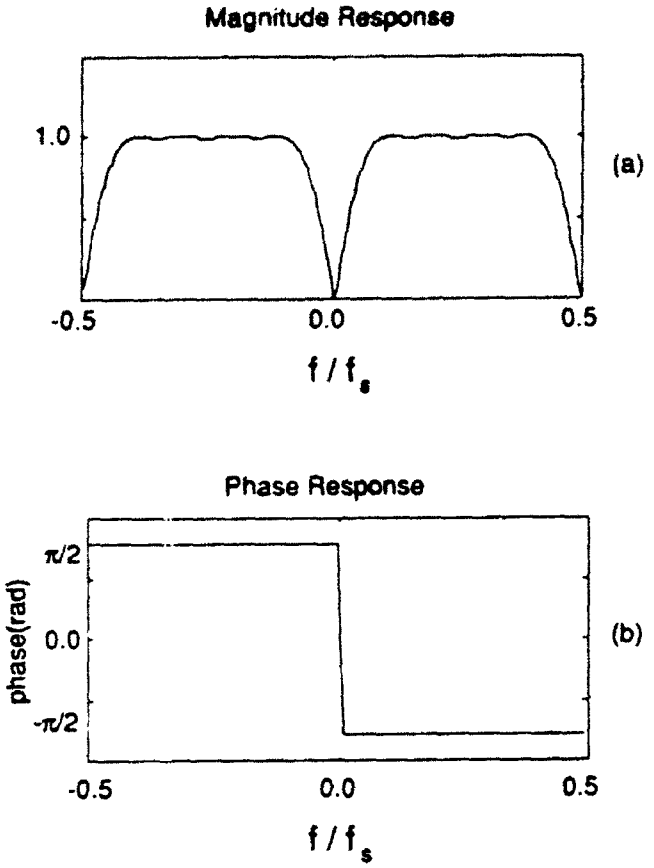


Fig. 9. Magnitude and phase response for the Hilbert transformer.

new function (which is related to the bicorrelation function), and locating the peaks associated with the interference fringes. This spatial carrier is mixed out of the complex bispectrum image producing a baseband image. The magnitude and (wrapped) phase of the bispectrum are computed from the standard formulas

$$B = (B_r^2 + B_i^2)^{1/2}, \tag{35}$$

$$\phi_B = \arctan \frac{B_i}{B_r}. \tag{36}$$

The two-dimensional phase is unwrapped by first unwrapping along a single row passing through the middle of the image. Each column is then unwrapped by working upward and downward from the central row. The removal of the spatial carrier suppresses the fringe information. Otherwise the unwrapped phase would exhibit a slope associated with the fringes that are superimposed upon the true bispectrum phase.

In this section we have shown that a Hilbert transform relation exists between the real and imaginary parts of the bispectrum, given reasonable causality constraints. We have further described an approach for recovering the magnitude and phase of the bispectrum for wide bandwidth waveforms. This process includes the use of a discrete Hilbert transform on images produced by the optical realization of the bispectrum. The method is quite general in that the only requirements of the image are that it be sufficiently sampled and that the bispectrum information can be separated from other components of the interference image. The performance of the method for several test cases is the subject of the next section.

V. Test Results

The Hilbert transform approach to bispectrum magnitude and phase recovery was tested with four basic test cases. The first two tests emphasized the magnitude estimation capability. The next two focused on the recovery of bispectrum phase. Four different quadratic transfer functions were used to generate the magnitude and phase profiles. For each case presented here, rf signals were generated as described in Section III and introduced into the optical system through the AO modulators. The output images were digitized and subsequently processed to estimate magnitude and phase.

A. Bispectrum Magnitude Recovery

The spectral envelopes of the signals used in the first magnitude test are shown in Fig. 6(b). The signal $f(t)$ is a series of sinc pulses 5 μs apart. In the frequency domain, the spectrum $F(f)$ is a set of uniform amplitude, constant-phase tones 200 kHz apart, with a total bandwidth of 6 MHz. The signal $g(t)$ is the output of a square-law device followed by a normalizing operation. Its spectrum is also uniform in amplitude, and it has a bandwidth of 12 MHz. $F(f)$ and $G(f)$ are quadratically related to each other and their cross bispectrum should be nonzero and uniform in amplitude over a 6 MHz × 6 MHz region of the bifrequency plane.

The isolated interference term (representing the real part of the bispectrum on a spatial carrier) for the first magnitude test is shown as a gray-scale plot in

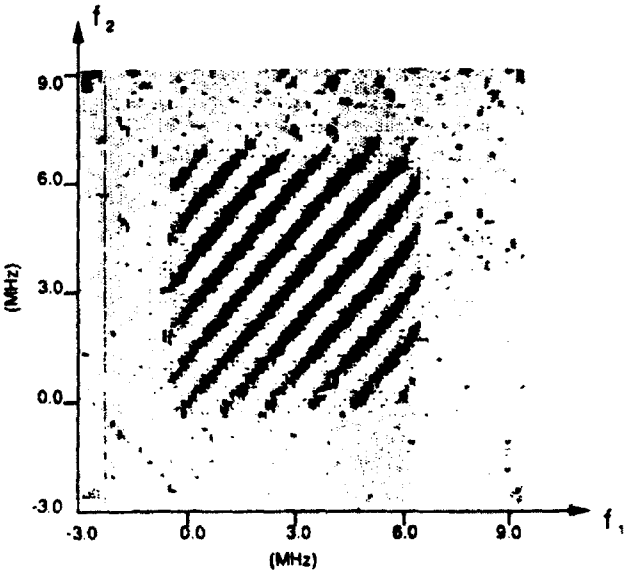


Fig. 10. Gray-scale plot of the isolated interference term for the first magnitude test.

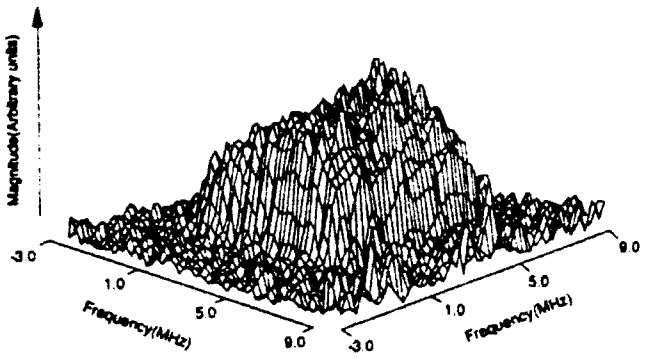


Fig. 11. Magnitude surface of the bispectrum for the first magnitude test.

Fig. 10. The fringes do appear to be fairly uniform. A three-dimensional plot of the magnitude of the bispectrum and a cross-sectional plot along the line $f_1 = 3$ MHz are shown in Figs. 11 and 12, respectively. These reflect the essential characteristics of the output we expect. Nonuniformity of the amplitudes of the interfering beams, a slight nonuniformity of focus across the image plane, and background noise contribute to the distortion of the image.

The second magnitude test employs the same input signal $f(t)$ as the first, but the signal $g(t)$ is the direct output of the square-law device without normalization. The spectral envelopes of these signals are shown in Fig. 6(a). Now $G(f)$ is triangular in form, representing the convolution of $F(f)$ with itself. This shape determines the expected cross-bispectrum output, which will look somewhat like a tent whose peak lies on the diagonal line $f_1 + f_2 = 6$ MHz.

Figure 13 shows the isolated interference term for the second magnitude test in the form of a gray-scale plot. Note that the upper right and lower left corners of the fringes are beginning to fade out. The bispectrum magnitude is plotted in Fig. 14. Cross-sectional plots along the lines $f_1 = f_2$ and $f_1 + f_2 = 6$ MHz are shown in Figs. 15(a) and 15(b), respectively. The first cross section clearly exhibits the triangular magnitude profile carried by $G(f)$. The second, along the

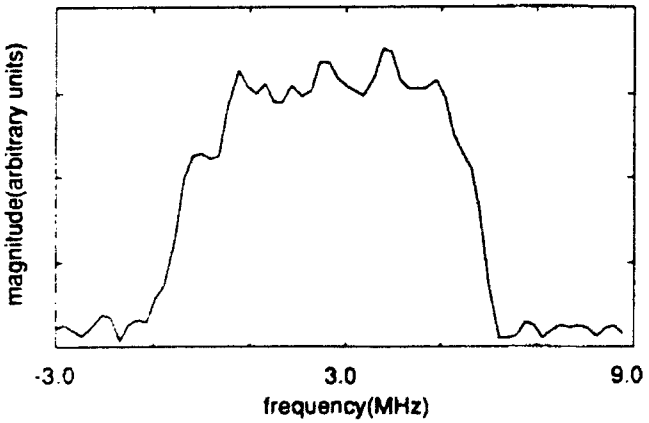


Fig 12. Cross section of the magnitude of the bispectrum from the first magnitude test ($f_1 = 3$ MHz).

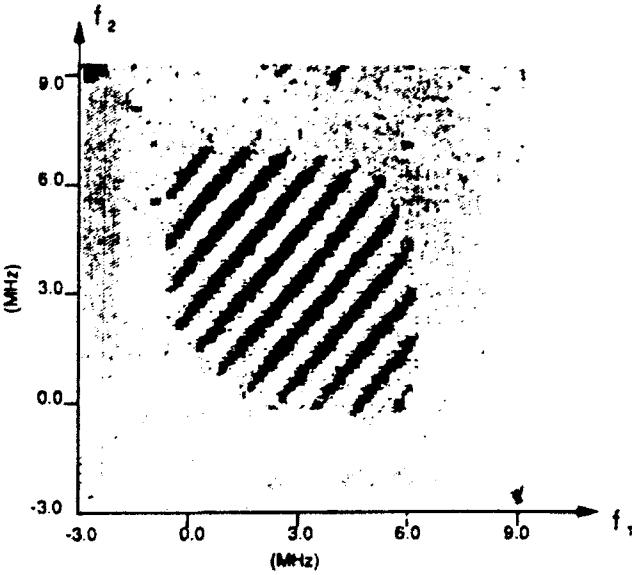


Fig. 13. Gray-scale plot of the isolated interference term for the second magnitude test.

peak of the tent, has a more rectangular form, as expected.

B. Bispectrum Phase Recovery

The phase test cases are based on signals with the same magnitude spectra as those of the first magnitude test. That is, $F(f)$ and $G(f)$ are 6- and 12-MHz wide, respectively, they are quadratically related, and they have uniform amplitudes in the frequency domain. The difference is that $G(f)$ no longer has a uniform phase profile. Specifically, its phase profile is parabolic, as illustrated in Fig. 6c. The expected form of the magnitude of the bispectrum should generally be the same as that of the first magnitude test. The bispectrum phase (unwrapped) should be a parabolic surface whose axis lies along the line $f_1 + f_2 = 6$ MHz.

The interference frame for the first phase test is shown in Fig. 16. Note that the spatial frequency of the fringes changes as one moves along the diagonal of the plot. The unwrapped phase surface for this test is shown in Fig. 17. Recall that prior to the phase computation the spatial carrier was mixed out. The domain of the plot is limited to the 6 MHz × 6 MHz

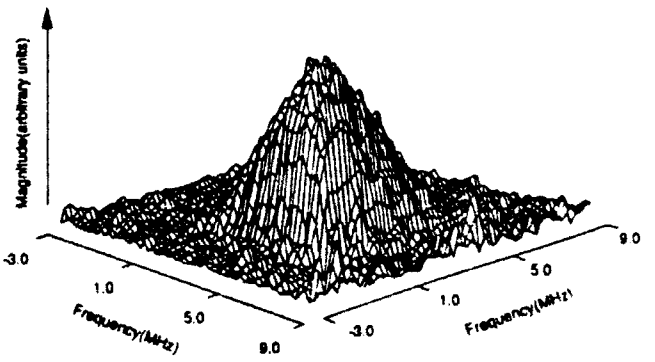


Fig. 14. Magnitude surface of the bispectrum for the second magnitude test.

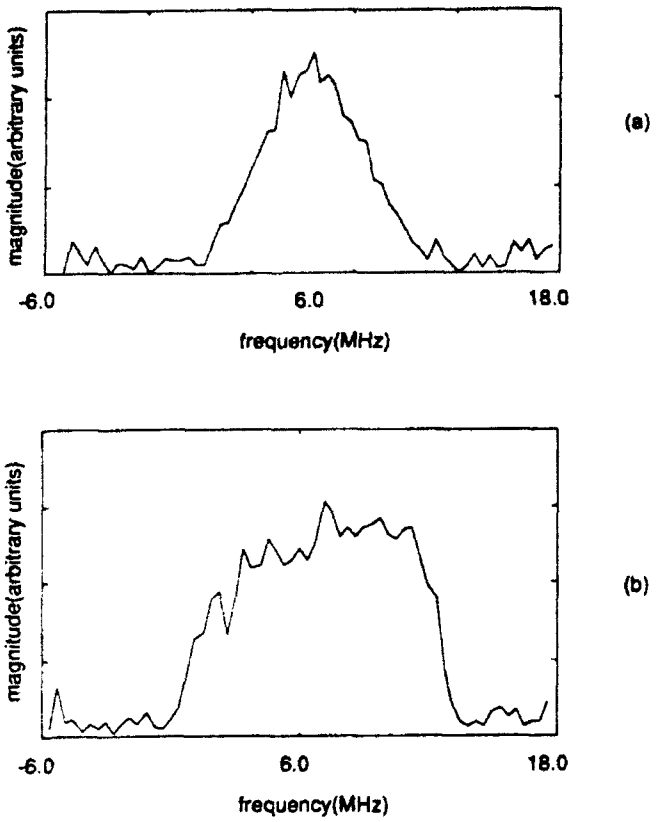


Fig. 15. Cross section of the magnitude of the bispectrum from the second magnitude test along (a) $f_1 = f_2$ and (b) $f_1 + f_2 = 6$ MHz.

region, where the bispectrum is nonzero. This results in a less confusing picture. The characteristic parabolic profile can be seen in Fig. 17 but becomes much more obvious in Fig. 18—a cross-sectional plot of the phase surface along $f_1 = f_2$. The equation used to

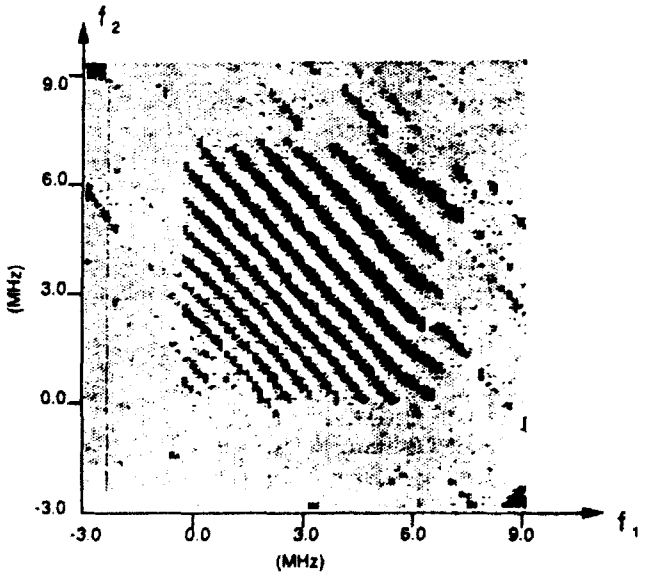


Fig. 16. Gray-scale plot of the isolated interference term for the first phase test.

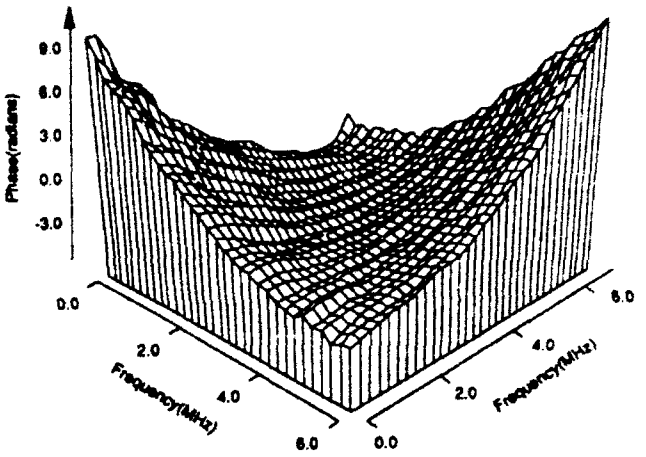


Fig. 17. Phase surface of the bispectrum for the first phase test.

generate the quadratic phase term is given by

$$\phi_q = \pi(f_1 + f_2 - 6\ \text{MHz})^2/10 \tag{37}$$

and is shown as a dashed curve. The solid curve depicts the measured phase values. Generally, the agreement between the two curves is quite good. There is a slight separation of the curves in the range from 0 to 4 MHz. This corresponds to the lower left corner of the interference pattern in Fig. 16. Note that the interference fringes tend to fade out in this region of higher fringe frequency. Therefore, the magnitude of the bispectrum is reduced in the region. The agreement between the measured phase values and the true phase profile is poorer due to the reduced signal-to-background noise ratio. This is the same effect as that seen with the loss of contrast in the time delay tests.[28] In the case of time delay, it was shown that the Gaussian beam profile acts as a weighting function that reduces the magnitude of the correlation between signals as the delay is increased. Unlike the effects of time delay, which modify the fringe frequency and contrast for the entire 6 MHz × 6 MHz region, the quadratic phase profile produces a position-dependent change in fringe frequency and contrast.

The second phase test employed a signal whose spectrum had a three-phase step profile as indicated

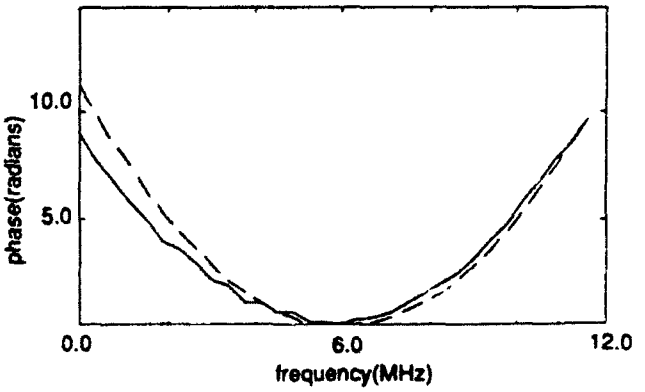


Fig. 18. Cross section of the phase of the bispectrum from the first phase test ($f_1 = f_2$).

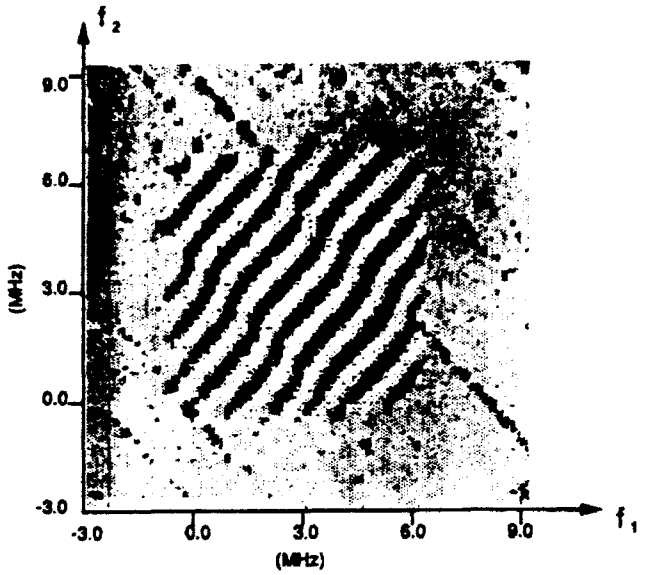


Fig. 19. Gray-scale plot of the isolated interference term for the second phase test.

by Fig. 6(d). The phase difference between successive steps was 120 deg (2.094 rad). The effects of this phase profile are clearly manifested in Fig. 19, which shows the interference frame for the test. Note that the fringe orientation was changed in this test so that the fringes met the phase steps at roughly a right angle. This makes it easier for the reader to see the effect of the phase transition in the interference pattern. It also illustrates that the procedure works regardless of fringe orientation. The phase surface for this test and a cross-sectional slice are shown in Figs. 20 and 21, respectively. The solid curve represents the measured phase values while the dashed curve is the expected phase profile. The phase values shown in Fig. 21 are averaged across each step. The phase differences between adjacent steps (for this particular slice) aare measured as 112 deg (1.95 rad) and 107 deg (1.86 rad). Both estimates are within 11% of the true value of 120 deg. The phase steps are slightly smaller and the transitions less sharp than we might expect.

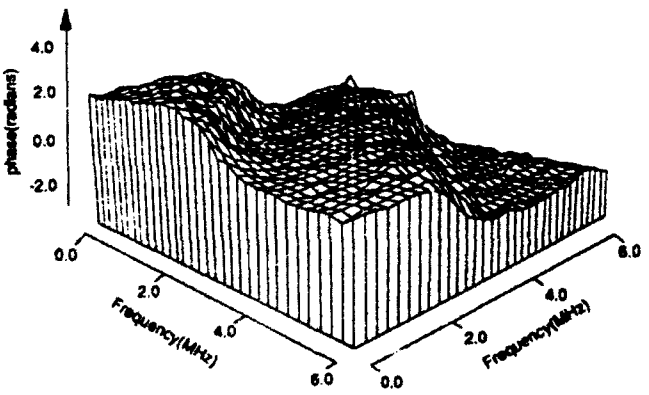


Fig. 20. Phase surface of the bispectrum for the second phase test.

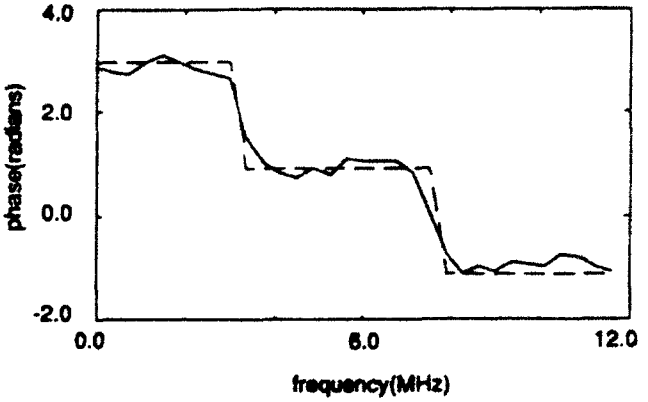


Fig. 21. Cross section of the phase of the bispectrum from the second phase test ($f_1 = f_2$).

This leads us to suspect that our interference term isolation procedure has filtered out more of the spectrum of the image than we intended for this particular case.

VI. Summary

The primary goal of the present research effort was to compute bispectrum magnitude and phase for wide bandwidth input waveforms. This has been accomplished through enhancements to an existing optical architecture, extension of the input signal generation capability, and through more intensive digital processing of the output images. The improvements to the optical configuration provided the capability of satisfying the frequency sum rule over an extended region of the bifrequency plane. A new signal generation approach was incorporated that is capable of creating wide bandwidth test signals, which are related by arbitrary quadratic transfer functions. Both developments are necessary for the implementation and testing of a wide bandwidth bispectrum processor.

The digital processing of the interference images that contain the bispectrum information took advantage of the spectral characteristics of the images. The interference term containing the real part of the bispectrum on a spatial carrier was isolated from the noninterference background terms by high-pass filtering the image. This avoided several shortcomings associated with an earlier approach in which multiple frames must be acquired and subtracted. The imaginary part of the bispectrum was recovered by performing a Hilbert transform on the isolated interference term. This is possible given the constraint that the signals being analyzed are causally related to each other. The magnitude and phase of the bispectrum were computed from actual interference images for several quadratic transfer functions. The magnitude tests were in good qualitative agreement with expectations. The agreement of the phase tests was excellent.

Future testing should be aimed at determining the performance limits of the optical realization of the bispectrum in terms of the accuracy of time delay[28,29] and magnitude and phase estimates. The ability of

the bispectrum to suppress additive Gaussian noise should be investigated in the context of the optical processor. The kind of rigorous testing envisioned requires that the processor be further optimized in several different areas. The quality of the optical output must be improved. This involves both the optics and the rf circuitry used to drive the modulators. The digital filters and Hilbert transformer used in this work were selected primarily for convenience. Some effort will be necessary to determine the optimal digital filters for the specific processor applications and performance requirements. The speed of operation is limited at present by the image acquisition and processing hardware. A reasonable and realistic goal for the processor in the near future would be to compute the bispectrum at a rate of 30 frames/s for a 10×10 MHz region of the bifrequency plane. This would involve algorithm optimization as well as hardware enhancements. In the future, we might consider other variations of the implementation of the bispectrum that put more of the computational load onto the optics. Since the bispectrum information exists as an interference term, one interesting concept would be to record this as a hologram from which the bispectrum information could be reconstructed as an amplitude distribution and subsequently processed optically. Another possibility is the use of an optical phase-modifying element that would allow simultaneous recovery of the real and imaginary parts of the bispectrum.

This research was supported in part through a Fannie and John Hertz Foundation Graduate Fellowship and a University of Texas Research Institute Grant. E. J. Powers' work on higher-order spectral analysis is supported by the U.S. Department of Defense Joint Services Electronics Program under contract AFOSR F49620-89-C-0044.

References

1. C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: a digital signal processing framework," Proc. IEEE **75**, 869–891 (1987).
2. W. T. Rhodes, "Acousto-optic signal processing: convolution and correlation," Proc. IEEE **69**, 65–79 (1981).
3. T. M. Turpin, "Spectrum analysis using optical processing," Proc. IEEE **69**, 79–92 (1981).
4. J. N. Lee, "Optical architectures for temporal signal processing," in *Optical Signal Processing*, J. L. Horner, ed. (Academic, San Diego, Calif., 1987), pp. 165–190.
5. J. Wilbur and F. J. Taylor, "An acousto-optic Wigner processor for time-varying signal analysis," Proc. IEEE. Lett. **75**, 427–428 (1987).
6. A. A. Canas, "The acousto-optic triple product processor and its applications," Opt. Quantum Electron. **19**, 79 (1987).
7. M. H. Kauderer, M. F. Becker, and E. J. Powers, "Acousto-optical bispectral processing," Appl. Opt. **28**, 627–637 (1989).
8. J. M. Florence and J. H. Song, "Real-time acousto-optic bispectral processor," in *Real Time Signal Processing VI*, K. Bromley, ed., Proc Soc Photo-Opt *Instrum Eng* **431**, 284 (1983).
9. Y. C. Kim and E. J. Powers, "Digital bispectral analysis and its applications to nonlinear wave interactions," IEEE Trans Plasma Sci. **PS-7**, 120–131 (1979).
10. E. J. Powers and R. W. Miksad, "Polyspectral measurement and analysis of nonlinear wave interactions," in *Nonlinear Wave Interactions in Fluids*, R. W. Miksad, T. R. Akylas, and T. Herbert, eds. (American Society of Mechanical Engineers, New York, 1987), pp. 9–16.
11. K. I. Kim, E. J. Powers, C. P. Ritz, R. W. Miksad, and F. J. Fischer, "Modeling of the nonlinear drift oscillations and moored vessels subject to non-Gaussian random sea-wave excitation," IEEE J. Ocean Eng. **OE-12**, 568–575 (1987).
12. K. I. Kim and E. J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," IEEE Trans. Acoust. Speech Signal Process. **36**, 1758–1769 (1988).
13. A. W. Lohmann and B. Wirnitzer, "Triple correlations," Proc IEEE **72**, 889–901 (1984).
14. J. L. Whitton, S. M. Elgie, H. Kugel, and H. Moldofsky, "Genetic dependence of the electroencephalogram bispectrum," Electroencephalogr. Clin. Neurophysiol. **60**, 293–298 (1985).
15. K. Sasaki, T. Sato, and Y. Nakamura, "Holographic passive sonar," IEEE Trans. Sonics Ultrason. **SU-24**, 193–200 (1977).
16. T. Sato and K. Sasaki, "Bispectral holography," J. Acoust. Soc. Am. **62**, 404–408 (1977).
17. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968).
18. M. Schetzen, *The Volterra and Weiner Theories of Nonlinear Systems*, (Wiley, New York, 1980).
19. J. S. Bendat and A. G. Piersol, "Spectral analysis of non-linear systems involving square-law operations," J. Sound Vib. **81**, 199–213 (1982).
20. A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, N. J., 1975).
21. L. R. Rabiner and R. W. Schafer, "On the behavior of minimax FIR digital Hilbert transformers," Bell Syst. Tech. J. **53**, 363–390 (1974).
22. D. Gabor, "Theory of communication," J. Inst. Electr. Eng Part 3 **93**, 429–457 (1946).
23. J. D. Jackson, *Classical Electrodynamics* (Wiley and Sons, New York, 1975).
24. M. Takeda, H. Ina, and S. Kobayashi, "Fourier transform method of fringe pattern analysis for computer-based topography and interferometry," J. Opt. Soc. Am. **72**, 156–160 (1982).
25. W. M. Macy, "Two-dimensional fringe pattern analysis," Appl. Opt. **22**, 3898–3901 (1983).
26. K. H. Womack, "Interferometric phase measurement using spatial synchronous detection," Opt. Eng. **23**, 391–395 (1984).
27. T. Kreis, "Digital holographic interference-phase measurement using the Fourier transform method," J. Opt. Soc. Am. A **3**, 847–855 (1986).
28. S. K. Kniffen, "Acousto-optical bispectrum processing of wide bandwidth signals," Ph.D. dissertation (The University of Texas at Austin, Austin, Tex., 1990; available from University Microfilms, Ann Arbor, Michigan).
29. S. K. Kniffen, M. F. Becker, and E. J. Powers, "Applications of an acousto-optical bispectrum processor," in *Optical Information Processing Systems and Architectures II*, B. Javidin, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1347**, 131–143 (1990).

J. Fluid Mech. (1992), vol. 236, pp. 385–413
Printed in Great Britain

Subharmonic growth by parametric resonance

By M. R. HAJJ, R. W. MIKSAD AND E. J. POWERS
College of Engineering, The University of Texas at Austin, Austin, TX 78712, USA

(Received 29 November 1990 and in revised form 18 June 1991)

An experimental investigation is conducted in order to quantify the nonlinear and parametric resonance mechanisms that are associated with the subharmonic growth in the transition to turbulence in plane mixing layers. Higher-order digital statistical analysis techniques are used to investigate the nonlinear and parametric mechanisms responsible for the energy transfer to the subharmonic. The results show that the dominant interaction is a parametric resonance mechanism between the fundamental and the subharmonic modes which leads to a pronounced growth of the subharmonic. Measurements also indicate that the fundamental, besides interacting with the subharmonic, is also engaged in redistributing its energy to the other Fourier components of the flow via nonlinear three-wave interactions. Local wavenumber measurements verify that frequency–wavenumber resonance matching conditions exist between the fundamental and subharmonic in the region where the subharmonic gains its energy by parametric resonance. The results are in general agreement with theoretical models by Kelly (1967), and Monkewitz (1988) on subharmonic growth.

1. Introduction

Two important features of the transition to turbulence in plane mixing layers are the subharmonic growth and vortex pairing. The emergence and subsequent growth of a flow component at the subharmonic frequency, beyond the initial linear instability region, is a good example of secondary instability, and represents the first step in the sequence of instabilities that lead to the final breakdown to turbulence. Evidence of the growth of the subharmonic mode has been given in many experimental investigations, such as those of Sato (1959), Browand (1966), Miksad (1972) and many others. The mechanisms of vortex roll-up and subsequent pairing represent the vorticity restructuring in the transitioning mixing layer. Evidence of vortex pairing has been observed in the flow visualization pictures of Winant & Browand (1974) and Brown & Roshko (1974). Ho & Huang (1982) used hydrodynamic instability-wave concepts and flow visualization to compare the development of the fundamental and subharmonic modes to the vortex pairing mechanism in mixing layers. Their results showed that the saturation of the fundamental mode accompanies the vortex roll-up process, and that the saturation of the subharmonic mode accompanies the vortex merging. Also, the analysis of Pierrehumbert & Widnall (1982) showed that the emergence of a subharmonic component, in the two-dimensional case, may correspond to the pairing of two neighbouring vortices. Ho (1982) suggested a simplified model in which the mixing layer grows only by the vortex pairing mechanism. In general, the enhancement or suppression of the spreading mixing layer seems to be dependent on the ability to control this vorticity restructuring mechanism and consequently the subharmonic

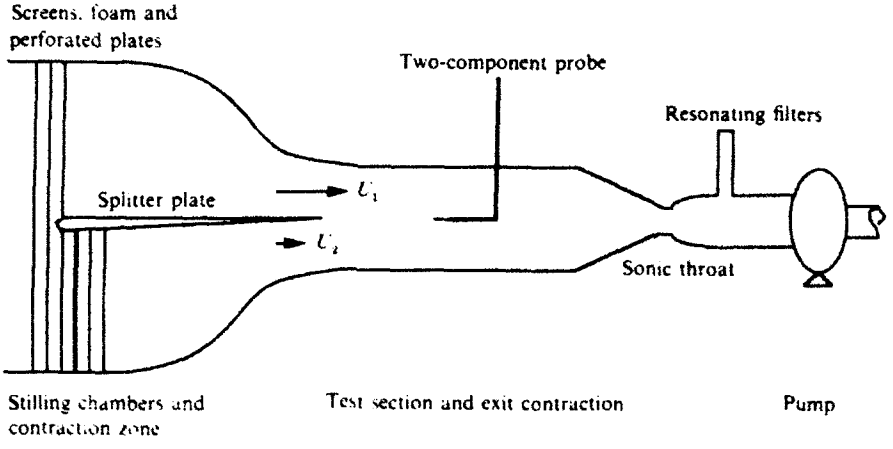


FIGURE 1. Schematic sketch of the wind tunnel apparatus.

condition in the wavenumber domain for efficient energy exchange via *parametric* resonance is established.

2. Experimental set-up

The experiments were conducted in a low-turbulence subsonic wind tunnel with a $30 \times 20 \times 150$ cm test section, see figure 1. The mixing layer was formed by merging two laminar streams as shown in figure 2. The upper and lower free-stream velocities were $U_1 = 7.17$ and $U_2 = 1.51$ m/s, respectively. This resulted in a velocity differential $\Delta U = 5.66$ m/s and a velocity ratio $R = (U_1 - U_2)/(U_1 + U_2) = 0.65$. The initial Reynolds number based on initial momentum thickness and upper free-stream velocity was 305. The free-stream turbulence intensity in the vicinity of the trailing edge of the splitter plate was $0.0005\Delta U$ in the high-speed stream, and $0.0011\Delta U$ in the low-speed stream. Most of the background turbulence was concentrated in frequencies below 40 Hz. Streamwise velocity fluctuations were measured using a DISA 56N/C hot-wire anemometry system. A special probe (Jones *et al.* 1988) with two sensing elements was used in these experiments to measure the spectral energy exchanges between two downstream locations, figure 2. The two sensing elements are separated in the streamwise direction by a distance $\Delta x = 0.1$ cm and in the spanwise direction by 0.15 cm, centre to centre. The cross-stream separation is zero. The small spanwise separation is used to isolate the downstream wire from the wake of the upstream wire. The hot-wire signals were DC and anti-alias filtered and were sampled with a CAMAC digital data acquisition system. The sampling frequency was set at 1000 Hz. The results shown here are for natural transition excited by random fluctuations in the flow. No external forcing was used. The measured frequency of the dominant instability mode of this transition was $f_0 = 215$ Hz.

3. Mean flow properties

The time-averaged flow properties were measured at midspan. The average velocity, $\bar{U} = \frac{1}{2}(U_1 + U_2)$ is equal to 4.34 m/s. The initial momentum thickness of the high-speed-side free-stream velocity, θ_0, is equal to 0.064 cm and the streamwise

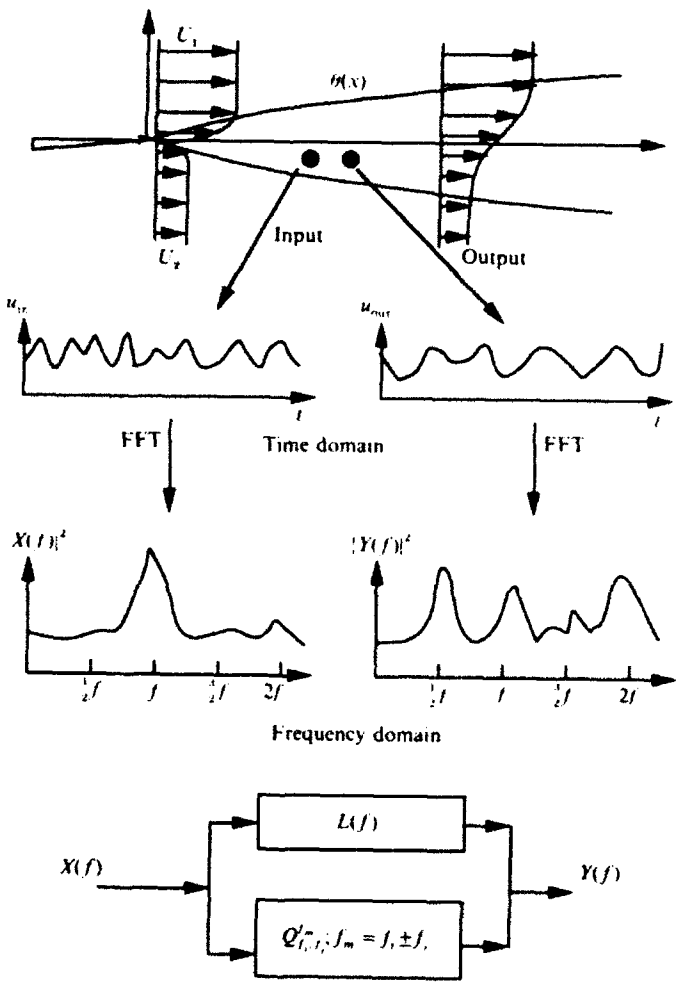


FIGURE 2. Schematic of the mixing layer and experimental approach.

wavelength of the dominant instability mode, λ_a, is equal to 1.98 cm. The Strouhal number of the dominant instability mode, $f_0\theta_0/\bar{U}$, is equal to 0.0317. In order to compare our results with data taken in experiments at different values of R, lengthscales are normalized with the ratio $R/\lambda_0 = 0.32$ as suggested by Huang & Ho (1990). The variations in the cross-stream direction of the longitudinal mean velocity and total r.m.s. fluctuations at different downstream locations are shown in figure 3. The mean velocity profiles are normalized according to the relation $2(\bar{U}(y)-\bar{U})/\Delta U$. The velocity fluctuations are normalized with ΔU. The mean flow profiles show a combination of a wake and a mixing layer up to $Rx/\lambda_0 = 1.0$. Beyond this location, they become similar to a hyperbolic tangent profile.

The vertical spreading of the mixing layer is measured by the local momentum thickness, defined as

$$\theta = \frac{1}{\Delta U^2}\left\{\int_{-\infty}^{0} \bar{U}(\bar{U}-U_2)\,dy + \int_{0}^{+\infty} \bar{U}(U_1-\bar{U})\,dy\right\}. \tag{1}$$

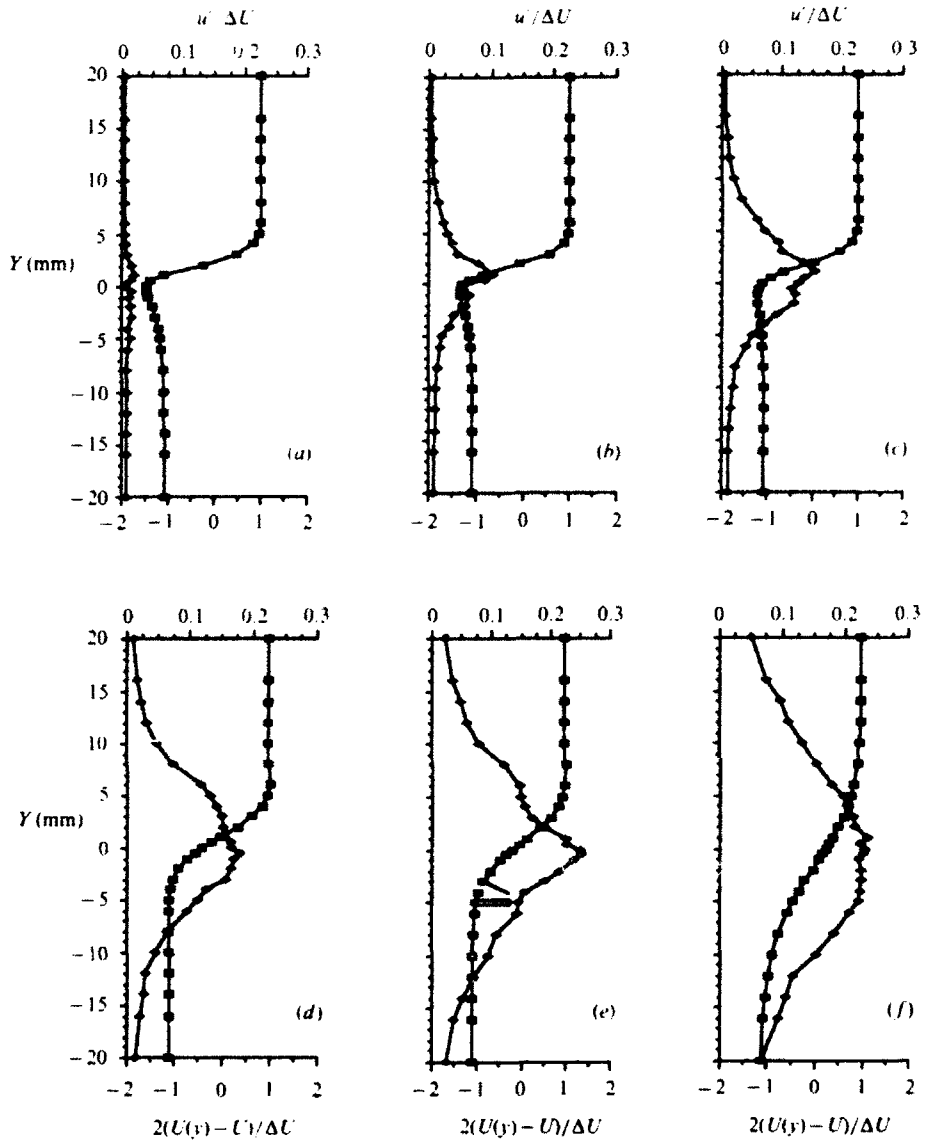


FIGURE 3. Cross-stream variations of the normalized mean velocity, $2(U(y)-U)/\Delta U$ (--⊡--), and longitudinal fluctuations, $u'/\Delta U$ (—◆—) at different downstream locations: (*a*) $Rx/\lambda_0 = 0.5$; (*b*) 1.0; (*c*) 1.5; (*d*) 2.0; (*e*) 2.7; (*f*) 5.0.

Similarly, an energy thickness, defined as

$$\epsilon = \frac{1}{\Delta U^3}\left\{\int_{-\infty}^{0} U(U^2-U_2^2)\,\mathrm{d}y + \int_{0}^{+\infty} U(U_1^2-U^2)\,\mathrm{d}y\right\} \tag{2}$$

is used as a measure of the flux of energy from the mean flow. The downstream development of the maximum u'_{rms} of the fundamental and subharmonic modes are shown in figure 4. The development of the energy thickness and the momentum thickness are shown in figure 5.

In figure 4, four basic regions of fundamental and subharmonic development are

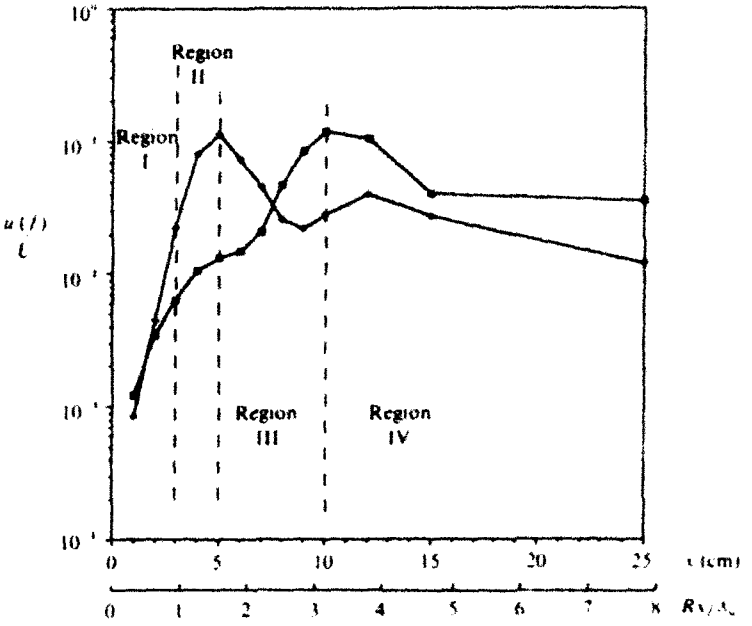


FIGURE 4. Variation of the r.m.s. amplitude of the fundamental (⊡) and subharmonic (◆) modes along max. u'_{rms} in the streamwise direction

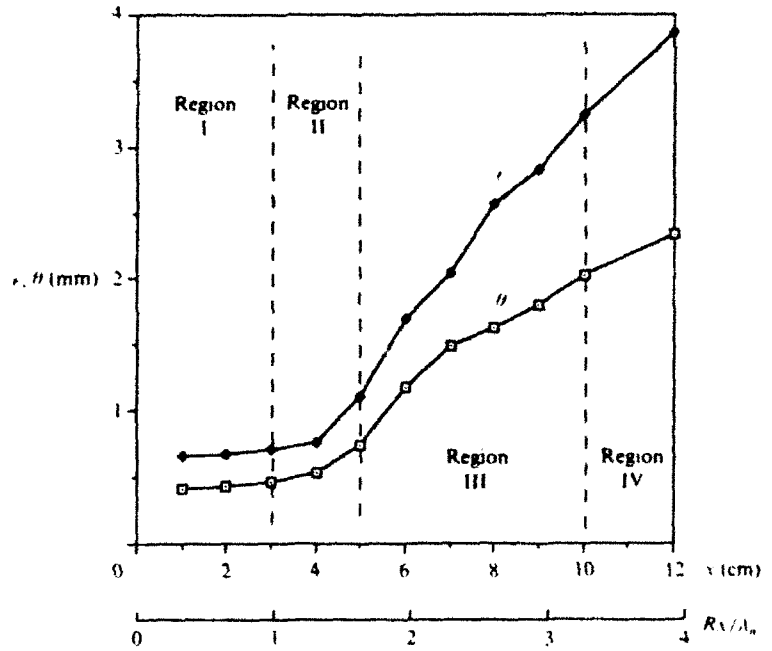


FIGURE 5. Momentum (—⊡—) and energy (—◆—) thicknesses, θ and δ, respectively, of the mixing layer *vs.* streamwise direction.

evident. For $Rx/\lambda_0 < 1.0$, the fundamental and subharmonic modes grow exponentially. The non-dimensional growth rate of the fundamental and the subharmonic, based on θ_0, are 0.10 and 0.047, respectively. These values are within 6% of the values predicted by the spatial linear instability analysis of Monkewitz &

Huerre (1982). The region limited by $Rx/\lambda_0 < 1.0$ represents the initial instability region where the fluctuations components grow exponentially, and will be referred to as Region I. At about $Rx/\lambda_0 = 1.0$, the subharmonic starts deviating from exponential growth. This location is chosen for reference because it is the last location at which the subharmonic amplitude deviates by less than 5% from the amplitude predicted by its initial exponential growth. Note that the choice of this criterion is subjective and is stated here for the sake of comparison. At this location, the r.m.s. amplitude of the longitudinal fluctuations at the fundamental frequency has an average value of $u'_{rms}/\bar{U} = 0.012$ (peak value = 0.016). This value is almost twice the critical fundamental amplitude of $0.0153\alpha^2 = 0.006$ predicted by Monkewitz (1988).

At $Rx/\lambda_0 = 1.6$, the fundamental achieves a saturation level of $0.113\bar{U}$. This level is one order of magnitude larger than that of the critical fundamental amplitude. Notice that up this location, figure 5 indicates that the vertical spreading of the mixing layer and the extraction of energy from the mean flow are very slow. The region between $Rx/\lambda_0 = 1.0$ and 1.6 represents a region where the fundamental mode continues to grow at a rate less than that predicted by the linear theory, to eventually equilibrate into finite-amplitude oscillations. This region will be referred to as Region II.

Between $Rx/\lambda_0 = 1.6$ and 3.2, the r.m.s. amplitude of the longitudinal fluctuations at the fundamental frequency decreases. However, the subharmonic component undergoes a second region of growth and achieves a saturation level of the same order of magnitude as the fundamental. The region limited by $Rx/\lambda_0 = 1.6$ and 3.2 will be referred to as the secondary instability region (Region III). In figure 5, we notice that a large spreading of the mixing layer also occurs over this region. The momentum thickness doubles between $Rx/\lambda_0 = 1.6$ and 3.2 indicating that the extraction of energy from the mean flow is increasing. If Kelly's (1967) interaction mechanism is correct, this implies that the energy is being extracted from the mean flow and transferred to the subharmonic by a parametric resonance mechanism. Beyond $Rx/\lambda_0 = 3.2$, the r.m.s. amplitudes of the longitudinal fluctuations at the fundamental and subharmonic modes start decreasing. This could be the result of the second subharmonic, $\frac{1}{4}f_0$, or the onset of the three-dimensional activities. This region will be referred to as Region IV.

4. Fundamental–subharmonic interaction

The primary aim of this investigation is to quantify the spectral energy exchanges arising from parametric or nonlinear wave–wave interactions that are responsible for the subharmonic growth in a plane mixing layer. These interactions can best be studied by using polyspectral analysis techniques. The role of these interactions in reinforcing the subharmonic can be quantified by measuring quantities such as the linear and nonlinear transfer functions. The connection between existing theoretical models of subharmonic growth and our use of the transfer function measurements is not obvious and the following heuristic model is presented to motivate this approach.

4.1. *Fundamental–subharmonic energy transfer*

As shown by many investigators and detailed by Mollo-Christensen (1971), the turbulent energy transfer from the mean flow to the primary instability components, and from the primary instability to the secondary instability components, and vice versa, occurs because of nonlinear terms that are present in the equations governing

energy exchange. The importance of these terms to the subharmonic generation process is evident in the equation governing energy exchanges between the different components of the velocity fluctuations. Stuart (1960). Monkewitz (1988), based on the experimental results of Ho & Huang (1982) and the discussion of Ho & Huerre (1984), assumed that subharmonic behaviour in mixing layers can be considered to be locally parallel in the saturation region of each mode in the subharmonic sequence. Kelly (1967) made a similar assumption in his temporal analysis. Following Monkewitz (1988) and Kelly (1967), the effect of interacting components on the growth of the subharmonic component in a parallel flow can be written as

$$\left\{\frac{\partial}{\partial t}\nabla^2+(U\nabla^2-U'')\frac{\partial}{\partial x}\right\}\psi(f_m)=\sum_{f_i\pm f_j=f_m}\sum\left\{-\left(\frac{\partial\psi(f_i)}{\partial y}\frac{\partial}{\partial x}-\frac{\partial\psi(f_i)}{\partial x}\frac{\partial}{\partial y}\right)\nabla^2\psi(f_j)\right.$$
$$\left.-\left(\frac{\partial\psi(f_j)}{\partial y}\frac{\partial}{\partial x}-\frac{\partial\psi(f_j)}{\partial x}\frac{\partial}{\partial y}\right)\nabla^2\psi(f_i)\right\},\quad f_m=f_i\pm f_j. \tag{3}$$

where $\psi(f_m)$ represents the subharmonic component of the stream function and $\psi(f_i)$ and $\psi(f_j)$ represent any two components whose interactions could lead to the subharmonic reinforcement. Note that the summation on the left-hand side is necessary to take into account all sum and difference interactions that could affect the subharmonic component.

The analyses of Kelly (1967) and Monkewitz (1988) showed that the manifestations of the parametric and nonlinear resonance mechanisms are different. In the nonlinear resonance mechanism, energy is passed to the subharmonic directly from the fundamental. In the parametric resonance mechanism, energy is not directly exchanged with the fundamental, but passed from the mean flow to the subharmonic as a result of the parametric variations introduced by the fundamental. These differences lead to different differential equations. While the nonlinear resonance mechanisms are modelled by nonlinear terms in the governing equations, the parametric resonance mechanism is a linear-like phenomenon in that it is modelled by linear terms with periodic coefficients resulting from the periodic distortions of the mean flow by the fundamental. It is important to note here that because coupling of the subharmonic with the fundamental mode is either directly or indirectly involved in both types of mechanisms, the phase coupling and wavenumber–frequency resonance conditions for the parametric interaction are similar to those required for nonlinear interactions.

Itoh (1977) showed that by decomposing the Fourier components of the stream function into a sum of linear eigenfunctions, (3) can be expressed as a wave coupling equation of the form

$$\frac{\mathrm{d}A(f_m)}{\mathrm{d}x}=\alpha(f_m)A(f_m)+\sum_{f_i\pm f_j=f_m}\sum\beta^{f_m}_{f_i f_j}A(f_i)A(f_j),\quad f_m=f_i\pm f_j. \tag{4}$$

where $A(f)$ is the amplitude of the linear eigenfunction of the component with frequency f.

This coupling equation describes the spectral change of the complex amplitude of the frequency component f_m due to linear and quadratically nonlinear mechanisms. $\alpha(f_m)$ represents the linear growth, and $\beta^{f_m}_{f_i f_j}$ is the nonlinear wave–wave coupling coefficient. The first step in determining these coefficients and hence the interaction physics, is to express (4) in terms of a transfer function equation that models these

physics. The coefficients of such an equation are then determined by measuring the changes in the complex amplitudes, $A(f)$, of the spectral component over a distance Δx. Equation (4) can then be transformed to an equation that relates the complex amplitude of the wave of frequency f_m at the downstream location, $x+\Delta x$, linearly to the complex amplitude of the same wave and quadratically to the complex amplitudes of all the interacting waves of frequencies f_i and f_j, that add or subtract to f_m, at the upstream location, x, see figure 2. Such a relationship is established by expressing (4) as

$$Y(f_m) = L(f_m)X(f_m) + \sum_{f_i \pm f_j = f_m} \sum Q^{f_m}_{f_i f_j} X(f_i)X(f_j), \quad f_m = f_i \pm f_j. \tag{5}$$

In this equation, $Y(f_m)$ and $X(f_m)$ are the complex Fourier amplitudes at frequency f_m of the fluctuations at two points in space in the transitioning mixing layer, see figure 2. The quantities $L(f_m)$ and $Q^{f_m}_{f_i f_j}$ are the linear and quadratic elements of the transfer function model, respectively, and are generally complex quantities. It is important to note here some of the characteristics of these transfer functions. First, the quadratic transfer function, unlike the linear transfer function, is a dimensional quantity. Second, the complex amplitude at the output frequency component is dependent on the product of the transfer functions and the complex amplitudes at the input frequencies. Third, the parametric reinforcement of f_m will manifest itself as a part of the dynamics represented by the linear portion of the transfer function since, as Nayfeh (1987) notes, the parametric approach to the fundamental-subharmonic interaction leads to linear equations with periodic or quasi-periodic coefficients. The subharmonic problem can thus be reduced to a simple transfer function system that is characterized by a parallel combination of linear and quadratically nonlinear transfer functions (see figure 2). Methods to estimate the linear and quadratic transfer functions are given by Kim & Powers (1988) and Ritz & Powers (1986).

4.2. *Wavenumber domain matching*

Instability-wave theory of parallel shear flows models the velocity fluctuations as travelling waves whose amplitudes may vary in time and space. The growth of these components can be affected by the linear instability mechanisms and by parametric and nonlinear interactions of the different components. Cross-bispectral analysis and transfer function modelling can be used to study the characteristics of the energy transfer due to these interactions between the different modes in the frequency domain. The spatial characteristics and wavenumber frequency matching conditions between interacting waves play a particularly important role in this process. Bretherton (1964), for example, studied wave triad resonance and amplitude growth in a homogeneous medium. For a one-dimensional nonlinear equation, he showed that wavenumber resonance as well as frequency resonance is important for two modes to interact quadratically to drive a third mode. In the case of mixing layers, Monkewitz (1988) set the condition that the fundamental and subharmonic modes should travel at the same phase speed for an efficient transfer of energy to occur. Basically, this requires wavenumber as well as frequency resonance between the two modes. This condition was also assumed in the parametric resonance model of Kelly (1967).

Different methods of extracting spatial characteristics from time series data have been applied by many investigators. Stegen & Van Atta (1970) used the signals from two points separated in space to measure phase speeds of the Fourier components in

grid turbulence. Beall, Kim & Powers (1982) devised a technique to compute the local energy spectrum as a function of frequency and wavenumber, from simultaneous measurements of the time series at two points in space. From such a spectrum, one can then compute the dispersion relationship and the power spectrum in the wavenumber domain. In the case of transitioning flows, the measurement of the phase difference between fluctuations at two points may be random, which results in turbulent broadening of the dispersion relation. Thus, the wavenumber value as a function of time can be treated as a random variable with a probability distribution about a mean for each frequency mode. Wavenumber measurements in the present experiments were made using a two-tensor probe as described by Jones *et al.* (1988) and the technique devised by Beall *et al.* (1982). Notice that the separation between the two sensors should be sufficiently small compared to a wavelength and coherence length to avoid spatial aliasing and phase incoherence, respectively. For a defined wavenumber, we have

$$\left.\begin{aligned} X(f) &= A(f)_x \exp\{i(k(f)_x + \phi(f)\}, \\ Y(f) &= A(f)_{x+\Delta x} \exp\{i(k(f)_{x+\Delta x} + \phi(f))\}. \end{aligned}\right\} \tag{6}$$

Since $X(f)$ and $Y(f)$ are measured simultaneously, the phase $\phi(f)$ at both points is the same. The sample cross-power spectrum is then

$$X^*(f)\,Y(f) = A^*(f)_x A(f)_{x+\Delta x} \exp\{i(k(f)\,\Delta x)\}. \tag{7}$$

The sample local wavenumber can then be related to the local phase of the sample cross-power spectrum and is given by

$$k(f) = \text{phase}\{X^*(f)\,Y(f)\}/\Delta x. \tag{8}$$

The computational technique for the wavenumber frequency spectrum, $S(k, f)$, the wavenumber spectrum $S(k)$ and the dispersion relationship are given by Beall *et al.* (1982). In this investigation, this technique is used to study the spatial characteristics of the transition to turbulence in plane mixing layers and in particular the parametric resonance mechanism that results in the growth of the subharmonic.

5. Results and discussion

The results presented in this paper are based on measurements of the streamwise velocity fluctuations at cross-stream locations corresponding to the maximum of the u'_{rms} fluctuations. Miksad (1972) presented a detailed evolution of the u'_{rms} profiles for the various instability modes involved in the transition process. Estimates of the growth rates of the subharmonic and fundamental modes by Miksad (1972) based on maximum u'_{rms} differ only slightly (10% or less) from those determined by cross-stream integration of total u'_{rms}. Measurements made at the maximum u'_{rms} avoid problems of riding up or down a modal profile slope, a problem encountered when measurements are made at a constant cross-stream distance. Jones (1983) showed that the cross-stream location of maximum u'_{rms} also closely coincides with the cross-stream location of maximum u^3_{rms}, which, according to bispectral analysis marks the location where maximum nonlinear or parametric coupling occurs between interacting modes.

The variations of the power spectrum of the streamwise fluctuations in the downstream direction, along maximum u'_{rms}, are shown in figure 6. In Region I, at $Rx/\lambda_0 = 1.0$, the spectrum exhibits peaks at the fundamental frequency, $f_0 =$ 215 Hz, and its subharmonic, $\frac{1}{2}f_0$. The large energy content at the subharmonic mode

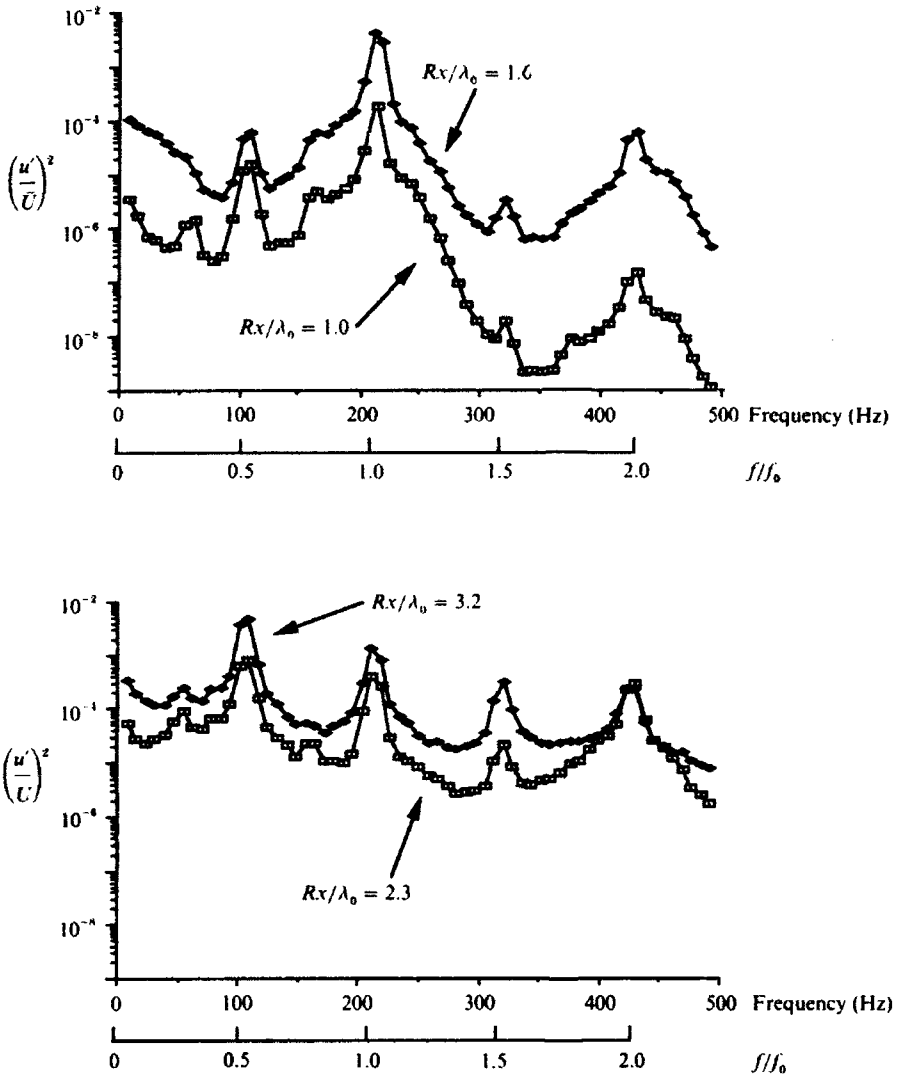


FIGURE 6. Variations of the power spectrum of the streamwise fluctuations in the downstream direction.

is due to the feedback of energy at this frequency from the vortex pairing mechanism. By $Rx/\lambda_0 = 1.6$ (i.e. Region II), both the fundamental and subharmonic modes have gained energy. Also, the band of fluctuations centred at $2f_0$ have gained a considerable amount of energy. The low-frequency components have also gained some energy. By $Rx/\lambda_0 = 2.3$ (i.e. Region III), the fundamental has lost some energy. However, the subharmonic continues to grow and the valleys between these modes and the other harmonics are filling rapidly.

The downstream development of the longitudinal fluctuations at the fundamental and subharmonic frequencies are shown in figure 4. Ho & Huang (1982) showed that the saturation of the fundamental corresponds to the vortex roll-up location, and the saturation of the subharmonic corresponds to the vortex merging location. In their experiment, the vortex roll-up occurred at $Rx/\lambda_0 = 2$ and the vortex merging at $Rx/\lambda_0 = 4$. In the present experiments, see figure 4, the fundamental and

subharmonic saturations occur at $Rx/\lambda_0 = 1.6$ and 3.2, respectively. The basic states of growth, equilibration and decay of the energy of the longitudinal fluctuations at the subharmonic and fundamental modes are observed and agree with the earlier results of Miksad (1972) and Ho & Huang (1982).

5.1. *Detection of coupled modes*

As the instability modes interact, they become coupled. The characteristics of energy transfers between coupled mode components cannot be determined by linear analysis. For example, the coherency function for a linear system, defined as

$$\gamma^2(f) = E[|X(f)\,Y(f)|^2]/\{E[|X(f)|^2]\,E[|Y(f)|^2]\},$$

where $E[\ldots]$ denotes an expected value, can only describe the linear relationship between the input $X(f)$ and the output $Y(f)$ at that same frequency. It cannot describe the interaction of two different frequency components, $X(f_i)$ and $X(f_j)$ at the input, to reinforce a third frequency component $Y(f_m)$ at the output, such that $f_m = f_i \pm f_j$. As noted by Powers & Miksad (1987), three-wave interactions can only be described by higher-order statistical moments, such as the cross-bispectrum.

The cross-bispectrum, defined as

$$B(f_i, f_j) = E[Y(f_m)\,X^*(f_i)\,X^*(f_j)],$$

where $f_m = f_i \pm f_j$ is a measure of the statistical dependence between the output spectral component at frequency f_m, $Y(f_m)$, and the input spectral components at frequencies f_i and f_j, $X(f_i)$ and $X(f_j)$, where $f_m = f_i \pm f_j$. The cross-bispectrum can only be non-zero if the frequency modes at f_i, f_j, and f_m are statistically dependent. In the case of three-wave interaction phenomena, this statistical dependence arises because of the phase coherence between the interacting waves. Thus, the product of the complex amplitudes of the three components will have a non-zero average over many realizations. The information that can be provided by the cross-bispectrum for the problem of subharmonic growth is of particular interest because it provides a means for detecting the presence of coupling between the fundamental and its subharmonic. A quantitative measure of this coupling is given by the cross-bicoherence squared

$$b^2(f_i, f_j) = |B(f_i, f_j)|^2/\{E[|X(f_i)\,X(f_j)|^2]\,E[|Y(f_m)|^2]\}.$$

It is important to note that the estimation of cross-bicoherence depends significantly on the number of realizations used to form an average. Figure 7, for example, shows that for one realization the cross-bicoherence will measure perfect coupling between the three modes under consideration. This is because in this case there is no averaging and the numerator and denominator are equal. As the number of realizations is increased, the level of bicoherence becomes lower. Finally, it converges to a constant value of 0.6 after 256 realizations. The results presented here show cross-bicoherence averaged over 256 realizations, each realization consisting of 128 samples.

Plots of the cross-bicoherence, along maximum $u'_{\rm rms}$, at different downstream locations are shown in figure 8. These locations were chosen to represent the different regions of growth of the fundamental and subharmonic modes. The values measured at $Rx/\lambda_0 = 0.32$ fall in the primary instability region or Region I. Values at $Rx/\lambda_0 = 1.6$ fall in Region II where the fundamental starts to saturate and the subharmonic has deviated from its exponential growth. Values shown for $Rx/\lambda_0 = 1.9$ and 2.3 fall in the secondary instability region or Region III. The results show that in

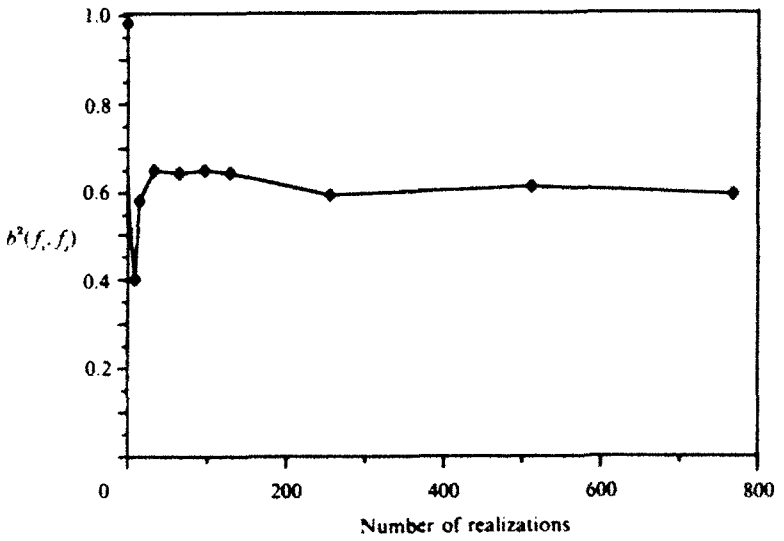


FIGURE 7. Convergence of the cross-bicoherence, $b^2(f_i, f_j)$, of a coherent frequency triplet as a function of the number of realizations.

region I, at $Rx/\lambda_0 = 0.32$, the strongest peak occurs at $(f_0, -\frac{1}{2}f_0)$. This shows that the fundamental mode and its subharmonic are phase coherent. This early phase coherence is the result of the feedback of energy from the downstream position where the subharmonic is emerging (vortex pairing is occurring), as suggested by Corke (1987). Although the phase coherency implies the possibility of nonlinear or parametric interaction and transfer of energy, we are still in the region of exponential growth where both modes are still growing at the linear rate predicted by the linear instability theory. The nonlinear effects are still minimal.

In Region II, at $Rx/\lambda_0 = 1.6$, the highest level of bicoherence appears at (f_0, f_0) and $(2f_0, -f_0)$ which indicates the production of the second harmonic via the self-interaction of the fundamental mode, as evidenced from the power spectra. Also, a new bicoherence pattern emerges where the contour level bands extend, away from (f_0, f_0) and $(2f_0, -f_0)$, to include fluctuations at $(f_0, f_0 - \delta f)$ and $(f_0 + \delta f, f_0)$ in the sum region and $(2f_0 \pm \delta f, -f_0)$ and $(2f_0 \pm \delta f, -f_0 \pm \delta f)$ in the difference region. These sideband bicoherencies indicate that the fundamental is coupling with fluctuations in its skirts to broaden the second harmonic. Note also the bicoherency between the fundamental and the subharmonic, $(f_0, \pm\frac{1}{2}f_0)$, and the fundamental and the $\frac{3}{2}$ harmonic, $(f_0, \pm\frac{3}{2}f_0)$. However, it is important to note that the level of bicoherence between the fundamental and its subharmonic, $(f_0, -\frac{1}{2}f_0)$, is lower in this region than in Region I. Note also that figure 4 shows that, in Region II, the subharmonic has deviated from exponential growth and started to equilibrate. Therefore, it appears that this equilibration of the subharmonic is accompanied by a reduction in the phase coherence between itself and the fundamental.

At the beginning of Region III, $Rx/\lambda_0 = 1.9$, where the subharmonic starts its second region of growth, the bicoherence level at $(f_0, -\frac{1}{2}f_0)$ is larger than that in Region II. This indicates that the growth of the subharmonic, in Region III, is accompanied by an increase in the level of its phase coupling with the fundamental. The above bispectral results demonstrate that the fundamental and subharmonic modes are phase-coupled over Regions I, II, and III of the transition. In Region I,

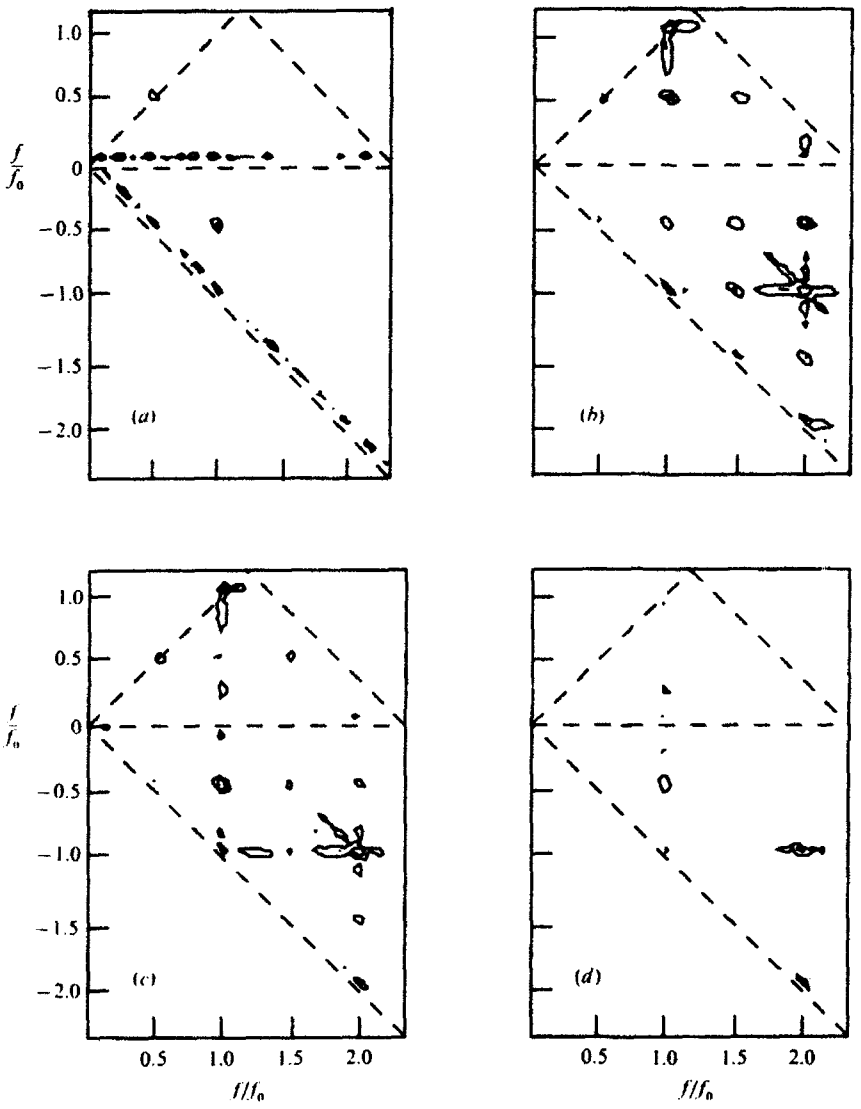


FIGURE 8. Cross-bicoherence spectra, $b^2(f_i, f_j)$, of the longitudinal fluctuations at different downstream locations that correspond to the different regions shown in figure 4. Contour levels are set at 0.3, 0.6, and 0.9. (*a*) Region I, $Rx/\lambda_0 = 0.32$; (*b*) Region II, $Rx/\lambda_0 = 1.6$; (*c*) Region III, $Rx/\lambda_0 = 1.9$; (*d*) Region IV, $Rx/\lambda_0 = 2.3$.

the coupling is a result of the feedback of energy at the fundamental and subharmonic frequencies from the vortex pairing (Corke 1987). In Region II, the level of bicoherence becomes lower as the fundamental starts to equilibrate and the growth rate of the subharmonic mode deviates from its initial exponential growth rate. As the subharmonic starts its second stage of growth in Region III, the level of bicoherence increases again. The existence of phase coupling between the fundamental mode, f_0, and its subharmonic, $\frac{1}{2}f_0$, provides a first step in establishing evidence of a nonlinear coupling or parametric resonance between the fundamental and its subharmonic.

5.2. *Linear and quadratic energy transfer analysis*

The cross-bicoherence analysis provides experimental evidence of the coupling between the fluctuations at the fundamental and subharmonic modes. However, cross-bicoherence measurements alone cannot provide information on the nature and efficiency of the coupling and its role in transferring energy to the subharmonic. It also cannot determine whether the measured three-wave couplings are due to interactions occurring at the location of measurement or are due to earlier upstream interactions which are advected past the probe by the flow. These questions can, to some extent, be answered by measuring the linear and quadratic transfer functions as shown in (5), see also figure 2. If we denote the output of the linear and quadratic transfer functions at frequency f_m by Y_L and Y_Q, respectively, the model output at frequency f_m is then given by

$$\left.\begin{aligned} &Y(f_m) = Y_L(f_m) + Y_Q(f_m), \\ \text{where}\quad &Y_L(f_m) = L(f_m)X(f_m), \quad Y_Q(f_m) = \sum_{f_i \pm f_j = f_m}\sum Q^{f_m}_{f_i f_j} X(f_i)X(f_j). \end{aligned}\right\} \tag{9}$$

Ritz & Powers (1986) used an iterative approach in the discrete frequency domain to estimate the linear and quadratic transfer functions. A more general approach by Kim & Powers (1988) that does not require iteration was used in our estimates of the linear and quadratic transfer functions. In this approach, the transfer functions are determined by solving two moment equations:

$$E[Y(f_m)X^*(f_m)] = L(f_m)E[X(f_m)X^*(f_m)] + \sum_{f_i \pm f_j = f_m}\sum Q^{f_m}_{f_i f_j} E[X(f_i)X(f_j)X^*(f_m)] \tag{10}$$

and

$$E[Y(f_m)X^*(f_k)X^*(f_l)] = L(f_m)E[X(f_m)X^*(f_k)X^*(f_l)] + \sum_{f_i \pm f_j = f_m}\sum Q^{f_m}_{f_i f_j} E[X(f_i)X(f_j)X^*(f_k)X^*(f_l)]. \tag{11}$$

where $f_m = f_i \pm f_j = f_k \pm f_l$. Note that these two equations permit one to express the linear transfer function, $L(f_m)$, and the quadratic transfer function, $Q^{f_m}_{f_i f_j}$ in terms of the various polyspectral moments that can be calculated from the Fourier Transforms of the input and output signals, respectively given by the upstream and downstream sensors of the probe. For example, the two terms on the left-hand sides of (10) and (11) are the cross-power spectrum $S_{xy}(f_m) = E[Y(f_m)X^*(f_m)]$ and the cross-bispectrum $S_{yxx}(f_k, f_l) = E[Y(f_m)X^*(f_k)X^*(f_l)]$. The terms on the right-hand sides are a hierarchy of spectral moments, specifically, the auto-power spectrum $S_{xx}(f_m) = E[X(f_m)X^*(f_m)]$, the auto-bispectrum $S_{xxx}(f_k, f_l) = E[X(f_m)X^*(f_k)X^*(f_l)]$ and a fourth-order spectral moment of the input $E[X(f_i)X(f_j)X^*(f_k)X^*(f_l)]$.

Once the linear and quadratic transfer functions are determined, one can predict the output spectrum, $S_{yy}(f_m)$, as

$$\begin{aligned} S_{yy}(f_m) &= E[Y(f_m)Y^*(f_m)] = E[(Y_L(f_m) + Y_Q(f_m))(Y_L(f_m) + Y_Q(f_m))^*] \\ &= E[Y_L(f_m)Y_L^*(f_m)] + E[Y_Q(f_m)Y_Q^*(f_m)] + 2\,\mathrm{Re}\,E[Y_L(f_m)Y_Q^*(f_m)] \\ &= S_L(f_m) + S_Q(f_m) + S_{LQ}(f_m). \end{aligned} \tag{12}$$

where $$S_L(f_m) = |L(f_m)|^2 E[|X(f_m)|^2] = |L(f_m)|^2 S_{xx}(f_m).$$

$$S_Q(f_m) = \sum_{f_i \pm f_j = f_m} \sum_{f_k \pm f_l = f_m} \{Q^{f_m}_{f_i f_j} Q^{f_m *}_{f_k f_l} E[X(f_i) X(f_j) X^*(f_k) X^*(f_l)]\}.$$

and $$S_{LQ}(f_m) = 2\,\mathrm{Re}\Big[\sum_{f_i \pm f_j = f_m} L(f_m) Q^{f_m *}_{f_i f_j} E[X(f_m) X^*(f_i) X^*(f_j)]\Big].$$

The 'predicted' output power spectrum is then the sum of the linear power spectrum, $S_L(f_m)$, the 'quadratic' power spectrum, $S_Q(f_m)$, and a linear–quadratic power spectrum, $S_{LQ}(f_m)$. It is important to note here that, in the set-up of the probe used in these experiments, the output power $S_{yy}(f_m)$ includes the input power $S_{xx}(f_m)$ which is advected by the mean flow from the first sensor to the second sensor. Because our interest is in the local characteristics of the linear and nonlinear growth mechanisms, we subtract the input power from the output power and consider only the change in the power spectrum, $\Delta S(f)$, between the two streamwise locations:

$$\begin{aligned}\Delta S(f_m) &= S_{yy}(f_m) - S_{xx}(f_m) = S_L(f_m) - S_{xx}(f_m) + S_Q(f_m) + S_{LQ}(f_m) \\ &= [|L(f_m)|^2 - 1] S_{xx}(f_m) + S_Q(f_m) + S_{LQ}(f_m) \\ &= S'_L(f_m) + S_Q(f_m) + S_{LQ}(f_m). \end{aligned} \tag{13}$$

where $S'_L(f_m) = [|L(f_m)|^2 - 1] S_{xx}(f_m)$. In this equation, the change in the energy content of the mode with frequency f_m is divided into three parts, namely the linear, quadratic, and linear–quadratic changes.

A power transfer rate, $\tau^2(f_m)$, can then be defined as the ratio of the estimated change in the energy of the mode with frequency f_m and the actual input energy of that mode over the distance Δx, $\Delta S(f_m)/(S_{xx}(f_m)\Delta x)$. The power transfer rate can then be divided into three parts, namely the linear power transfer rate, $\tau^2_l(f_m) = [|L(f_m)|^2 - 1]$, the quadratic power transfer rate, $\tau^2_Q(f_m)$, and the linear–quadratic power transfer rate, $\tau^2_{lq}(f_m)$. These are given as the ratios of $S'_L(f_m)$, $S_Q(f_m)$ and $S_{LQ}(f_m)$ over $S_{xx}(f_m)$, respectively. For a perfect system with no errors, all the energy change is accounted for by the three parts and the power transfer rate is equal to one. Also, as noted in §4, the linear transfer component, $\tau^2_l(f_m)$, includes the effect of the linear growth and the energy transfer via the parametric mechanism to the mode of frequency f_m. The transfer of energy via nonlinear resonance of two input modes f_i and f_j to an output mode at f_m is included in the quadratic power transfer component, $\tau^2_Q(f_m)$.

5.2.1. *Quadratic energy transfer*

As noted in §4, the nonlinear transfer of energy from two input modes f_i and f_j to an output mode at f_m depends on the quadratic transfer function $Q^{f_m}_{f_i f_j}$. This quantity is a complex function with amplitude and phase. From (5), it can be shown that the quadratic transfer function is given by

$$Q^{f_m}_{f_i f_j} = (Y(f_m) - L(f_m) X(f_m))/X(f_i) X(f_j), \quad f_m = f_i \pm f_j. \tag{14}$$

Note that the quadratic transfer function is inversely dependent on the amplitudes of input modes, $X(f_i)$ and $X(f_j)$. Figure 9 shows the logarithmic contour plots of the magnitude of the quadratic transfer functions, $|Q^{f_m}_{f_i f_j}|^2$, measured at locations that correspond to the four regions defined above. Note that the overall details of $|Q^{f_m}_{f_i f_j}|^2$ plots change gradually in the downstream direction. For instance, the magnitude of the quadratic transfer function is very low at $(f_0, -\frac{1}{2}f_0)$, at each downstream location. From (14), it is evident that this is in part due to the fact that the energy

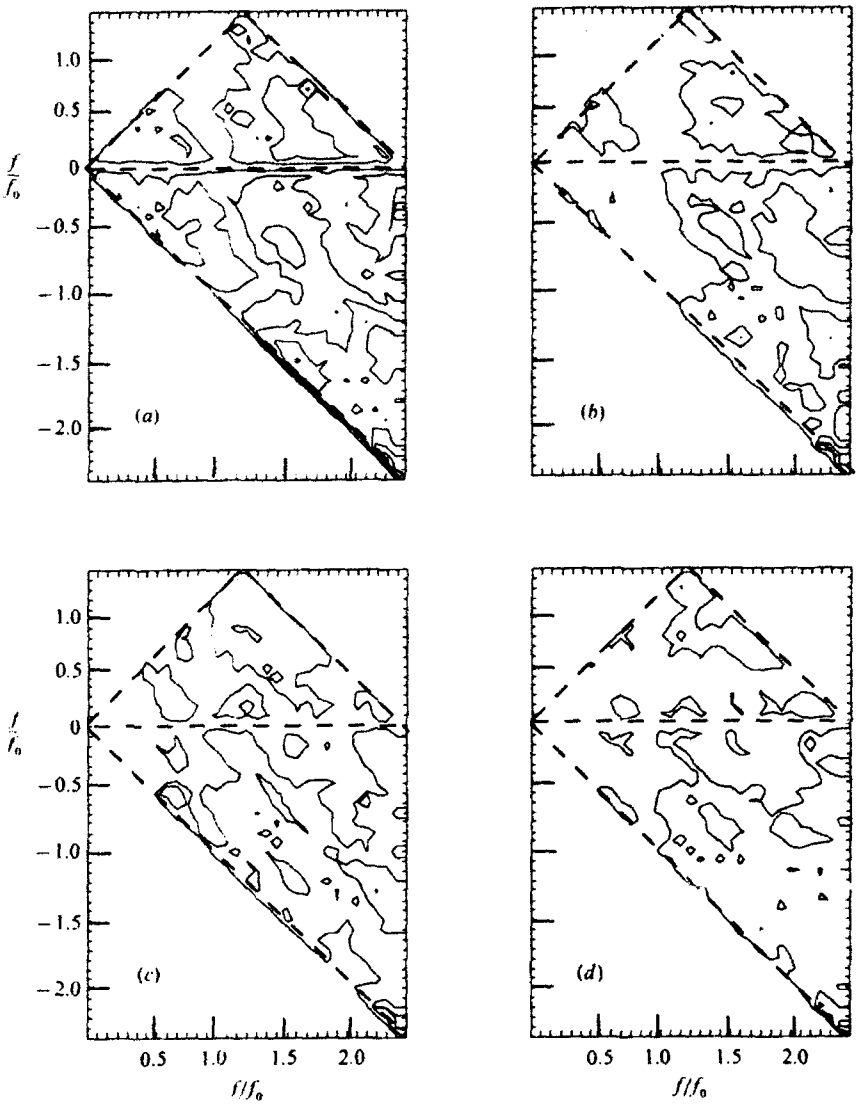


FIGURE 9. Logarithmic contour plots of the magnitude of the quadratic transfer functions, $|Q^{m}_{f_i f_j}|^2$, of the longitudinal fluctuations at different downstream locations that correspond to the different regions shown in figure 4. Contour levels are set at 0.5, 1.5, and 2.5. (*a*) Region I. $Rx/\lambda_0 = 0.32$; (*b*) Region II. $Rx/\lambda_0 = 1.6$; (*c*) Region III. $Rx/\lambda_0 = 2.6$; (*d*) Region IV. $Rx/\lambda_0 = 3.2$.

of the input modes at f_0 and $\frac{1}{2}f_0$ is large. However, it is important to note that, although the large energy levels at f_0, $\frac{1}{2}f_0$ and $\frac{3}{2}f_0$ are comparable, the magnitude of the quadratic transfer function is larger at $(\frac{3}{2}f_0, -f_0)$ and $(2f_0, -f_0)$ than at $(f_0, -\frac{1}{2}f_0)$. This suggests that the efficiency of the quadratic energy transfer via nonlinear resonance is larger for the $(\frac{3}{2}f_0, -f_0)$ and $(2f_0, -f_0)$ interactions than for the $(f_0, -\frac{1}{2}f_0)$ (i.e. fundamental—subharmonic) interaction. The sensitivity of the quadratic transfer function to the energy level of the input modes is also apparent from the fact that, at all downstream locations, measured values of the quadratic transfer function magnitude are typically large for pairs of frequency components that have low input energy. This effect stems from the fact that the quadratic transfer function is

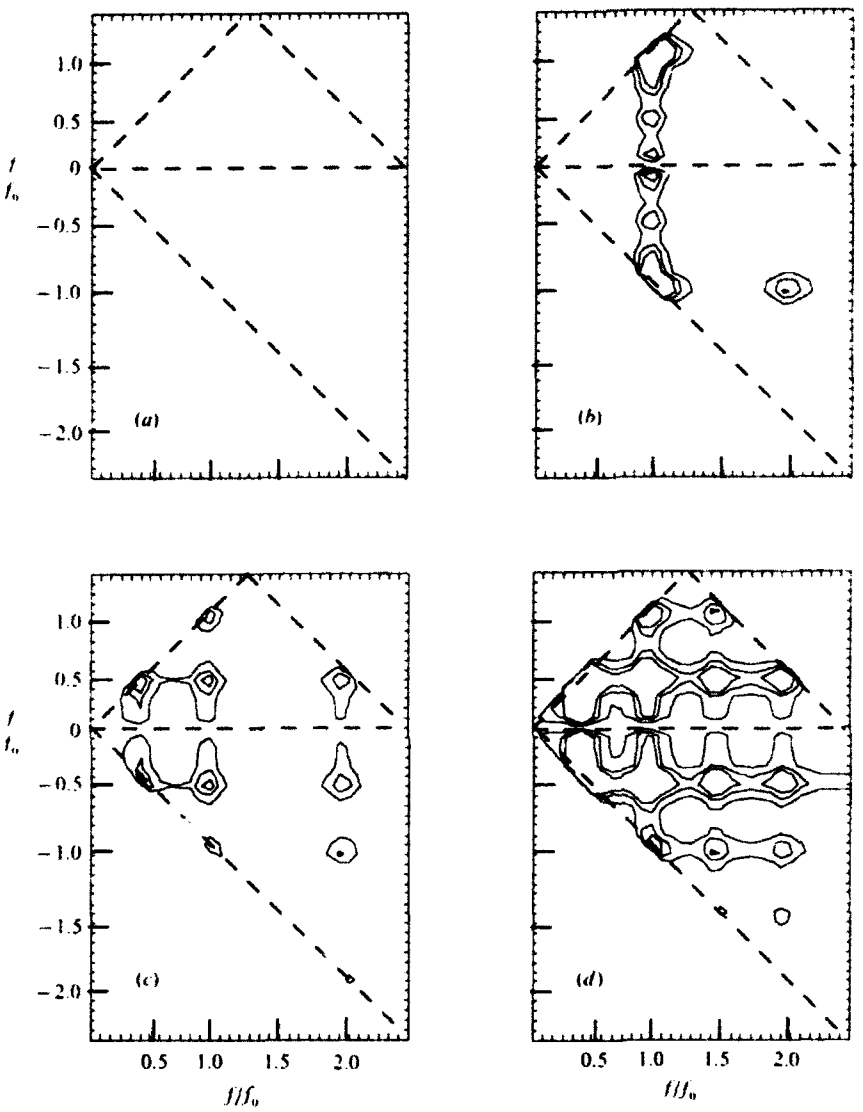


FIGURE 10. Contours plots of the interaction potential, $|X(f_i)X(f_j)|^2$, of the longitudinal fluctuations at different downstream locations that correspond to the different regions shown in figure 4. Contour levels are set at 0.2, 0.5, and 0.8. (*a*) Region I, $Rx/\lambda_0 = 0.32$; (*b*) Region II, $Rx/\lambda_0 = 1.6$; (*c*) Region III, $Rx/\lambda_0 = 2.6$; (*d*) Region IV, $Rx/\lambda_0 = 3.2$.

dimensional, as evident by (14), and is inversely dependent on the energy level of the interacting modes.

Miksad, Hajj & Powers (1989) proposed that a good measure of the amplitude dependence of $Q^{f_m}_{f_i f_j}$ is given by the magnitude of the coupled input energy level, $|X(f_i)X(f_j)|^2$, which will be referred to as the 'interaction potential'. This quantity is a measure of the potential amount of energy that interacting modes have for nonlinear energy transfer. Low-energy modes have less energy to transfer to a third mode than high-energy modes do. Figure 10 shows the frequency domain distribution of the interaction potential at four downstream locations that correspond to the different regions of the transition, as defined in §3. Note that the upper and lower

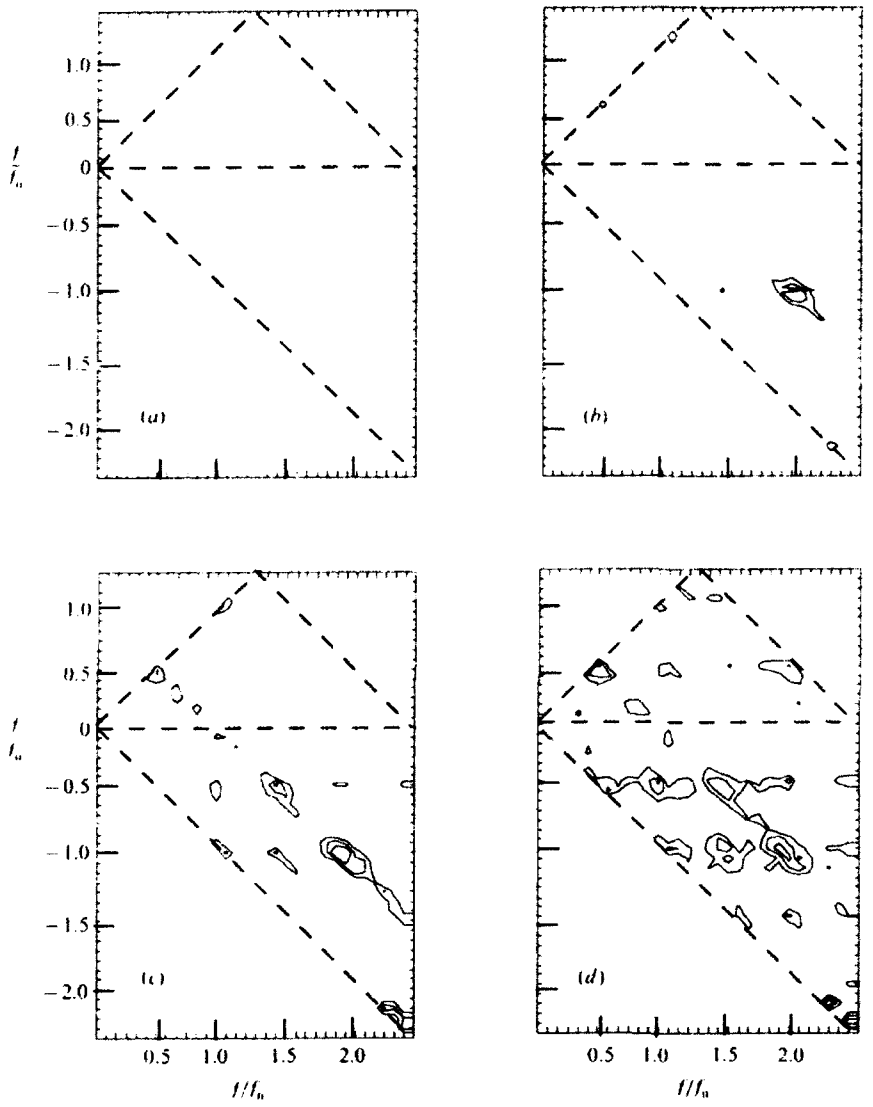


FIGURE 11. Contours of the weighted non-dimensional quadratic transfer functions, $Q^2(f_i, f_j)$, of the longitudinal fluctuations along $\max u'_{\mathrm{rms}}$ under natural excitation at different downstream locations that correspond to the different regions shown in figure 4. Contour levels are set at 0.3, 0.6, and 0.9. (*a*) Region I, $Rx/\lambda_0 = 0.32$; (*b*) Region II, $Rx/\lambda_0 = 1.6$; (*c*) Region III, $Rx/\lambda_0 = 2.6$; (*d*) Region IV, $Rx/\lambda_0 = 3.2$.

triangular regions are symmetrical, since $X(-f) = X^*(f)$ and therefore $|X(-f)|^2 = |X(f)|^2$. Note that in Region II, most of the interaction potential is centred at (f_0, f_0) indicating a significant input potential for the transfer of energy to the first harmonic. The actual energy transfer the first harmonic can be clearly seen in the power spectra and bicoherence plots. The large interaction potential at $(f_0, -\frac{1}{2}f_0)$ indicates that significant energy transfer to the subharmonic can also occur in these regions.

A measure of nonlinear energy transfer can be obtained by weighting the dimensional quadratic transfer function, $Q^2_{f_i f_j}$, by the energy potential of the

interacting input modes, $|X(f_i)X(f_j)|^2$, to obtain a non-dimensional quadratic transfer function such as

$$Q'^2(f_i, f_j) = E[|Q^{f_m}_{f_i f_j}|^2 (|X(f_i)X(f_j)|^2)^{\frac{1}{2}}], \quad f_m = f_i \pm f_j. \tag{15}$$

Contour plots of the weighted non-dimensional transfer functions, along maximum u'_{rms}, are shown in figure 11. Pairs that show high $Q'^2(f_i, f_j)$ characterize those interactions which are potentially most efficient in transferring energy. Note that, at $Rx/\lambda_0 = 0.32$, there are no indications of any prospect for energy transfer between any components. This location corresponds to Region I, the region of primary instability. These results agree with the predictions of the linear stability theory that all modes grow independently in the early region of the transition. More important to the question of subharmonic generation is the fact that although the cross-bicoherency plots indicate nonlinear coupling starting at $Rx/\lambda_0 = 0.32$, the $|Q'|^2$ measurements indicate that the prospect for quadratic energy transfer to the subharmonic in this region is not significant. This demonstrates that the growth of the subharmonic in Region I is not due to a nonlinear interaction mechanism. The first involvement of the subharmonic in nonlinear interaction appears in the sum region at $Rx/\lambda_0 = 1.6$, at the end of Region II. Recall that by this location, the fundamental has started to equilibrate and the subharmonic growth rate has deviated from its initial exponential growth. However, the involvement of the subharmonic component in nonlinear energy transfer does not become significant until $Rx/\lambda_0 = 2.6$, where the contour levels of $|Q'|^2$ at $(f_0, -\frac{1}{2}f_0)$ and $(\frac{3}{2}f_0, -f_0)$ are larger than at earlier locations. This indicates that, up to this location, the prospect for quadratic energy transfer to the subharmonic is very low. Further downstream, at $Rx/\lambda_0 = 3.2$ in Region III, the subharmonic, the fundamental and their harmonics have entered into a wide variety of interactions with other fluctuations and the $|Q'|^2$ contours indicate the initiation of a redistribution of subharmonic and fundamental mode energy to the different components of the flow.

5.2.2. *Linear and parametric energy transfer*

The above measurements of the quadratic transfer function and its weighted non-dimensional function show an important result. Both the quadratic transfer function, $|Q|^2$, and its weighted form, $|Q'|^2$ show large magnitudes at $(2f_0, -f_0)$. This indicates efficient transfer of energy between these modes. On the other hand, the level of the magnitudes of these functions at the fundamental subharmonic, $(f_0, -\frac{1}{2}f_0)$, indicate that the prospects for quadratic transfer of energy due to nonlinear resonance from the fundamental to the subharmonic are low. The question that arises is then: What is the role of the linear transfer function that includes the parametric effects on the growth of the subharmonic? This question can be answered by examining the linear transfer function and its role in the transfer of energy to the subharmonic and by comparing it to the corresponding role of the quadratic transfer function.

The growth rate of the fundamental, subharmonic and $\frac{3}{2}$ harmonic due to linear and linear-like effects, at different downstream locations, are given by the plot of the linear power transfer rate, $\tau_l^2(f_m)$, as shown in figure 12. Note that, in Region I, the linear power transfer rates of the three modes are constant and equal to the growth rates as predicted by the linear instability theory of Monkewitz & Huerre (1982). In Region II, between $Rx/\lambda_0 = 1.0$ and $Rx/\lambda_0 = 1.6$, the level of the linear power transfer rate, $\tau_l^2(f_m)$, of these modes starts to decrease. This decrease in the growth of these modes is expected because of the spreading of the mixing layer. Beyond

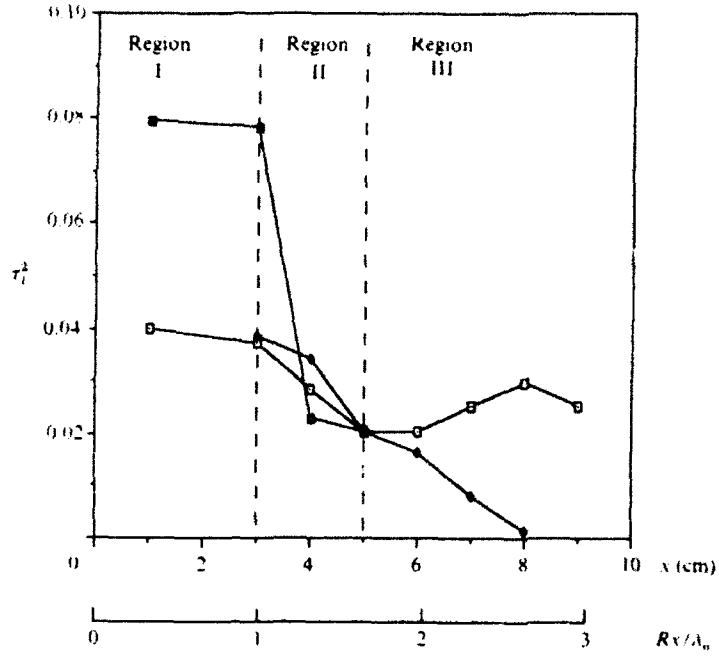


FIGURE 12. Linear power transfer rate, $\tau_l^2(f)$, of the fundamental, f_0 (■), the subharmonic, $\frac{1}{2}f_0$(⊡) and the $\frac{3}{2}$ harmonic $\frac{1}{2}f$ (◆) *vs.* downstream distance.

Region II, only the linear power transfer of the subharmonic and $\frac{3}{2}$ harmonic are shown. The fundamental growth is negative and therefore is not shown. In Region III, between $Rx/\lambda_0 = 1.6$ and $Rx/\lambda_0 = 2.6$, the linear power transfer rate, $\tau_l^2(f_m)$, of the $\frac{3}{2}$ harmonic continues to decrease. In contrast, the linear power transfer rate of the subharmonic increases. Only a linear-like mechanism, such as the parametric mechanism, may explain the increase in the linear power transfer rate of the subharmonic. These results indicate that the growth of the subharmonic in Region III is accompanied by an increase in the linear transfer of power as a result of the parametric resonance mechanism.

The total energy change of the subharmonic, $\Delta S(\frac{1}{2}f_0)$ at different downstream locations is shown in figure 13(*a*). Also shown on the same plot are the portions of $\Delta S(\frac{1}{2}f_0)$ that are due to linear and quadratic effects, $S_L(\frac{1}{2}f_0)$ and $S_Q(\frac{1}{2}f_0)$, respectively. Since the parametric resonance mechanism is a linear-like mechanism, its activity will appear in the linear term of (5). The transfer of energy to the subharmonic due to nonlinear interactions, $S_Q(\frac{1}{2}f_0)$, is two to three orders of magnitude less than the energy transfer by parametric linear-like mechanisms, $S_L(\frac{1}{2}f_0)$. Although both parametric and nonlinear mechanisms are simultaneously influencing the subharmonic, the results clearly show that the change in the power of the subharmonic mode is mainly due to a parametric resonance and only small nonlinear effects are present.

The nature of the energy transfer to the $\frac{3}{2}$ harmonic in Region III stands in contrast to that of the subharmonic. Figure 13(*b*) shows the total output energy change of the $\frac{3}{2}$ harmonic, $\Delta S(\frac{3}{2}f_0)$ at different downstream locations. The contributions of $S_L(\frac{3}{2}f_0)$ and $S_Q(\frac{3}{2}f_0)$ to the overall energy change, $\Delta S(\frac{3}{2}f_0)$, are also shown on the same figure. Note that in Region I the linear energy transfer to $\frac{3}{2}f_0$, $S_L(\frac{3}{2}f_0)$, is dominant. However, as we move further downstream to Region III,

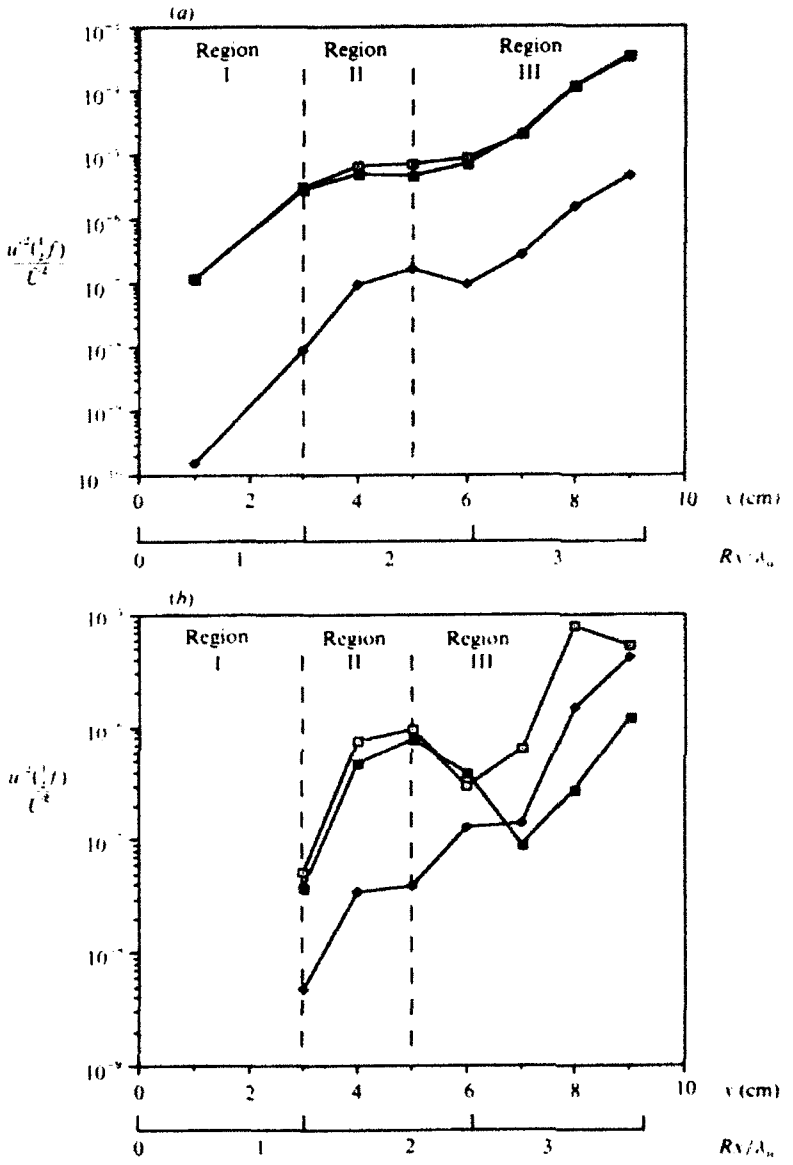


FIGURE 13(*a, b*). For caption see facing page.

quadratic energy transfer to $\frac{3}{2}f_0$, $S_Q(\frac{3}{2}f_0)$, becomes the dominant factor. This suggests that nonlinear interactions are more important in Region III in transferring energy to $\frac{3}{2}f_0$ than are parametric or linear mechanisms.

The differences between the mechanisms responsible for the energy transfer to the subharmonic and to the $\frac{3}{2}$ harmonic are also evident in the comparisons of the ratio of the quadratic to linear contributions, S_Q/S_L, for $\frac{1}{2}f_0$ and $\frac{3}{2}f_0$ at the different downstream locations, as shown in figure 13(*c*). Notice that while this ratio is of the order of 0.001 to 0.01 in the case of the subharmonic, $\frac{1}{2}f_0$, it is of the order of 0.1 to 10 in the case of $\frac{3}{2}f_0$, depending on the downstream location. This again suggests that

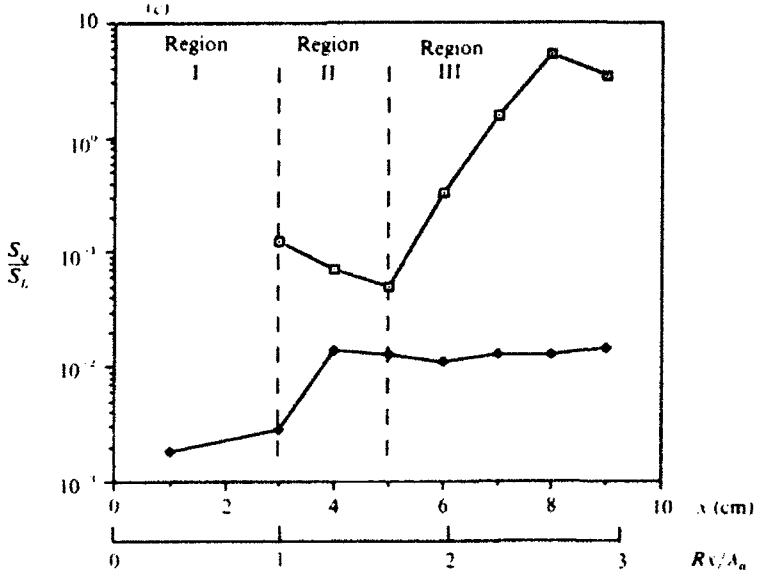


FIGURE 13. (*a*) Total power change, $\Delta N(\frac{1}{2}f_0)$, and linear and quadratic power transfer, $S_L(\frac{1}{2}f_0)$ and $S_Q(\frac{1}{2}f_0)$, respectively, *vs.* downstream distance; and Total power change, $\Delta N(\frac{3}{2}f_0)$, and linear and quadratic power transfer, $S_L(\frac{3}{2}f_0)$ and $S_Q(\frac{3}{2}f_0)$, respectively, *vs.* downstream distance (- ⊡ -, ΔN; -■-, S_L; -◆-, S_Q). (*c*) Ratio of quadratic to linear power transfer rate of the subharmonic, $\frac{1}{2}f_0$ (- ◆ -), and the $\frac{3}{2}$ harmonic, $\frac{3}{2}f_0$ (- ⊡ -) *vs.* downstream distance.

in Region III, nonlinear interactions are primarily responsible for the growth of $\frac{3}{2}f_0$ and that parametric effects are primarily responsible for subharmonic growth.

5.3. *Wavenumber–frequency matching*

The spatial characteristics of the transitioning mixing layer are obtained using the technique described in §4. These characteristics are seen in the estimates of the local wavenumber–frequency spectra $S(k, f)$ that are shown in figure 14. From these plots we can detect some features of the dispersion relationship. At $Rx/\lambda_0 = 0.32$, only the fundamental and subharmonic modes have a definite dispersion relationship. By $Rx/\lambda_0 = 1.6$, we notice the generation of the second harmonic. Also notice that the harmonic bands, at f_0, $\frac{1}{2}f_0$ and $2f_0$ are confined to a relatively narrow range of wavenumbers. By $Rx/\lambda_0 = 2.3$, we notice the broadening of the spectrum along the harmonic components and the valleys.

In order to gain better insight into the spatial characteristics of the different components of the flow, plots of the local wavenumber spectra are shown in figure 15. In the primary instability region or Region I, at $Rx/\lambda_0 = 1.0$, most of the energy is contained in a band of wavenumbers centred at $k_0 = 3.4$ rad/cm, which corresponds to a wavelength of 1.83 cm. This value is very close to the wavelength of the fundamental instability, $\lambda_0 = 1.98$ cm. By $Rx/\lambda_0 = 2.3$, or in Region II, different wavenumber components have gained some energy, leading to broadening of the spectrum. In Region III, between $Rx/\lambda_0 = 2.6$ and 3.2, the wavenumber spectra show that most of the energy is contained in a band of wavenumbers centred at $k = 1.7$ rad/cm ($= \frac{1}{2}k_0$), which is half the value at which the initial peak is located. Therefore, as the maximum energy in the frequency spectrum, in figure 6, shifts from f_0 to $\frac{1}{2}f_0$, the peak in the wavenumber spectra, in figure 15, shifts from k_0 to $\frac{1}{2}k_0$. This

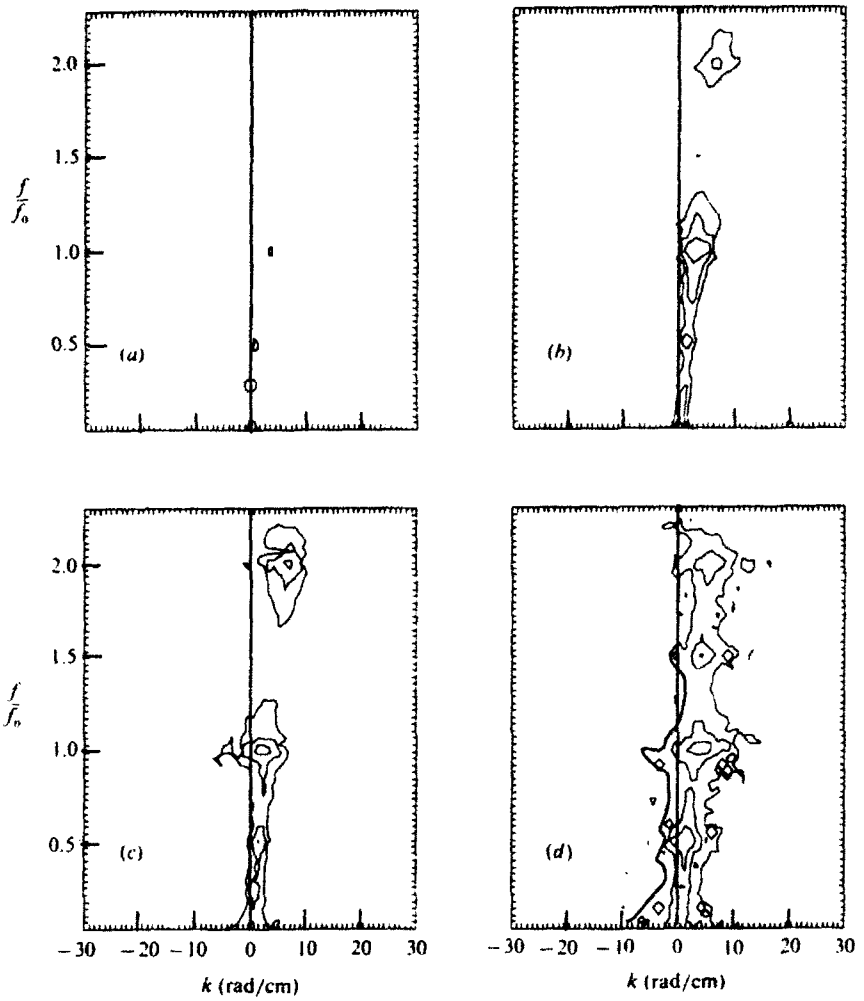


FIGURE 14. Wavenumber–frequency spectra, $S(k, f)$, of the longitudinal fluctuations at different downstream locations that correspond to the different regions shown in figure 4. Contour levels are set at 10^{-6}, 10^{-5}, and 10^{-4}. (*a*) $Rx/\lambda_0 = 0.32$; (*b*) 1.6; (*c*) 2.3; (*d*) 3.2.

shift occurs after the primary instability region of the transition, and coincides with the second region of subharmonic growth. The merging of the two vortices reduces the frequency of their passage from f_0 to $\frac{1}{2}f_0$ and increases their characteristic size from $l_0 = 2\pi/k_0$ to $2l_0$. The shifting of the centre of the band from k_0 to $\frac{1}{2}k_0$ represents an increase in the characteristic size of the vortices and therefore provides a spatial indication of the vortex merging process.

An important condition for parametric resonance to take place is that the subharmonic and the fundamental modes travel at the same phase speed. This condition was assumed by Kelly (1967) and postulated by Monkewitz (1988). Figure 16 shows the variation of the normalized phase speeds, $(\omega/k)/\bar{U}$, of the fundamental and the subharmonic in the downstream direction. Notice the good matching between the two modes beyond $Rx/\lambda_0 = 1.6$ (i.e. Region III). This shows that the condition of wavenumber resonance is satisfied beyond the initial region of growth

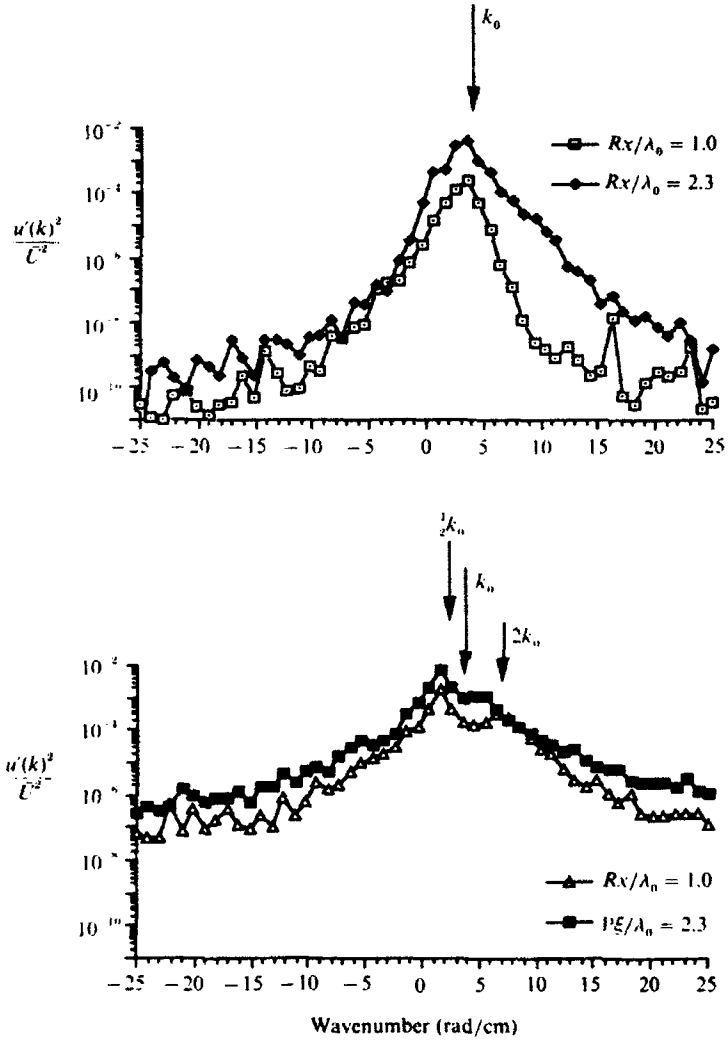


FIGURE 15. Wavenumber spectra of the longitudinal fluctuations at downstream locations that correspond to the different regions shown in figure 4.

of the subharmonic. This is in agreement with the requirements of the analysis of Kelly (1967) and Monkewitz (1988).

Monkewitz (1988) showed that the amplitude of the fundamental mode must attain a critical value before it becomes phase locked with the subharmonic and the growth rate of the subharmonic is modified. At this location the phase speed of the subharmonic is close to unity, i.e. the phase speed of the neutral fundamental. Monkewitz (1988) also showed that when the fundamental amplitude is larger than or equal to the critical amplitude, the growth of the subharmonic and its phase speed are dependent on the phase difference, $\Delta\phi$, between the fundamental and the subharmonic (see figures 6 and 7 in Monkewitz 1988). These figures show that the optimal conditions for the subharmonic growth occur when the phase difference between the two modes is close to zero. Monkewitz also showed that when the fundamental amplitude is decreasing, the growth of the subharmonic is enhanced regardless of the initial phase difference (see figure 9 in Monkewitz 1988). The results

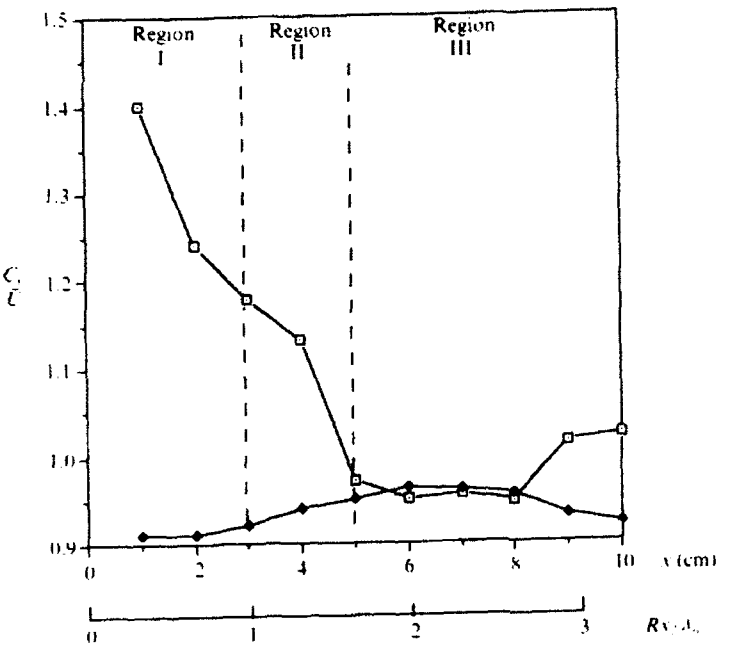


FIGURE 16. Downstream variations of the normalized phase speed, C_r/U, of the fundamental (◆) and subharmonic (⊡) modes.

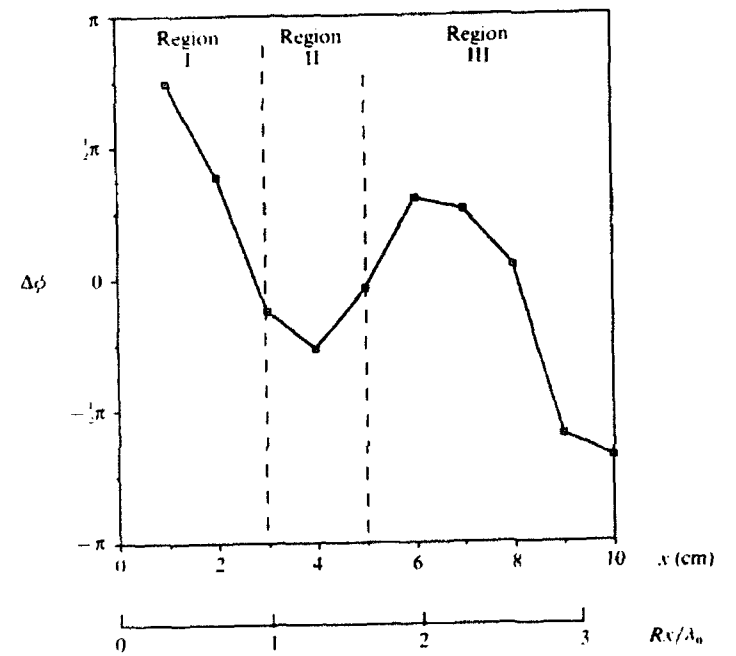


FIGURE 17. Downstream variations of the phase difference, $\Delta\phi$, between the fundamental and subharmonic modes.

of our experiments show that when the fundamental reaches its critical amplitude the growth rate of the subharmonic is decreasing, the phase speed of the subharmonic is not equal to unity and the energy transfer between the fundamental and the subharmonic is not efficient. However, these conditions for the subharmonic growth are met further downstream, beyond $Rx/\lambda_0 = 1.6$. Figure 17 shows the variation in the phase difference between the fundamental and the subharmonic in the downstream direction. At $Rx/\lambda_0 = 1.0$, where the fundamental amplitude is close to the critical amplitude, the phase difference between the fundamental and the subharmonic close to zero. By comparison to figure 6 of Monkewitz (1988), we note that at this phase difference, the growth of the subharmonic is suppressed over a small distance before it starts growing again. The reduced growth rate at $Rx/\lambda_0 = 1.0$ agrees with the predictions of Monkewitz (1988). At $Rx/\lambda_0 = 1.6$, where the fundamental amplitude is equal to five times the critical amplitude, the phase difference is still close to zero. By comparison with figure 7 of Monkewitz (1988), we note that the subharmonic growth is reduced at this phase angle. This also agrees with the reduced growth rate at this location as seen in figure 4. This reduced growth is also due to the low bicoherence (figure 7) which indicates low coupling and large phase difference variations. Further downstream, at $Rx/\lambda_0 = 2.3$, the amplitude of the fundamental is equal to twice that of the critical amplitude. The subharmonic amplitude is increasing significantly. However the fundamental amplitude is decreasing. By comparison with figure 9 of Monkewitz (1988), we note that the subharmonic should be growing at any phase difference. This result also agrees with the energy transfer measurements that show efficient energy transfer from the fundamental to the subharmonic at this location.

6. Conclusions

The nonlinear and parametric characteristics that are associated with the subharmonic generation in the transition to turbulence in plane mixing layers are detected and quantified. In the primary instability region of the transition, all instability modes, including the subharmonic, grow independently. No significant transfer of energy between any modes is detected. Measurements of the power spectra and the associated linear and quadratic transfer functions show that beyond this initial region, the spatial growth of the subharmonic supports the parametric approach which is described by a linear model with varying time coefficients. On the other hand, the same measurements show that the growth of $2f_0$ and $\frac{3}{2}f_0$ is best described by the quadratic nonlinear interaction approach. The growth rate of the subharmonic starts modifying about the location where $Rx/\lambda_0 \approx 1$. At this location, the fundamental amplitude is equal to $0.012\bar{U}$ and has not yet started to saturate. This shows that the parametric resonance between the fundamental and the subharmonic can start to affect subharmonic growth before the fundamental reaches its saturation amplitude. However, the most efficient transfer of energy to the subharmonic occurs further downstream as the fundamental mode reaches an amplitude close to its saturation amplitude. Measurements also indicate that the fundamental, besides interacting with the subharmonic, is also engaged in redistributing its energy via nonlinear interactions to other components such as the first harmonic and the valleys. Local wavenumber measurements verify that frequency–wavenumber resonance matching conditions exist between the fundamental and subharmonic in regions where efficient energy exchange to the subharmonic takes place. These results are in general agreement with theoretical models by Kelly (1967) and Monkewitz (1988).

This paper is based in part upon work supported by the Texas Advanced Technology Program under Grant No. ATP 3280, and in part by the National Science Foundation under Grant No. MSM-8211205. The digital signal processing techniques were developed under the auspices of the Department of Defense Joint Services Electronics Program under the Air Force Office of Scientific Research (AFOSR) Contract No. F49620-89-C-0044.

REFERENCES

BEALL, J. M., KIM, Y. C. & POWERS, E. J. 1982 Estimation of wavenumber and frequency spectra using fixed probe pairs. *J. Appl. Phys.* **53**, 3933–3940.

BRETHERTON, F. P. 1964 Resonant interactions between waves. The case of discrete oscillations. *J. Fluid Mech.* **20**, 457–479.

BROWAND, F. K. 1966 An experimental investigation of the instability of an incompressible separated shear layer. *J. Fluid Mech.* **26**, 281–307.

BROWN, G. L. & ROSHKO, A. 1974 On density effects and large structure in turbulent mixing layers. *J. Fluid Mech.* **64**, 775–816.

CORKE, T. C. 1987 Measurements of resonant phase locking in unstable axisymmetric jets and boundary layers. In *Nonlinear Wave Interactions in Fluids* (ed. R. W. Miksad, T. R. Akylas & T. Herbert). The Winter Annual Meeting of ASME, AMD vol. 87, pp. 37–65.

HO, C. M. 1982 Local and global dynamics of free shear layers. In *Proc. Symp. on Numerical and Physical Aspects of Aerodynamics Flow* (ed. T. Cebeci), pp. 521–533. Springer.

HO, C. M. & HUANG, L. S. 1982 Subharmonics and vortex merging in mixing layers. *J. Fluid Mech.* **119**, 443–473.

HO, C. M. & HUERRE, P. 1984 Perturbed free shear layers. *Ann. Rev. Fluid Mech.* **16**, 365–424.

HUANG, L. S. & HO, C. M. 1990 Small-scale transition in a plane mixing layer. *J. Fluid Mech.* **210**, 475–500.

ITOH, N. 1977 Nonlinear stability of parallel flows with subcritical Reynolds numbers. Part 1. An asymptotic theory valid for small amplitude disturbances. *J. Fluid Mech.* **82**, 455–467.

JONES, F. L. 1983 An experimental study of nonlinear wave interactions and modulations during transition of a symmetric wake. Ph.D. dissertation, LTT-83-02, The University of Texas at Austin.

JONES, F. L., RITZ, C. P., MIKSAD, R. W., POWERS, E. J. & SOLIS, S. R. 1988 Measurement of the local wavenumber and frequency spectrum in a plane wake. *Exps. in Fluids* **6**, 365–372.

KELLY, R. E. 1967 On the stability of an inviscid shear layer which is periodic in space and time. *J. Fluid Mech.* **27**, 657–689.

KIM, K. I. & POWERS, E. J. 1988 A digital method of modeling quadratically nonlinear systems with a general random input. *IEEE Trans. Acoust., Speech, Signal Process.* **36**, 1758–1769.

MIKSAD, R. W. 1972 Experiments on the non-linear stages of free shear layer transition. *J. Fluid Mech.* **56**, 695–719.

MIKSAD, R. W. 1973 Experiments on nonlinear interactions in the transition of a free shear layer. *J. Fluid Mech.* **59**, 1–21.

MIKSAD, R. W., HAJJ, M. R. & POWERS, E. J. 1989 Measurements of nonlinear transfer functions for subharmonic generation in mixing layers. *AIAA* 89-0980.

MIKSAD, R. W., JONES, F. L., POWERS, E. J., KIM, Y. C. & KHADRA, L. 1982 Experiments on the role of amplitude and phase modulations during the transition to turbulence. *J. Fluid Mech.* **123**, 1–29.

MOLLO-CHRISTENSEN, E. 1971 Physics of turbulence flows. *AIAA J.* **9**, 1217–1228.

MONKEWITZ, P. A. 1988 Subharmonic resonance, pairing and shredding in the mixing layer. *J. Fluid Mech.* **188**, 223–252.

MONKEWITZ, P. A. & HUERRE, P. 1982 The influence of the velocity ratio on the spatial instability of mixing layers. *Phys. Fluids* **25**, 1137–1143.

NAYFEH, A. H. 1987 On secondary instabilities in boundary layers. In *Stability of Time Dependent and Spatially Varying Flows* (ed. D. L. Dwoyer & M. Y. Hussaini), pp. 18–49. Springer.

NIKITOPOULOS, D. E. & LIU, J. T. C. 1987 Nonlinear binary-mode interactions in a developing mixing layer. *J. Fluid Mech.* **179**, 345–370.

PIERREHUMBERT, R. T. & WIDNALL, S. E. 1982 The two- and three-dimensional instabilities of a spatially periodic shear layer. *J. Fluid Mech.* **114**, 59–82.

POWERS, E. J. & MIKSAD, R. W. 1987 Polyspectral and measurements and analysis of nonlinear wave interactions. In *Nonlinear Wave Interactions in Fluids* (ed. R. W. Miksad, T. R. Akylas & T. Herbert). The Winter Annual Meeting of ASME, AMD-vol. 87, pp. 37–65.

RAETZ, G. S. 1959 A new theory of the cause of transition in fluid flows. *Norair Rep.* NOR-59-383, Hawthorne, CA.

RITZ, C. P. & POWERS, E. J. 1986 Estimation of nonlinear transfer functions for fully developed turbulence. *Physica* D **20**, 320–334.

RITZ, C. P., POWERS, E. J., MIKSAD, R. W. & SOLIS, R. S. 1988 Nonlinear spectral dynamics of a transitioning flow. *Phys. Fluids* **31**, 3577–3588.

SATO, H. 1959 Further investigation on the transition of two dimensional separated layers at subsonic speed. *J. Phys. Soc. Japan* **14**, 1797–1810.

STEGUN, G. R. & VAN ATTA, C. W. 1970 A technique for phase speed measurements in turbulent flow. *J. Fluid Mech.* **43**, 689–699.

STUART, J. T. 1960 On the nonlinear mechanisms of hydrodynamic stability. *J. Fluid Mech.* **4**, 1–21.

WINANT, C. D. & BROWAND, F. K. 1974 Vortex Pairing: the mechanism of turbulent mixing layer growth at moderate Reynolds number. *J. Fluid Mech.* **63**, 237–255.

ICASSP-92

VOLUME 2

S_2 ANU

SPEECH PROCESSING 2
AUDIO
NEURAL NETWORKS
UNDERWATER ACOUSTICS

1992
IEEE INTERNATIONAL CONFERENCE
ON
ACOUSTICS, SPEECH AND
SIGNAL PROCESSING

MARCH 23-26, 1992
THE SAN FRANCISCO MARRIOTT
SAN FRANCISCO, CALIFORNIA



IEEE

92CH3103-9

Sponsored by
THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS,
SIGNAL PROCESSING SOCIETY

TWO-TONE VS. RANDOM PROCESS INPUTS FOR NONLINEAR DISTORTION ESTIMATION

Y.S. Cho and E. J. Powers

Department of Electrical and Computer Engineering, and Electronics Research Center
The University of Texas at Austin, Austin, Texas 78712-1084

ABSTRACT

This paper compares two methods (two-tone input approach and random process input approach) for the estimation of harmonic and intermodulation distortions of nonlinear systems. The random input approach, where a nonlinear system is modeled by a second-order Volterra series, is examined in terms of its statistical properties, and its advantages and limitations over the classical two-tone input approach. Experimental results are shown where these two approaches are applied to evaluate second-order distortion of a loudspeaker and to compare the performance of these approaches in terms of Volterra kenels and distortion factors.

1. Introduction

Harmonic and intermodulation distortions, which are most significant in high-fidelity systems, usually result from nonlinear input-output characteristics of some component. The measurement of nonlinear distortion, which is the subject of this paper, has traditionally been conducted by applying sinusoidal signals and measuring the extent to which harmonic or combination tones are generated in the output of the component under test. In contrast to the sinusoidal input approach [1], a random input approach has been developed to analyze nonlinear distortion since a random signal well approximates a frequency-multiplexed input such as a speech signal [2],[3]. Now that two different approaches are available for nonlinear distortion estimation, the following questions arise: Are the distortion products estimated by the random input approach the same as the ones estimated by the sinusoidal input approach? If so, under what conditions? And what are the advantages and limitations of each approach? In this paper, we address these questions by (a) expressing these two approaches by a Volterra series representation [4], (b) investigating the statistical properties of the transfer function estimators in the random input approach, (c) comparing results (in terms of Volterra kernels and distortion factors) obtained by both approaches, and (d) discussing the advantages and limitations of both approaches. Finally, experimental results of applying the two-tone input approach and random input approach to quantify second-order distortion of a loudspeaker are presented.

2. Quadratic System Modeling with Random Input

Consider a frequency domain Volterra series representation of a quadratically nonlinear time-invariant system.

$$Y(f_m) = H_1(f_m)X(f_m) + \sum_{i+j=m} H_2(f_i, f_j)X(f_i)X(f_j) + \varepsilon(f_m) \quad (1)$$

where $X(f_m)$, $Y(f_m)$, and $\varepsilon(f_m)$ are the N-point discrete Fourier transforms of the input, output, and model error, respectively. The second-order factor of intermodulation distortion (D_{IM2}) at frequency $f_1 \pm f_2$ is defined as

$$D_{\mathrm{IM2}} = 20\log[\mathrm{IM2}] = 20\log\left[\left|\frac{V_0(f_1 \pm f_2)}{V_0(f_1)}\right|\right] \quad [\mathrm{dB}] \quad (2)$$

where $V_0(f_1 \pm f_2)$ is the output voltage at the frequency $f_1 \pm f_2$. This term may be expressed in terms of the frequency domain Volterra kernels in the form

$$D_{\mathrm{IM2}} = 20\log\left[\left|\frac{H_2(f_1, \pm f_2)}{H_1(f_1)}\right| V_{\mathrm{I}}\right] \quad [\mathrm{dB}] \quad (3)$$

where V_{I} is the amplitude of both input voltages ($v_i(t) = V_{\mathrm{I}}\cos 2\pi f_1 t + V_{\mathrm{I}}\cos 2\pi f_2 t$). Likewise, the second-order harmonic distortion factor D_{H2}, which is defined as

$$D_{\mathrm{H2}} = 20\log[\mathrm{H2}] = 20\log\left[\left|\frac{V_0(2f)}{V_0(f)}\right|\right] \quad [\mathrm{dB}] \quad (4)$$

0-7803-0532-9/92 $3.00 © 1992 IEEE

may be expressed in terms of the Volterra kernels as

$$D_{H2} = 20\log\left[\frac{1}{2}\left|\frac{H_2(f,f)}{H_1(f)}\right|V_1\right] \quad [\text{dB}] \tag{5}$$

Thus, the second-order distortion factors of a component under test can be obtained easily at various frequencies if the linear transfer function and quadratic transfer function of the component can be estimated simultaneously at those frequencies.

By minimizing the mean square error $E[|\varepsilon(f_m)|^2]$ one can obtain an optimum estimation of the transfer functions as follows:

$$E[X^*(m)Y(m)] = H_1(m)E[|X(m)|^2] + \sum_{i+j=m} H_2(i,j)E[X^*(m)X(i)X(j)] \tag{6}$$

$$E[X^*(k)X^*(l)Y(m)] = H_1(m)E[X^*(k)X^*(l)X(m)] + \sum_{i+j=m} H_2(i,j)E[X^*(k)X^*(l)X(i)X(j)] \tag{7}$$

When the system input is a zero-mean Gaussian random signal, the linear and quadratic transfer functions are obtained separately and determined by the various spectra up to third order. However, it was shown in [5] that commonly used random inputs, such as those provided by a noise generator, cannot automatically be regarded as sufficiently Gaussian for nonlinear identification studies. Thus, a practical digital method of estimating nonlinear distortion of a system without assuming any particular input statistics was presented [3],[6].

For non-Gaussian inputs, one has to solve (6) and (7) simultaneously with respect to $H_1(m)$ and $H_2(i,j)$. These equations can be rewritten using vector notation as follows:

$$E[\mathbf{X}^*(m)Y(m)] = E[\mathbf{X}^*(m)\mathbf{X}^T(m)]\mathbf{H}(m) \tag{8}$$

where $\mathbf{X}(m)$ and $\mathbf{H}(m)$ are of matrix form ($E[\mathbf{X}^*(m)\mathbf{X}^T(m)]$ consists of second-, third-, and fourth-order spectral moments, and $\mathbf{H}(m)$ consists of $H_1(m)$ and $H_2(i,j)$). The optimal solution in the mean square sense is given by [6]

$$\mathbf{H}(m) = E[\mathbf{X}^*(m)\mathbf{X}^T(m)]^{-1}E[\mathbf{X}^*(m)Y(m)] \tag{9}$$

Here, we assume we have a set of M realizations or records for $x(n)$ and $y(n)$ consisting of N samples each. If $X_k(m)$, $Y_k(m)$, and $\varepsilon_k(m)$ denote the DFT's of $x_k(n)$, $y_k(n)$, and $e_k(n)$, respectively, then the optimal estimator of the transfer function is given by

$$\hat{\mathbf{H}}(m) = [\sum_{k=1}^{M} \mathbf{X}_k^*(m)\mathbf{X}_k^T(m)]^{-1}[\sum_{k=1}^{M} \mathbf{X}_k^*(m)Y_k(m)] \tag{10}$$

where $\mathbf{X}_k(m)$ is of matrix form consisting of various higher-order spectral ments, up to fourth order, formed by the k-th realization. Provided that $\varepsilon_k(m)$ is independent of $X_k(m)$, it can be easily shown that

$$E[\hat{\mathbf{H}}(m)] = \mathbf{H}(m) \tag{11}$$

Provided that $\varepsilon_k(m)$ is independent of $\varepsilon_{k'}(m)$, where $k \neq k'$, with variance σ_ε^2, the covariance matrix of $\hat{\mathbf{H}}(m)$ is obtained as follows:

$$\text{cov}\ \hat{\mathbf{H}}(m) = \sigma_\varepsilon^2[\sum_{k=1}^{M} \mathbf{X}_k^*(m)\mathbf{X}_k^T(m)]^{-1} \tag{12}$$

The result (12) can be used to determine how the variance of the estimate decrease with the number of realizations. Furthermore, it can be shown that the estimator in (10) is unbiased and consistent if (a) there is no noise at the input point (b) noise at the output is uncorrelated (c) the system under consideration is appropriately modeled.

3. Bichromatic vs. Random Process Estimates

In order to calculate the response of the quadratically nonlinear system subject to two sinusoidal inputs with frequency f_1 and f_2 ($f_1 > f_2$), let the input be

$$x(t) = \sum_{i=1}^{2} |V_{I_i}|\cos[2\pi f_i t + \theta_i] \tag{13}$$

which can also be expressed as

$$x(t) = \frac{1}{2}\sum_{i=-2}^{2} V_{I_i}\exp[j2\pi f_i t] \tag{14}$$

where

$$V_{I_{-i}} = V_{I_i}^*, \quad V_{I_0} = 0, \quad f_{-i} = -f_i$$

The linear response to the inputs specified by (13) is given by

$$y_1(t) = \sum_{i=1}^{2} |V_{I_i}||H_1(f_i)|\cos[2\pi f_i t + \theta_i + \phi_1(f_i)] \tag{15}$$

The quadratic response to the input specified by (14) is given by

$$y_2(t) = \frac{1}{2}\sum_{i_1=-2}^{2}\sum_{i_2=-2}^{2} V_{I_{i_1}}V_{I_{i_2}}H_2(f_{i_1},f_{i_2})\exp[j2\pi(f_{i_1}+f_{i_2})t] \tag{16}$$

Since $H_2(-f_1,-f_2) = H_2^*(f_1,f_2)$ and $H_2(f_1,f_2) = H_2(f_2,f_1)$ in real systems, (16) can be expressed as

$$\begin{aligned} y_2(t) = {} & \sum_{i=1}^{2} |V_{I_i}|^2|H_2(f_i,f_i)|\cos[2\pi(2f_i)t + 2\theta_i + \phi_2(f_i,f_i)] \\ & + 2|V_{I_1}||V_{I_2}||H_2(f_1,f_2)|\cos[2\pi(f_1+f_2)t + \theta_1 + \theta_2 + \phi_2(f_1,f_2)] \\ & + 2|V_{I_1}||V_{I_2}||H_2(f_1,-f_2)|\cos[2\pi(f_1-f_2)t + \theta_1 - \theta_2 + \phi_2(f_1,-f_2)] \\ & + \sum_{i=1}^{2} |V_{I_i}|^2 H_2(f_i,-f_i) \end{aligned} \tag{17}$$

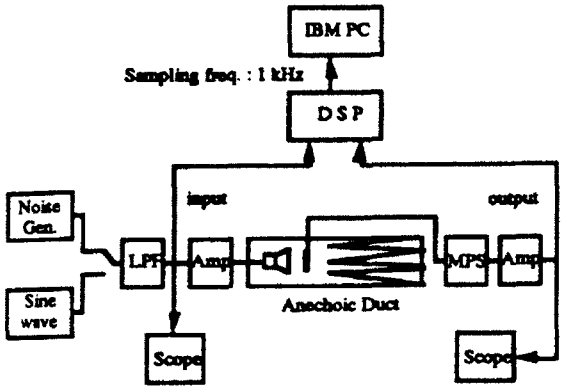


Figure 1: Block diagram of distortion analysis set-up

The first term in (17) denotes second-order harmonic distortions, the second term denotes an intermodulation distortion at the sum frequency, the third term denotes an intermodulation distortion at the difference frequency, and the last term denotes the DC component. Note that the output at the difference frequency, $f_1 - f_2$, contains not only the intermodulation distortion at that frequency but also the second-order harmonic distortion due to frequency f_2 if $f_1 - f_2 = 2f_2$. These harmonic and intermodulation distortions (or $H_2(f_1, -f_2)$ and $H_2(f_2, f_2)$) can be identified by supplying only one sinusoidal signal first to measure harmonic distortion (or $H_2(f_2, f_2)$) and by subtracting the harmonic distortion from the total distortion subject to two sinusoidal signals to obtain intermodulation distortion.

4. Experimental Results

The second-order Volterra approch described in the previous section has been utilized to model, identify, and quantify the linear response and a second-order distortion product of a loudspeaker in the low-frequency region. The arrangement of Fig. 1 was used to measure the distortion performance of the compression driver typical of those used to drive acoustic horns. The noise source used puts out a flat noise spectrum bandlimited to 500 Hz. A low pass filter with a cutoff frequency of 420 Hz was inserted before the amplifier which was adjusted to drive the compression driver at 10 W. A Digital Signal Processor (DSP 56000) took 98,304 data samples from the input and output signal at a 1 kHz sampling rate. These were divided up into 384 segments of 128 samples each to get the spectral estimates in (10).

Since the input signal from the noise generator to the nonlinear loudspeaker system is not sufficiently Gaussian [5], one has to solve (6) and (7) simultaneously for $H_1(m)$ and $H_2(i,j)$. The linear and quadratic transfer functions are estimated according the solution given by (10). Contour plot of the estimated $H_2(i,j)$ (smoothed over adjacent frequencies), accounting for second-order distortion, is shown in Fig.2(bottom). Note that it is enough to plot only the upper and lower triangular region of $H_2(f_i, f_j)$ in the f_i and f_j plane because of various symmetry properties. Values in the upper triangular region, where $f_i \geq 0$, $f_j \geq 0$, $f_i \geq f_j$, represent distortions at sum frequencies, while values at the lower triangular region, where $f_i \geq 0$, $f_j \leq 0$, $f_i \geq |f_j|$ represents distortions at difference frequencies. The values located on the 45° line in the upper triangular region correspond to harmonic distortion while the others correspond to intermodulation distortions.

To compare the proposed technique with the conventional analog two-tone method, $H_2(i,j)$ is measured by the method described in Section 3. Two oscillators, both of them having the same amplitude, are applied to the amplifier. The output of the amplifier, which generates a sinusoidal signal at 10 W, is connected to the component under test to calculate the distortions at sum frequencies and difference frequencies by (17). $H_2(i,j)$ in Fig.2(top), measured by the two-tone input method, shows good agreement with the one obtained using the random input approach except for relatively small fluctuations.

Distortion factors of the loudspeaker can be calculated by substituting the estimated transfer functions into (3) and (5). Reasonable agreement between the two-tone input method and the random input method is obtained except for small fluctuations. For additional discussion see [3].

However, there are trade-offs between the two-tone input approach and random input approach on estimating second-order distortion. In the two-tone input approach, it is possible to measure accurate second-order distortions, although measuring such distortions over the band of frequencies of interest is a tedious and time-consuming job. On the other hand, the random input approach does not require repeating the same experiments by changing the frequency of input since distortions over the band of frequencies of interest can be estimated with one data set. Another advantage of the random input approach is that one can quantify distortion over the entire frequency band using higher-order coherence spectra [3]. Also, random signals are a closer approximation to actual input signals, e.g., music and speech. The disadvantage of the random input approach lies in the fact that it requires a large number (e.g., hundreds) of ensemble averages to obtain good estimates.

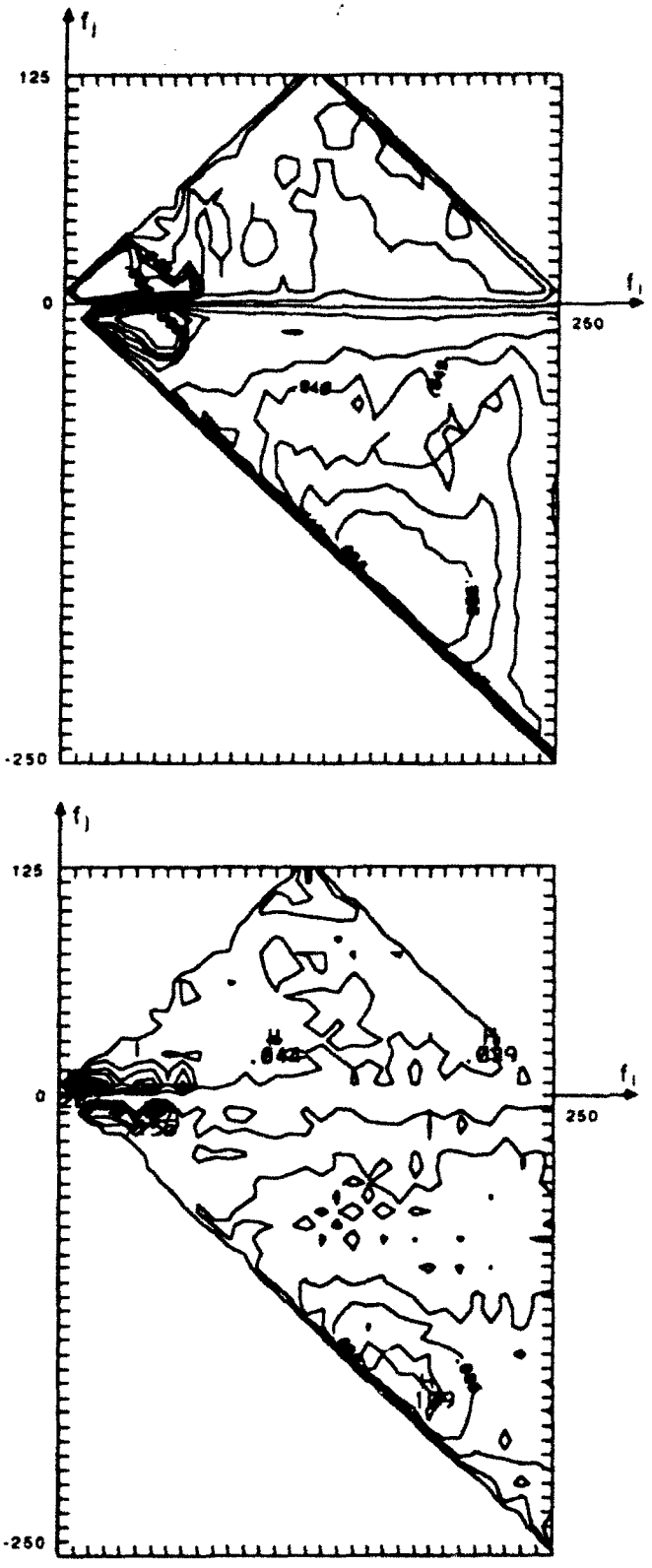


Figure 2: Contour plot of the loudspeaker quadratic transfer function $H_2(i,j)$ experimentally estimated using: (a) two-tone input method (top), (b) random input method (bottom)

5. Summary

The results of this investigation indicate that distortion products of a system can be estimated with relatively high accuracy by the random input approach if there are a sufficient number of realizations of raw data, minimal input noise, no unmodeled dynamics, and uncorrelated output noise. The results obtained in this paper can be also easily extended to handle higher-order (e.g., third-order) distortion of the system. Because of the generality of the Volterra approach, the results of this investigation can be applied to measure nonlinear distortion introduced by any type of system, including audio systems and communication systems.

Acknowledgement

This work is supported by the DOD Joint Service Electronics Program under contract AFOSR F49620-89-C-0044, and in part (Volterra kernel estimation) by Department of Navy Contracts N00167-88-K-0049 and N00014-88-K-0638.

References

[1] F.F. Mazda, *Electronic Instruments and Measurement Techniques*, Cambridge: Cambridge University Press, 1987.

[2] J.S. Bendat, *Nonlinear System Analysis and Identification From Random Data*, Wiley-Interscience, New York, 1990.

[3] Y.S. Cho, S.B. Kim, E.L. Hixson, and E.J. Powers, "A digital technique to estimate second-order distortion using higher-order coherence spectra," (to be appeared on May 1992 issue) *IEEE Tran. on Signal Processing*.

[4] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear System*, New York: Wiley, 1980.

[5] Y.S. Cho, W.T. Oh, S.B. Kim, and E.J. Powers, "Testing for Gaussianity in nonlinear system identification," *IEEE 1990 Int. Symp. Circuits and Systems*, vol. 2, pp. 1450-1453, 1990.

[6] K.I. Kim and E.J. Powers, "A digital method of modeling quadratically nonlinear systems with a general random input," *IEEE Trans. on ASSP*, vol. 36, no. 11, pp. 1758-1769, Nov. 1988.

ICASSP-92

VOLUME 5

D_2EV

DIGITAL SIGNAL PROCESSING 2
ESTIMATION
VLSI

1992
IEEE INTERNATIONAL CONFERENCE
ON
ACOUSTICS, SPEECH AND
SIGNAL PROCESSING

MARCH 23-26, 1992
THE SAN FRANCISCO MARRIOTT
SAN FRANCISCO, CALIFORNIA



IEEE
92CH3103-9

Sponsored by
THE INSTITUTE OF ELECTRICAL AND
ELECTRONICS ENGINEERS,
SIGNAL PROCESSING SOCIETY

UTILIZATION OF CONDITIONED HIGHER-ORDER SPECTRA FOR NONLINEAR SYSTEM IDENTIFICATION

S. B. Kim and E. J. Powers

Department of Electrical and Computer Engineering, and
Electronics Research Center
The University of Texas at Austin
Austin, Texas 78712, U. S. A.

ABSTRACT

For higher-order Volterra systems with arbitrary stationary inputs, an orthogonalization scheme, based upon ordered sets of conditioned orthogonal higher-order input vectors in the frequency domain using Gram-Schmidt method is presented. A new set of nonlinear transfer functions for the conditioned orthogonal Volterra system is obtained from a coordinate transformation of a Volterra system based upon non-orthogonal higher-order input vectors. The orthogonal model has the same minimum mean square error as the original system but without hybrid coherences, thus allowing physically meaningful interpretation of nonlinearities of systems.

1. Introduction

In 1961, Tick[1] developed expressions for linear and quadratic frequency-domain Volterra kernels in terms of cross-power spectra and cross-bispectra, respectively. These results are valid for Gaussian excitations only and lead to an orthogonal Volterra model. In recognition the fact that many experimental signals used in nonlinear system identification are not sufficiently Gaussian to make use of Tick's results[1], Kim and Powers[2] developed an approach to estimate linear and quadratic frequency-domain Volterra kernels which is valid for general (i.e., nonGaussian as well as Gaussian) random inputs. However, their model is not orthogonal. Thus, when using the model to "predict" the power spectrum of the response of a nonlinear physical system, one must deal with "interference" terms, corresponding to cross-product interactions between the outputs of the linear and quadratic components of the model, which greatly hinder a physical interpretation of the model. For example, the "interference" terms may result in a positive or negative contribution to the output power spectrum, depending on the relative phases of the outputs of the linear and quadratic components of the model. Thus, the objective of this paper is to present an approach for Volterra kernel estimation which is valid for general random inputs while at the same time eliminating the presence of the interference terms commonly associated with non-orthogonal models.

The approach is based upon an extension to higher-order Volterra systems of the work done by Dodds and Robson[3], and Bendat[4], for the multiple input/output linear system identification problem. Recently, Archer and Gardner developed methods which exploit cyclostationary inputs to obtain orthogonality properties associated with the Volterra kernels[5]. Korenberg and Paarmann studied orthognal search method to find significant model terms to system identification in the time domain[6].

In this paper, the Gram-Schmidt orthogonalization procedure is utilized to generate conditioned orthogonal polyspectral inputs in the frequency domain for arbitrary stationary inputs with arbitrary statistics. Coordinate transformations are utilized to relate the orthogonal Volterra kernels to the non-orthogonal Volterra kernels. This allows one to develop an orthogonal model of the Volterra series so that each order term in the series is orthogonal to the others. Thus, a partitioning of the output power spectra and a simple iterative representation of system transfer functions becomes possible without introducing interference phenomena.

We derive higher-order conditioned auto- and cross-spectral moment matrices in terms of conventional higher-order spectral moment matrices. We also demonstrate that the optimum solution for the nonlinear system transfer functions of the conditioned orthogonal model has the same minimum mean square error as the conventional nonlinear system transfer functions of a non-orthogonal Volterra model estimated using minimum mean square error criteria.

0-7803-0532-9/92 $3.00 © 1992 IEEE

2. Conditioned Orthogonal Input Vector for Higher-Order Volterra System

A higher-order nonlinear Volterra system can be regarded as a multiple inputs/single-output linear system. Partial coherence between the inputs exist since the multiple high-order inputs are generated from a single input time series. In such cases, high-order nonlinear system can be partially modeled as low-order nonlinear system, and vice versa. Thus, it hinders a physical interpretation of the system identification.

In order to remove the partial coherence between the multiple high-order inputs, thus rendering them orthogonal, one can utilize the Gram-Schmidt orthogonalization procedure. Such conditioning of multiple input vectors allows us to find the system transfer functions straightforwardly, and the relationship between the transfer functions of the orthogonal Volterra system and the non-orthogonal Volterra system can be easily found.

A frequency-domain rth-order discrete Volterra series representation of the nonlinear system can be written as follows

$$\begin{aligned} Y(m) - N(m) &= H_1(m)X(m) \\ &+ \sum_{i+j=m} H_2(i,j)X(i)X(j) \\ &+ \sum_{i+j+k=m} H_3(i,j,k)X(i)X(j)X(k) + \cdots \\ &+ \sum_{i_1+\cdots+i_r=m} H_r(i_1,\cdots,i_r)X(i_1)\cdots X(i_r) \\ &= \mathbf{H}(m)\mathbf{X}(m) \end{aligned} \tag{1}$$

where $X(m)$ and $N(m)$ are the discrete Fourier transforms of the input $x(t)$ and the noise $n(t)$, respectively. The row vector $\mathbf{H}(m)$ and the column vector $\mathbf{X}(m)$ are given by

$$\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_r] \tag{2}$$

$$\mathbf{X}^t = [\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_r] \tag{3}$$

where the superscript 't' is the transpose of a matrix. In Eqs. (2) and (3), the row vectors $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_r$ are linear, quadratic, and rth-order transfer functions, respectively. The column vectors $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_r$ are Fourier transforms of the linear, quadratic, and rth-order inputs $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_r$, respectively. For example, the linear transfer function and input are given by

$$\mathbf{H}_1(m) = H_1(m) \tag{4}$$

$$\mathbf{X}_1(m) = X(m) \tag{5}$$

where $0 \le m \le M$, and M is the Nyquist frequency. The quadratic transfer function and input vectors can be written as

$$\mathbf{H}_2(m) = [\ \cdots H_2(i,j) \cdots\ ] \tag{6}$$

$$\mathbf{X}_2(m) = \begin{bmatrix} \vdots \\ I(i,j)X(i)X(j) \\ \vdots \end{bmatrix} \tag{7}$$

where $\lceil (m+1)/2 \rceil \le i = m - j \le M$, $\lceil \ell \rceil$ is the largest integer less than ℓ, $I(i,j) = 2$ for $i \ne j$, and $I(i,j) = 1$ otherwise. The cubic transfer function and input vectors are given by[7]

$$\mathbf{H}_3(m) = [\ \cdots H_3(i,j,k) \cdots\ ] \tag{8}$$

$$\mathbf{X}_3(m) = \begin{bmatrix} \vdots \\ I(i,j,k)X(i)X(j)X(k) \\ \vdots \end{bmatrix} \tag{9}$$

where the indices i, j, and k span

$$\begin{aligned} \lceil (m+2)/3 \rceil = m_1 \le\ & i \le M \\ \lceil (m - m_1 + 1)/2 \rceil \le\ & j \le m, \text{ for each } i \\ & k = m - i - j \end{aligned}$$

The symmetry factor $I(i,j,k)$ is

$$I(i,j,k) = \begin{cases} 6 & \text{if } i \ne j \ne k \\ 3 & \text{if two of three indices are same} \\ 1 & \text{if } i = j = k \end{cases}$$

For the higher-order input vectors defined above, their polyspectral matrices are defined by

$$\mathbf{S}_{ji} = \langle \mathbf{X}_j \mathbf{X}_i^{\wedge} \rangle \tag{10}$$

where $\langle \cdot \rangle$ represents a expectation operator, and the superscript '$\wedge$' denotes a Hermitian of a matrix.

The alternative orthogonal model can be constructed by replacing the original given input records by ordered sets of conditioned input records $\{\mathbf{Z}_i = \mathbf{X}_{i \cdot i-1}\}$, $i = 1, 2, \ldots, r$. The dependence upon frequency f of functions is omitted for simplicity in notation. The conditioned input $\mathbf{Z}_i$ represents $\mathbf{X}_i$ in which the linear effects of $\mathbf{X}_1$ to $\mathbf{X}_{i-1}$ have been removed by the Gram-Schmidt orthogonalization procedure.

Suppose a nonlinear Volterra system with multiple input vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_r$, which are given as Eqs. (5), (7), etc. Clearly, the system can be represented with any linear combination of these input vectors. Let

$$\mathbf{Z}_1 = \mathbf{X}_1 \tag{11}$$

$$\mathbf{Z}_2 = \mathbf{X}_2 - \mathbf{L}_{21}\mathbf{Z}_1 \tag{12}$$

We wish to choose $\mathbf{L}_{21}$ so that $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are orthogonal in such a way that $\langle \mathbf{Z}_2 \mathbf{Z}_1^{\wedge} \rangle = 0$. Then one can show that

$$\mathbf{L}_{21} = \langle \mathbf{X}_2 \mathbf{Z}_1^{\wedge} \rangle \langle \mathbf{Z}_1 \mathbf{Z}_1^{\wedge} \rangle^{-1} = \mathbf{S}_{21} \mathbf{S}_{11}^{-1} \tag{13}$$

where $\mathbf{S}_{21}$ is the auto-bispectrum and $\mathbf{S}_{11}$ is the auto-power spectrum of the input $\mathbf{X}_1$. Notice that the expression given for $\mathbf{L}_{21}$ is the same as the transfer function when $\mathbf{X}_2$ is modeled with $\mathbf{X}_1$. Thus one can interpret the orthogonalized input $\mathbf{Z}_2$ as representing $\mathbf{X}_2$ in which the linear effect of $\mathbf{X}_1$ has been removed.

Now, one can repeat the orthogonalization procedure by adding a new orthogonal basis vector and applying orthogonal conditions each time. Therefore, in general, for an rth-order nonlinear Volterra system, one can write $\mathbf{X}_j$ in terms of conditioned orthogonal input vectors $\mathbf{Z}_i$, where $i, j = 1, \ldots, r$, as follows:

$$\mathbf{X}_j = \sum_{i=1}^{j} \mathbf{L}_{ji} \mathbf{Z}_i \tag{14}$$

with $\langle \mathbf{Z}_i \mathbf{Z}_j \rangle = 0$ for $i \neq j$. Furthermore, one can easily obtain

$$\mathbf{L}_{ji} = \langle \mathbf{X}_j \mathbf{Z}_i^{\wedge} \rangle \langle \mathbf{Z}_i \mathbf{Z}_i^{\wedge} \rangle^{-1} \tag{15}$$

In this alternative conditioned system, one can regard the noise term N as the output Y $(= \mathbf{X}_{r+1})$ conditioned on all of the input records $\mathbf{X}_1$ to $\mathbf{X}_r$, and thus, is given by $\mathbf{Z}_{r+1}$. In this way, one can extend the system order to $r+1$ in Eqs. (14) and (15). Notice that the noise N and the output Y of the conditioned orthogonal system remain same as those of the original system.

3. New Higher-Order Transfer Functions for Conditioned Orthogonal Volterra System

As we have new conditioned orthogonal coordinate system with the basis of $\mathbf{Z}_i$, one can obtain new components of the conditioned input vectors, i.e. transfer functions$\mathbf{G}_i$, according to the coordinate transformation from the system with the basis $\mathbf{X}_j$ to the orthogonal system with the basis $\mathbf{Z}_i$. Using the two different coordinate systems, one can represent the output Y as follows,

$$\begin{aligned} Y - N &= \mathbf{GZ} \\ &= \mathbf{HX} \\ &= \mathbf{HLZ} \end{aligned} \tag{16}$$

where the vectors $\mathbf{G}$ and $\mathbf{Z}$ are given by

$$\mathbf{G} = [\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_r] \tag{17}$$

$$\mathbf{Z}^t = [\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_r] \tag{18}$$

The row vectors $\mathbf{G}_1$, $\mathbf{G}_2$, and $\mathbf{G}_r$ are linear, quadratic, and rth-order transfer functions (or components) of the conditioned orthogonal system, respectively. The column vectors $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_r$ are linear, quadratic, and rth-order conditioned orthogonal high-order inputs, respectively.The transformation matrix $\mathbf{L}$ is a block unit lower triangular matrix given by

$$\mathbf{L} = \begin{bmatrix} \mathbf{I}_{11} & & & \\ \mathbf{L}_{21} & \mathbf{I}_{22} & & \\ \mathbf{L}_{31} & \mathbf{L}_{32} & \mathbf{I}_{33} & \\ \vdots & & & \\ \mathbf{L}_{r1} & \mathbf{L}_{r2} & \cdots & \mathbf{L}_{rr} \end{bmatrix} \tag{19}$$

where $\mathbf{I}_{ii}$ $(= \mathbf{L}_{ii})$ is an identity matrix, and $\mathbf{L}_{ji}$ is given in Eq. (15). From Eq. (16), one can immediately obtain the new set of conditioned orthogonal transfer functions $\mathbf{G}$,

$$\mathbf{G} = \mathbf{HL} \tag{20}$$

$$= \langle Y \mathbf{Z}^{\wedge} \rangle \langle \mathbf{Z}\mathbf{Z}^{\wedge} \rangle^{-1} \tag{21}$$

where $\langle \mathbf{Z}\mathbf{Z}^{\wedge} \rangle$ is a block diagonal matrix with ith diagonal element $\langle \mathbf{Z}_i \mathbf{Z}_i^{\wedge} \rangle$. Now the orthogonal transfer function $\mathbf{G}$ can be easily calculated using the conditioned polyspectral matrix.

For a cubically nonlinear Volterra system, the conditioned orthogonal transfer functions can be obtained using Eqs. (19) and (20),

$$\begin{aligned} \mathbf{G}_1 &= \mathbf{H}_1 + \mathbf{H}_2 \mathbf{L}_{21} + \mathbf{H}_3 \mathbf{L}_{31} \\ \mathbf{G}_2 &= \mathbf{H}_2 + \mathbf{H}_3 \mathbf{L}_{32} \\ \mathbf{G}_3 &= \mathbf{H}_3 \end{aligned}$$

Note that the cubic transfer functions of two alternative systems are identical. It is also possible to reverse order of the input vectors. Then, the linear transfer functions of the two alternative systems become identical, and the higher-order transfer functions are modified successively.

Since, $\mathbf{L}_{ii}$ is an identity matrix, one can rewrite Eq. (20) for $\mathbf{H}_i$,

$$\mathbf{H}_i = \mathbf{G}_i - \sum_{j=i+1}^{r} \mathbf{H}_j \mathbf{L}_{ji} \tag{22}$$

For a cubically nonlinear Volterra system,

$$\begin{aligned} \mathbf{H}_3 &= \mathbf{G}_3 \\ \mathbf{H}_2 &= \mathbf{G}_2 - \mathbf{H}_3 \mathbf{L}_{32} \\ \mathbf{H}_1 &= \mathbf{G}_1 - \mathbf{H}_2 \mathbf{L}_{21} - \mathbf{H}_3 \mathbf{L}_{31} \end{aligned}$$

Thus, one can find the transfer functions of the original system recursively using the new set of transfer functions of the conditioned orthogonal system.

In order to find explicit expressions for the transfer functions of the conditioned system using Eq. (20), the conditioned polyspectral matrices are needed in terms of the polyspectral matrices of the original input vectors, which will be described in the next section.

4. Recursive Relationship of Conditioned Polyspectral Matrix

A recursive relationship between the conditioned polyspectral matrix can be found by rewriting Eq. (14) as follows,

$$\begin{aligned} \mathbf{X}_j &= \sum_{\ell=1}^{j} \mathbf{L}_{j\ell}\mathbf{Z}_\ell \\ &= \sum_{\ell=1}^{k} \mathbf{L}_{j\ell}\mathbf{Z}_\ell + \mathbf{X}_{j\cdot k} \end{aligned} \tag{23}$$

Then, one can write

$$\mathbf{X}_{j\cdot k} = \mathbf{X}_{j\cdot k-1} - \mathbf{L}_{jk}\mathbf{X}_{k\cdot k-1} \tag{24}$$

where $\mathbf{X}_{k\cdot k-1} = \mathbf{Z}_k$. By taking an expectation after multiplying $\mathbf{X}_i$ in Eq. (24) for $j \geq i \geq k$, one can obtain

$$\mathbf{S}_{ji\cdot k} = \mathbf{S}_{ji\cdot k-1} - \mathbf{L}_{jk}\mathbf{S}_{ki\cdot k-1} \tag{25}$$

where $\mathbf{S}_{ji\cdot k}$ is a conditioned polyspectral matrix obtained from jth-order and ith-order input vectors in which the linear effects of $\mathbf{X}_1$ to $\mathbf{X}_k$ have been removed.

As an example, for a cubic nonlinear system, one can find the following from Eqs. (15) and (25),

$$\mathbf{L}_{21} = \mathbf{S}_{21}\mathbf{S}_{11}^{-1}, \quad \mathbf{L}_{31} = \mathbf{S}_{31}\mathbf{S}_{11}^{-1}, \quad \mathbf{L}_{32} = \mathbf{S}_{32\cdot 1}\mathbf{S}_{22\cdot 1}^{-1}$$

and,

$$\begin{aligned} \mathbf{S}_{32\cdot 1} &= \mathbf{S}_{32} - \mathbf{S}_{31}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \\ \mathbf{S}_{22\cdot 1} &= \mathbf{S}_{22} - \mathbf{S}_{21}\mathbf{S}_{11}^{-1}\mathbf{S}_{12} \end{aligned}$$

Using Eq. (20) and the above identities, we can see the conditioned orthogonal system is equivalent to the original non-orthogonal Volterra system. Thus, if $\mathbf{H}$ is a set of optimum transfer functions in the original non-orthogonal system, then $\mathbf{G}$ is also a set of transfer functions in the conditioned orthogonal system with the same minimum mean square error.

The decomposition method derived in this paper reconfigures the partially coherent higher-order nonlinear input vectors of a Volterra system into a set of mutually incoherent input vectors so that the total model output power consists of the sum of the outputs due to each of its conditioned orthogonal input vector. The advantage of this method is that the performance of a system model can be improved successively by increasing the order of nonlinearity of the system model without modification of the results of previous model.

ACKNOWLEDGEMENT

This work is supported in part by the Texas Advanced Technology Program Project 003658-392, NSF Grant CDR-8721512, Department of Navy Contract N00014-88-K-0638, and the Joint Services Electronics Program AFOSR Contract F-49620-89-C-0044.

REFERENCES

[1] L. J. Tick, "*The Estimation of Transfer Functions of Quadratic System*," Technometrics, Vol. 3, pp. 563-567, 1961.

[2] K. I. Kim, and E. J. Powers, "*A Digital Method of Modelling Quadratically Nonlinear Systems with a General Random Input*," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 36, pp. 1758-1769, 1988.

[3] C. J. Dodds and J. D. Robson, "*Partial Coherence in Multivariate random processes*," Jnl. Sound and Vibration, Vol. 42, pp. 243-249, 1975.

[4] J. S. Bendat, "*System Identification from Multiple Input/Output Data*," Jnl. Sound and Vibration, Vol. 49, pp. 293-308, 1976.

[5] T. L. Archer and W. A. Gardner, "*New Methods for Identifying the Volterra Kernels of a Nonlinear System*," Twenty-Fourth Annual Asiloma Conference on Signals, Systems and Computers, Nov. 5-7, 1990.

[6] M. K. Korenberg and L. D. Paarmann, "*Orthogonal Approaches to Time-Series Analysis and System Identification*," IEEE Signal Processing Magazine, pp. 29-43, July, 1991.

[7] S. W. Nam, E. J. Powers, and S. B. Kim, "*On the Identification of a Third-Order Volterra Nonlinear System Using a Frequency Domain Block RLS Adaptive Algorithm*," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Albuquerque, New Mexico, pp. 2407 - 2410, 1990.

Edward J. Powers

ISSPA 92

THE THIRD INTERNATIONAL SYMPOSIUM ON SIGNAL PROCESSING AND ITS APPLICATIONS

PROCEEDINGS

August 16 - 21, 1992
ANA Hotel Gold Coast
Queensland Australia

Incorporating: **HOSSPA 92**
INTERNATIONAL CONFERENCE ON
HIGHER-ORDER STATISTICAL SIGNAL PROCESSING WITH
APPLICATION TO NONLINEAR, NONGAUSSIAN,
NONSTATIONARY SIGNALS AND SYSTEMS

IEEE

IASTED

The Institution of Engineers, Australia

QUT
School of Electrical and Electronic Systems Engineering
Queensland University of Technology

Chairman: Prof B Boashash
Signal Processing Research Centre, QUT, Brisbane, Queensland

Editor: Dr D A Gray

ISSPA 92, Signal Processing and its Applications.
Gold Coast, Australia, August 16-21, 1992.
Editor: D Gray

Time Domain Cubic System Identification Using Higher-Order Correlations of I.I.D. Signals

C-H Tseng
E J Powers
Department of Electrical and Computer Engineering
The University of Texas at Austin, Texas, USA

Identification of cubically nonlinear systems with an i.i.d. input is investigated in this paper. A simple algorithm for estimating the Volterra kernels of cubic systems is derived and presented. It will be shown that the computational complexity is greatly reduced compared to conventional methods.

1. INTRODUCTION

Volterra models have been widely used to identify nonlinear systems. Identification of nonlinear systems using such models involves estimation of the Volterra kernels, which is usually computationally intensive. A simple algorithm to estimate the Volterra kernels of quadratic nonlinear systems with an i.i.d. (independent, identically distributed) input has been developed recently[1]. But some nonlinear systems (for example, a nonlinear communication channel[2]) can not be modeled effectively by a quadratic Volterra series. For these nonlinear systems, Volterra models higher than second order are required. The objective of this paper is to develop a simple algorithm to estimate the Volterra kernels of cubically nonlinear systems, when subjected to an i.i.d. input. If conventional methods are used to identify the Volterra kernels of cubic systems, estimation of higher-order correlations with various time lags up to 6th order is necessary. In addition, the required calculation of an inverse of the input correlation matrix is computationally intensive, especially when the system time delay is long. By using the properties of higher-order correlations of an i.i.d. signal, we can greatly reduce the complexity of the estimation of Volterra kernels. A simple algorithm for estimating the Volterra kernels of cubic systems is derived and presented. This algorithm needs neither the estimation of higher-order correlations for various time lags nor the calculation of an inverse matrix. It will be shown that the reduction of complexity is significant.

2. VOLTERRA MODELING OF NONLINEAR SYSTEMS

A third-order Volterra model can be written as follows:

$$y(n) = \sum_{i=0}^{N-1} h_1(i)x(n-i) + \sum_{i=0}^{N-1}\sum_{j=i}^{N-1} h_2(i,j)x(n-i)x(n-j)$$

$$+\sum_{i=0}^{N-1}\sum_{j=i}^{N-1}\sum_{k=j}^{N-1} h_3(i,j,k)x(n-i)x(n-j)x(n-k)+e(n) \qquad (1)$$

Where $x(n)$ and $y(n)$ are the input and the output of a cubically nonlinear system, $h_1(i)$, $h_2(i,j)$, and $h_3(i,j,k)$ are the linear, quadratic, and cubic Volterra kernels respectively, and $e(n)$ is the additive noise. In this third-order Volterra model, although the relationship between the input and the output is nonlinear, the relationship between the output and the filter coefficients is linear. Therefore, the Volterra kernels can be estimated by solving the following equation[3][4]:

$$\mathbf{Rh} = \mathbf{p} \qquad (2)$$

where

$$\mathbf{h} = [h_1(0),\ldots,h_3(N-1,N-1,N-1)]^T \qquad (3)$$

$$\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}(n)^T] \qquad (4)$$

$$\mathbf{p} = E[\mathbf{x}(n)y(n)] \qquad (5)$$

$$\mathbf{x}(n) = [x(n),\ldots,x(n-N+1),x^2(n),\ldots,x^2(n-N+1), x^3(n),\ldots,x^3(n-N+1)]^T \qquad (6)$$

It is noticed that the input vector $\mathbf{x}(n)$ includes not only linear terms, but also quadratic and cubic terms. Therefore, the input autocorrelation matrix $\mathbf{R}$ consists of higher-order moments of the input up to 6th order.

3. PROPERTIES OF AN I.I.D. SIGNAL

The i.i.d. signal plays an important roll in many engineering areas. For example, in digital communications, the input signal is usually considered to be i.i.d. Thermal noise and noise generators, which are frequently used in acoustics, can also be regarded as i.i.d.[6]. For a zero-mean i.i.d. signal, the second-order correlation can be determined as follows:

$$E[x(n_1)x(n_2)] = \begin{cases} \mu_2 & \text{if } n_1 = n_2 \\ 0 & \text{otherwise} \end{cases} \qquad (7)$$

where the property $E[x(n_1)x(n_2)] = E[x(n_1)]E[x(n_2)]$ if $n_1 \neq n_2$ for an i.i.d. signal has been used, and μ_2 is equal to $E[x^2(n)]$. Correlations higher than second-order can be determined in a similar manner. Using the above properties, it can be shown that the higher-order correlations of an i.i.d. signal up to 6th order can be completely characterized by μ_2, μ_3, μ_4, μ_5, and μ_6, where μ_i denotes the ith-order correlation of $x(n)$, $E[x^i(n)]$.

4. THE CUBIC I.I.D. ALGORITHM

In the general case, the Volterra kernels of a cubic system can be estimated by solving Eq. 2. For a cubic system, the input autocorrelation matrix $\mathbf{R}$ includes higher-order correlations of the input for various time lags up to 6th order. It is also noticed that the estimation process involves calculation of the inverse of the $\mathbf{R}$ matrix, $\mathbf{R}^{-1}$,

which results in $O(L^3)$ complexity, where L is the total number of filter coefficients. L is usually very large, especially when the order of the nonlinear system is high (say, cubic) and/or the memory of the system is long. Therefore the whole estimation process is very computationally intensive. Some research has been devoted to relieving the computational burden. A simple algorithm for identifying quadratic nonlinear systems with an i.i.d. input has been developed[1]. But in some situations, the result of using quadratic Volterra model is not satisfactory. For example, in digital communications, a nonlinear communication channel can be modeled by odd-order terms of a Volterra series[2]. To take such nonlinear effects into account, adding at least the cubic term is required. For these nonlinear systems, a simple algorithm to estimate Volterra kernels of cubically nonlinear systems is highly desirable.

For a cubically nonlinear system with an i.i.d. input, the kernel estimation process can be simplified by using the properties of an i.i.d. signal mentioned in section 3. The input correlation matrix $\mathbf{R}$ for such a system can be completely determined as long as μ_2, μ_3, μ_4, μ_5, and μ_6 are known. In addition, since most elements of the $\mathbf{R}$ matrix are actually zero, the $\mathbf{R}$ matrix can be reduced to a sparse matrix. Thus, with some algebra, a simple algorithm for estimating the Volterra kernels of cubic systems can be derived. In this algorithm, the cubic Volterra kernels are estimated by the following formula:

$$\begin{aligned} h_3(k,l,m) &= E[y(n)x(n-k)x(n-l)x(n-m)]/\mu_2^3, \\ &\quad 0 \le k < l < m \le N-1 \end{aligned} \tag{8}$$

$$\begin{aligned} h_3(k,l,l) &= \{\mu_3 E[y(n)x(n-k)x(n-l)] - \mu_2 E[y(n)x(n-k)x^2(n-l)] \\ &\quad +\mu_2^2 E[y(n)x(n-k)]\}/r, \\ &\quad 0 \le k,l \le N-1, k \ne l \end{aligned} \tag{9}$$

where

$$r = \mu_3^2\mu_2 - \mu_2^2(\mu_4 - \mu_2^2) \tag{10}$$

After the cubic Volterra kernels are determined, the quadratic Volterra kernels can be estimated as follows:

$$\begin{aligned} h_2(k,l) &= \{E[y(n)x(n-k)x(n-l)] - \mu_2\mu_3[h_3(k,l,l) \\ &\quad +h_3(l,k,k)]\}/\mu_2^2, \\ &\quad 0 \le k < l \le N-1 \end{aligned} \tag{11}$$

Similarly, the linear Volterra kernels can be determined as long as the cubic and the quadratic kernels are known:

$$\begin{aligned} h_1(k) &= \{E[y(n)x(n-k)] - \mu_2^2 H_3(k) - \mu_3 h_2(k,k) \\ &\quad -\mu_4 h_3(k,k,k)\}/\mu_2, \\ &\quad 0 \le k \le N-1 \end{aligned} \tag{12}$$

where $H_3(k)$ is defined as follows:

$$\begin{aligned} H_3(k) &= \sum_{i=0, i\ne k}^{N-1} h_3(k,i,i), \\ &\quad 0 \le k \le N-1 \end{aligned} \tag{13}$$

Note that in this algorithm the Volterra kernels are expressed in terms of μ_2, μ_3, μ_4, μ_5, μ_6 and higher-order cross-correlations between the input and the output. There is no need to calculate the higher-order correlations of the input for various time lags. Furthermore, since there are no matrix operations involved, the computational complexity is greatly reduced.

5. CONCLUSION

A simple algorithm for estimating the Volterra kernels of cubic systems subjected to an i.i.d. input is presented in this paper. One of the applications of the i.i.d. algorithm is in the digital communication area, where the input signal for digital communication is usually considered to be i.i.d.

6. ACKNOWLEDGEMENT

This work is supported by the Joint Services Electronics Program through AFOSR F49620-89-c-0044.

References

[1] Y. S. Cho and E. J. Powers, "*Estimation of Quadratically Nonlinear Systems with an I.I.D. Input*," IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 5, pp. 3117-3120, 1991.

[2] S. Benedetto, E. Biglieri and V. Castellani, "*Digital Transmission Theory*," Prentice-Hall, Inc., 1987.

[3] T. Koh and E.J. Powers, "*Second-Order Volterra Filtering and Its Application to Nonlinear System Identification*," IEEE Trans on Acoustics, Speech, and Signal Processing, Vol. 33, No. 6, pp. 1445-1455, Dec. 1985.

[4] K. H. Kim, S. B. Kim, and E. J. Powers, "*Fast RLS Algorithms for General Filters*," Proceedings of the Fifth European Signal Processing Conference, Barcelona, Spain, September 1990.

[5] C. E. Davila and A. J. Welch, "*A Second-Order Adaptive Volterra Filter with Rapid Convergence*," IEEE Trans on Acoustics, Speech, and Signal Processing, Vol. 35, No. 9, pp. 1259-1263, Sept. 1987.

[6] F. F. Mazda, "*Electronic Instruments and Measurement Techniques*," Cambridge: Cambridge University Press 1987.

Proceedings of the Second (1992) International Offshore and Polar Engineering Conference
San Francisco, USA, 14-19 June 1992
Copyright © 1992 by The International Society of Offshore and Polar Engineers
ISBN 1-880653-00-1 (Set); ISBN 1-880653-03-6 (Vol III)

IDENTIFICATION OF QUADRATIC DRIFT RESPONSE OF TLP'S USING CONDITIONED ORTHOGONAL QFRF'S

S.B. Kim, E.J. Powers and R.W. Miksad
University of Texas at Austin
Austin, Texas, USA

F.J. Fischer
Shell Development Company
Houston, Texas, USA

ABSTRACT

We present a new orthogonal approach for the estimation of nonlinear FRF's which is valid for general random waves (i.e., nonGaussian as well as Gaussian) while at the same time removing the presence of the interference terms associated with non-Gaussian waves. The approach of this paper is illustrated by using it to quantify the linear and quadratically nonlinear dynamical response of TLP's to random nonGaussian sea wave excitation, and by comparing the performance of the conventional non-orthogonal Volterra model with that of the new conditioned orthogonal model.

KEY WORDS: TLP's, Orthogonal Volterra System Model, NonGaussian Interference, Nonlinear wave interaction, Polyspectral Analysis

INTRODUCTION

This paper is concerned with the application of digital bispectral analysis techniques (Powers and Miksad, 1987) to experimentally detect and quantify nonlinear wave interaction phenomena associated with quadratic response of marine structures such as tension leg platforms (TLP) in nonGaussian random seas, without "interference" of linear wave force and quadratic wave drift force due to the nonGaussian nature of the wave excitation (S. B. Kim et al., 1989). The presence of the "interference" terms greatly hinder a physical interpretation of the model. For example, the "interference" terms may result in a positive or negative contribution to the power spectrum of the response predicted by the model, depending on the relative phases of the linear and quadratic components of the model (K. I. Kim et al., 1987). Thus, the objective of this paper is to present a new approach for the estimation of nonlinear FRF's which is valid for general random waves (i.e., nonGaussian as well as Gaussian) while at the same time removing the presence of the interference associated with non-Gaussian wave.

In this paper, we derive conditioned orthogonal polyspectral moment matrices of the wave excitation in terms of conventional polyspectral moment matrices utilizing the Gram-Schmidt orthogonalization procedure which do not depend on the wave statistics. Some accounts of the useful ideas of conditioned (or residual) spectra and partial coherence has been given in the books of Jenkins and Watts (Jenkins and Watts, 1968) and of Bendat and Piersol (Bendat and Piersol, 1971). Such concepts are generalized by Dodds and Robson (Dodds and Robson, 1975) for multivariate stationary random processes. Bendat (Bendat, 1976) utilized the concepts for linear system identification from multiple input/output data. Even though they developed the ideas from a spectral density matrix from a set of mutually incoherent white noise sources, it is consistent with the ideas of Gram-Schmidt orthogonalization procedure.

We generalize the concept of conditioned orthogonal spectra for nonlinear system identification with higher-order Volterra system model. Such an orthogonalised model is different from Barrett (or Wiener) orthogonal series in such a way that the latter is constructed from a hierarchy of functionals of increasing order that are orthogonal to each other with respect to a white Gaussian (or Gaussian) input and whose sum characterizes the system (Barrett, 1963). In our model, an orthogonalization procedure is applied to the multiple order input vectors for arbitrary input statistics. Using this orthogonal model of the Volterra series, a partitioning of the response power spectra and a simple iterative representation of FRF's becomes possible without introducing interference phenomena. We obtain the relationship between the new FRF's of the conditioned orthogonal system and the FRF's of the non-orthogonal Volterra system for general random wave excitation by utilizing the concept of coordinate transformation.

Finally, we illustrate the approach of this paper by using it to quantify the linear and quadratically nonlinear dynamical response of tension leg platforms to random nonGaussian sea wave excitation, and by comparing the performance of the conventional non-orthogonal Volterra model and associated "prediction" of the TLP response power spectra with that of the new conditioned orthogonal model described in this paper. Of particular importance is the fact that the conditioned orthogonal model allows one to decompose the observed response into its linear and quadratic components without the uncertainty associated with the nonGaussian interference term.

ORTHOGONALIZATION OF THE HIGHER-ORDER INPUT VECTORS

This is a generalization of multi-input linear system identification with conditioned inputs carried out by Bendat (Bendat, 1976), to include higher-order Volterra system with arbitrary inputs. The necessity to include nonGaussian inputs in higher-order Volterra models is discussed in the references of K. I. Kim (K. I. Kim et al., 1987) and S. B. Kim (S. B. Kim et al., 1989), and in the collection of works in the area of *nonGaussian* signal processing (Wegman et al., 1989). Such generalization using multi-input system concepts is possible since one can regard higher-order Volterra series as linear in terms of the Volterra kernels (or frequency response functions), though it is nonlinear in terms of the input. In our model, each higher-order input is regarded as an independent input so that a higher-order nonlinear time-invariant Volterra system becomes a multiple input/output system. In this paper, since we are dealing with quadratic drift response of a TLP due to random waves, we are going to limit ourselves to second-order Volterra system, even though one can generalize the model to any order nonlinear Volterra system.

Lets first consider a nonlinear, time invariant analytic system which can be described with a second-order Volterra series,

$$\begin{aligned} y(t) - n(t) &= \int h_1(\tau)x(t-\tau)d\tau \\ &+ \int\int h_2(\tau_1,\tau_2)x(t-\tau_1)x(t-\tau_2)d\tau_2 d\tau_1 \end{aligned} \quad (1)$$

where $x(t)$, $y(t)$, and $n(t)$ are the time series of the input wave, the output response, and the error (or noise) of the model, and $h_1(\tau)$ and $h_2(\tau_1, \tau_2)$ are the linear and the quadratic Volterra kernels, respectively.

Let's look at the system described in Eq. (1) in the frequency domain by performing a discrete Fourier transform. Then, the second-order discrete Volterra series representation of the nonlinear system can be written as follows

$$Y(m) = H_1(m)X(m) + \sum_{i+j=m} H_2(i,j)X(i)X(j) + N(m) \quad (2)$$

where $X(m)$, $Y(m)$, $N(m)$, $H_1(m)$, $H_2(i,j)$ are discrete Fourier transforms of $x(t)$, $y(t)$, $n(t)$, $h_1(\tau)$, $h_2(\tau_1,\tau_2)$, respectively.

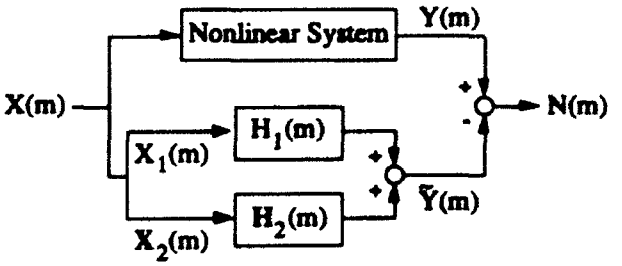


Figure 1: Second-Order Multiple Input System Model.

Notice that Eq. (2) is linear in terms of the frequency response functions $H_1(m)$, $H_2(i,j)$, though the equation is nonlinear with respect to the input wave $X(m)$. Thus, by regarding $X(i)X(j)$ as a separate input, one can treat the second-order nonlinear system as a two-input linear system, as shown in Fig. 1,

$$\begin{aligned} Y(m) &= \hat{Y}(m) + N(m) \\ &= \mathbf{H}_1(m)\mathbf{X}_1(m) + \mathbf{H}_2(m)\mathbf{X}_2(m) + N(m) \end{aligned} \quad (3)$$

where $\hat{Y}(m)$ is the model output. In Eq. (3), the linear frequency response function $\mathbf{H}_1(m)$ and the first-order input $\mathbf{X}_1(m)$ are given by

$$\mathbf{H}_1(m) = H_1(m) \quad (4)$$

$$\mathbf{X}_1(m) = X(m) \quad (5)$$

where $0 \le m \le M$, and M is the Nyquist frequency. The quadratic frequency response function $\mathbf{H}_2(m)$ and second-order input vector $\mathbf{X}_2(m)$ can be written as

$$\mathbf{H}_2(m) = [\ \cdots H_2(i,j) \cdots\ ] \quad (6)$$

$$\mathbf{X}_2(m) = \begin{bmatrix} \vdots \\ I(i,j)X(i)X(j) \\ \vdots \end{bmatrix} \quad (7)$$

where $\lceil (m+1)/2 \rceil \le i = m - j \le M$, $\lceil \ell \rceil$ is the largest integer less than ℓ, $I(i,j) = 2$ for $i \ne j$, and $I(i,j) = 1$ otherwise.

In general, the two inputs $\mathbf{X}_1(m)$ and $\mathbf{X}_2(m)$ are not independent, especially for nonGaussian input where the cross-coherence between the two inputs $(\mathbf{X}_1(m)\mathbf{X}_2^{\Lambda}(m))$, which is the auto-bispectrum of the input wave, is not zero. The superscript 'Λ' denotes a Hermitian of a matrix and $(\cdot)$ represents a expectation operator.

As shown in Fig. 2, the alternative orthogonal model can be constructed by replacing the original input records by ordered sets of conditioned input records $\mathbf{Z}_1(m)$ and $\mathbf{Z}_2(m)$, where the conditioned input $\mathbf{Z}_2(m)$ represents $\mathbf{X}_2(m)$ in which the linear effects of $\mathbf{X}_1(m)$ have been removed by the Gram-Schmidt orthogonalization procedure.

Suppose a second-order Volterra system with two-input vectors $\mathbf{X}_1(m)$ and $\mathbf{X}_2(m)$ which are given as Eqs. (5) and (7).

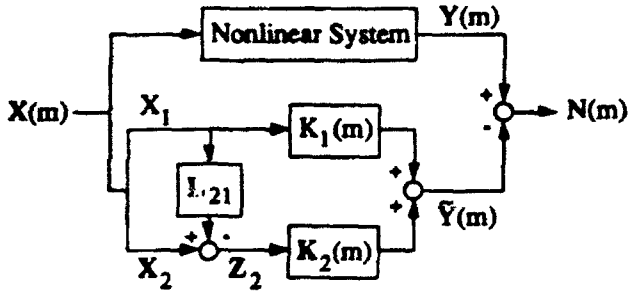


Figure 2: Second-Order Conditioned Orthogonal Multiple Input System Model.

Clearly, the system can be represented with any linear combination of these input vectors. Let

$$\mathbf{Z}_1(m) = \mathbf{X}_1(m) \tag{8}$$
$$\mathbf{Z}_2(m) = \mathbf{X}_2(m) - \mathbf{L}_{21}(m)\mathbf{Z}_1(m) \tag{9}$$

We wish to choose $\mathbf{L}_{21}(m)$ so that $\mathbf{Z}_1(m)$ and $\mathbf{Z}_2(m)$ are orthogonal in such a way that $\langle \mathbf{Z}_2(m)\mathbf{Z}_1^A(m)\rangle = \mathbf{0}$. Then one can show that

$$\begin{aligned}\mathbf{L}_{21}(m) &= \langle \mathbf{X}_2(m)\mathbf{Z}_1^A(m)\rangle \langle \mathbf{Z}_1(m)\mathbf{Z}_1^A(m)\rangle^{-1} \\ &= \mathbf{S}_{21}(m)\mathbf{S}_{11}^{-1}(m)\end{aligned} \tag{10}$$

where $\mathbf{S}_{11}(m)$ is the auto-power spectrum and $\mathbf{S}_{21}(m)$ is the auto-bispectral vector of the input wave $\mathbf{X}_1(m)$, which are given as

$$\mathbf{S}_{11}(m) = \langle X(m)X^*(m)\rangle \tag{11}$$
$$\mathbf{S}_{21}(m) = \langle \mathbf{X}_2(m)\mathbf{Z}_1^A(m)\rangle = \begin{bmatrix} \vdots \\ I(i,j)\langle X(i)X(j)X^*(m)\rangle \\ \vdots \end{bmatrix} \tag{12}$$

where $i + j = m$. Notice that the expression given for $\mathbf{L}_{21}(m)$ is the same as the frequency response function when $\mathbf{X}_2(m)$ is modeled with $\mathbf{X}_1(m)$. Thus one can interpret the orthogonalized input $\mathbf{Z}_2(m)$ as representing $\mathbf{X}_2(m)$ in which the linear effect of $\mathbf{X}_1(m)$ has been removed.

HIGHER-ORDER FRF'S FOR CONDITIONED ORTHOGONAL VOLTERRA SYSTEM

Now, one can obtain the new linear and the quadratic frequency response functions $\mathbf{K}_1(m)$ and $\mathbf{K}_2(m)$ from the conditioned orthogonal coordinate system with the basis of $\mathbf{Z}_1(m)$ and $\mathbf{Z}_2(m)$,

$$Y(m) = \mathbf{K}_1(m)\mathbf{Z}_1(m) + \mathbf{K}_2(m)\mathbf{Z}_2(m) + N(m) \tag{13}$$

where $\mathbf{Z}_1(m)$ and $\mathbf{Z}_2(m)$ are linear and quadratic conditioned orthogonal inputs, respectively. Notice that the noise $N(m)$ of the conditioned orthogonal system remains the same as that of the original system.

Considering the orthogonal properties of the two conditioned inputs $\mathbf{Z}_1(m)$ and $\mathbf{Z}_2(m)$, one can find the new FRF's $\mathbf{K}_1(m)$ and $\mathbf{K}_2(m)$ of the condioned orthogonal system using least mean square criterion in terms of the original input spectral matrices as follows

$$\mathbf{K}_1(m) = \mathbf{S}_{y1}(m)\mathbf{S}_{11}^{-1}(m) \tag{14}$$
$$\mathbf{K}_2(m) = \bar{\mathbf{S}}_{y2}(m)\bar{\mathbf{S}}_{22}^{-1}(m) \tag{15}$$

In Eq. (15), the numerator $\bar{\mathbf{S}}_{y2}(m)$ is the wave-surge cross bispectrum where the apparent contribution due to nonGaussian quadratic wave-wave interference is subtracted. The denominator $\bar{\mathbf{S}}_{22}(m)$ is the forth-order polyspectrum due to four wave interactions where the apparent contribution due to nonGaussian quadratic wave-wave interference is subtracted. Notice that the fourth-order spectral matrix $\bar{\mathbf{S}}_{22}(m)$ consists of an element $\langle X(i)X(j)X(k)^*X(l)^*\rangle$, where $i + j = k + l = m$.

Thus, $\mathbf{K}_2(m)$ is the QFRF with arbitrary input wave without the apparent contribution due to nonGaussian wave-wave interference. Also note that $\mathbf{K}_1(m)$ is identical with the LFRF for Gaussian input. The QFRF $\mathbf{K}_2(m)$ of the conditioned orthogonal model, which is identical with the QFRF for non-orthogonal model, is reduced to the QFRF for Gaussian model when the bispectrum $\mathbf{S}_{12}(m)$ (or $\mathbf{S}_{21}(m)$) is zero.

One can relate the new FRF's $\mathbf{K}_1(m)$ and $\mathbf{K}_2(m)$ of the conditioned orthogonal system with the FRF's $\mathbf{H}_1(m)$ and $\mathbf{H}_2(m)$ of the Volterra system with nonGaussian inputs, according to the coordinate transformation from the system with the basis $\mathbf{X}_1(m)$ and $\mathbf{X}_2(m)$ to the orthogonal system with the basis $\mathbf{Z}_1(m)$ and $\mathbf{Z}_2(m)$. For a quadratically nonlinear Volterra system, one can obtain the conditioned orthogonal frequency response functions using Eqs. (8) $\sim$ (10),

$$\begin{aligned}\mathbf{K}_1(m) &= \mathbf{H}_1(m) + \mathbf{H}_2(m)\mathbf{L}_{21}(m) \\ \mathbf{K}_2(m) &= \mathbf{H}_2(m)\end{aligned} \tag{16}$$

Note that the QFRF's of two alternative quadratic Volterra systems are identical. Using the Eqs. (10) $\sim$ (12), (14) and (15), one can find the expression for the linear and quadratic model power spectra $\mathbf{P}_1(m)$ and $\mathbf{P}_2(m)$ for the orthogonal Volterra system as follows

$$\mathbf{P}_1(m) = \mathbf{S}_{y1}(m)\mathbf{S}_{11}^{-1}(m)\mathbf{S}_{1y}(m) \tag{17}$$
$$\mathbf{P}_2(m) = \bar{\mathbf{S}}_{y2}(m)\bar{\mathbf{S}}_{22}^{-1}(m)\bar{\mathbf{S}}_{y2}^{A}(m) \tag{18}$$

The the total model power spectrum $\mathbf{S}_{yy}(m)$ for the orthogonal Volterra model becomes

$$\begin{aligned}\mathbf{S}_{yy}(m) &= \mathbf{P}_1(m) + \mathbf{P}_2(m) \\ &= \mathbf{H}_1(m)\mathbf{S}_{11}(m)\mathbf{H}_1^A(m) + \mathbf{H}_2(m)\mathbf{S}_{22}(m)\mathbf{H}_2^A(m) \\ &+ 2Re[\mathbf{H}_2(m)\mathbf{S}_{21}(m)\mathbf{H}_1^A(m)] \\ &= \mathbf{P}_L(m) + \mathbf{P}_Q(m) + \mathbf{P}_{LQ}(m)\end{aligned} \tag{19}$$

which is the same as the total model power spectrum for the non-orthogonal Volterra model. Thus, two alternative systems have the same model error power, and both are optimum systems. In

Eq. (19), $\mathbf{P}_L(m)$, $\mathbf{P}_Q(m)$ and $\mathbf{P}_{LQ}(m)$ are the linear, quadratic, and hybrid model power spectra for the non-orthogonal Volterra system, respectively.

The coherence functions, which are defined as the ratios of the model power spectra to the measured output power spectrum, are then calculated using Eqs. (17) ~ (19),

$$\begin{aligned}\gamma^2_{yy}(m) &= \gamma^2_1(m) + \gamma^2_2(m) \\ &= \gamma^2_L(m) + \gamma^2_Q(m) + \gamma^2_{LQ}(m) \qquad (20)\end{aligned}$$

where $\gamma^2_1(m)$ and $\gamma^2_2(m)$ are the linear and the quadratic coherence functions of the conditioned orthogonal system, and $\gamma^2_L(m)$, $\gamma^2_Q(m)$, and $\gamma^2_{LQ}(m)$ are the linear, quadratic, and hybrid cohernece functions for the non-orthogonal Volterra system, respectively.

Notice that the expressions of $\mathbf{P}_1(m)$ and $\mathbf{P}_2(m)$ in Eqs. (17) and (18) are positive definite so that the linear and the quadratic coherence functions $\gamma^2_1(m)$ and $\gamma^2_2(m)$ of the conditioned orthogonal system are bounded between zero and one. However, for non-orthogonal system, the boundness is not guaranteed so that the linear and the quadratic coherence functions $\gamma^2_L(m)$ and $\gamma^2_Q(m)$ of the non-orthogonal system can be greater than one. Furthermore, the hybrid coherence function $\gamma^2_{LQ}(m)$ can be negative depending upon the phase of $\mathbf{H}_2(m)\mathbf{S}_{21}(m)\mathbf{H}_1^*(m)$. Such characteristics of coherence functions of non-orthogonal system cause great difficulties in meaningful physical interpretation of the analysis results.

EXPERIMENTAL RESULTS

In this section we compare the results of the polyspectral analysis using non-orthogonal second-order Volterra model with the results of the analysis using orthogonal Volterra model. The experimental data we obtained from a scaled (1:54) model test of a prototype TLP anchored in 1500 feet of water and is described elsewhere (S. B. Kim et al., 1989). The TLP was excited using 45^0 uni-directional irregular wave of significant wave height 40 feet. Waves were generated according to the Pierson-Moskowitz spectrum. The wave and surge data were sampled at a rate of 17 Hz which resulted in a data record of 4260 samples. Since our major interest was in the low-frequency resonance drift oscillation of TLP surge, we resampled the data using a reduced sampling rate of 2.125 Hz for the higher-order spectral analysis.

We analyzed the surge response, which contains both linear wave response around 0.5 Hz and the response due to quadratic difference frequency interactions at the surge resonance band around 0.07 Hz (S. B. Kim et al., 1989), using the two alternative system model. In Fig. 3, we present the linear and quadratic coherence functions of wave-surge time series of the TLP using the non-orthogonal model (Fig. 3-a) and the conditioned orthogonal model (Fig. 3-b). Overall features of the two plots are similar in the sense that both the quadratic coherence functions $\gamma^2_Q(f)$ of the non-orthogonal model and $\gamma^2_2(f)$ of the conditioned orthogonal model are quite high in the frequency bands below 0.2 Hz and above 0.8 Hz, whereas the linear coherence functions $\gamma^2_L(f)$ and $\gamma^2_1(f)$ are negligible in these same frequency bands.

On the other hand, the linear coherence functions $\gamma^2_L(f)$ of the non-orthogonal model and $\gamma^2_1(f)$ of the conditioned orthogonal model are close to unity in the wave-excitation band centered at approximately 0.5 Hz. However, there are frequency zones where $\gamma^2_L(f)$ and $\gamma^2_Q(f)$ in Fig. 3-a are larger than one and $\gamma^2_{LQ}(f)$ is negative. Furthermore, there is large variation of coherence with frequency. Those behaviors of the coherence functions obtained using a non-orthogonal model greatly hinders physical interpretation of the analysis results. In the case of the conditioned orthogonal model, as one can see in Fig. 3-b, the linear and quadratic coherence functions $\gamma^2_1(f)$ and $\gamma^2_2(f)$ are bounded between zero and one at all frequencies and there is no hybrid coherence due to nonGaussian interference. They are also much more stable, as one can see in the figure.

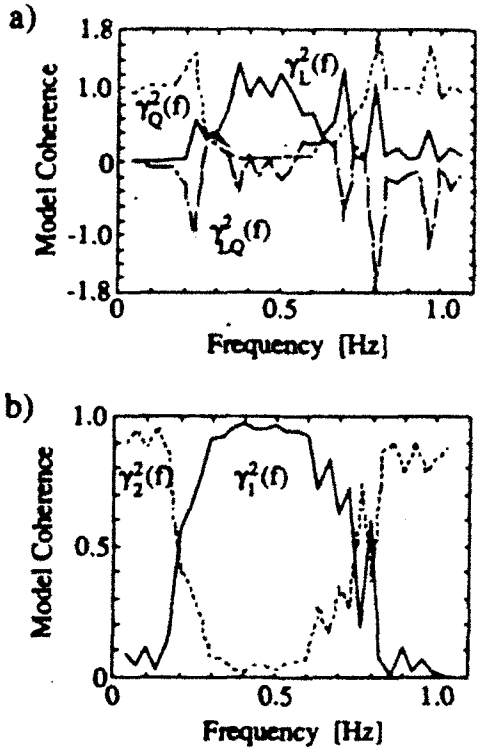


Figure 3: TLP Surge Model Coherence Functions. a) Non-Orthogonal Model, b) Conditioned Orthogonal Model

In Fig. 4, we compare the linear, the hybrid and the quadratic model power spectra $P_L(f)$, $P_{LQ}(f)$, and $P_Q(f)$ of wave-surge time series of the TLP using the non-orthogonal model (Fig. 4-a) with the linear and the quadratic model power spectra $P_1(f)$ and $P_2(f)$ of the conditioned orthogonal model (Fig. 4-b). In Fig. 4-a, the absolute value of the hybrid power spectrum $|P_{LQ}(f)|$ is plotted due to the fact that it is negative at some frequencies. The magnitude of the hybrid power spectrum cannot be greater than the sum of the linear and quadratic model powers, however, it is certainly greater than the linear model power within the

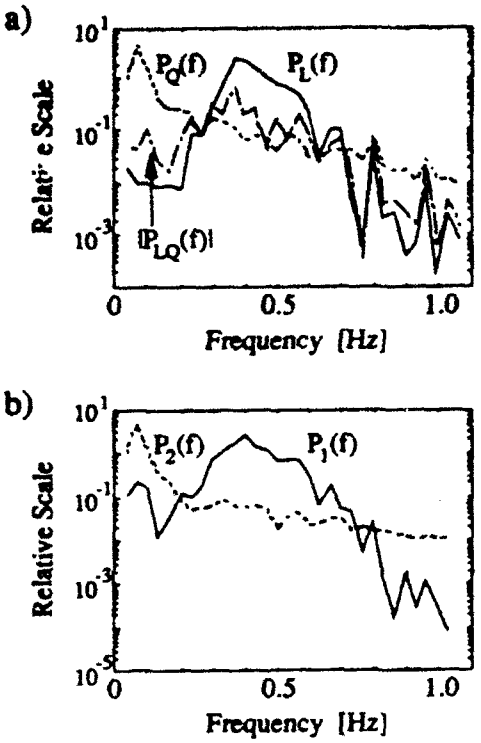


Figure 4: TLP Surge Model Power Spectra. a) Non-Orthogonal Model, b) Conditioned Orthogonal Model

frequency band where quadratic interactions are dominant in our example. On the other hand, we don't see such ill effects in the model power spectra estimated using the conditioned orthogonal model shown in Fig. 4-b, though the total system model power spectra of the two systems are identical as we have proved in the previous section.

As we discussed earlier, the successful outcome of the conditioned orthogonal model results from the fact that the apparent nonGaussian interference effects are assigned in such a way that no hybrid interactions between the linear and the quadratic models exist utilizing the orthogonal procedure, whereas in the non-orthogonal model the effects are shared by both the linear and the quadratic system models. In fact, there is no unique way to assign such nonGaussian interference effects. However, we follow the idea that no results from the linear system analysis should be changed due to the effects, which one can see from the fact that we have the same expression of the LFRF $\mathbf{K}_1(f)$ as the LFRF obtained assuming linear system model only (i.e., $\mathbf{H}_2(f) \equiv 0$). Thus, one can successively improve the system model by adding higher-order terms in the Volterra system without modifying the lower order FRF's.

SUMMARY

By utilizing digital polyspectral analysis techniques incorporated with higher-order Volterra system models, we are able to identify the effects on the response of a TLP due to the nonGaussian interference of wave-wave interactions. This was possible with the new tools we have developed to remove the interference effects by conditioning the input wave so that each higher-order input wave vectors are orthogonal each other. We derived the new conditioned input vectors and their polyspectra using original input polyspectral matrix, and the relationship between the FRF's of the orthogonal system and of the non-orthogonal system is obtained using the concept of coordinate transformation. The application to the scaled model TLP test experimental data yielded improved results over the non-orthogonal model without ill effects due to nonGaussian interference.

ACKNOWLEDGEMENT

The authors' work on nonlinear hydrodynamics is sponsored by *ONR contract N00014-88-4-0638*, and on applications of nonlinear system identification to TLP's by the Texas Advanced Technology Program (Grants 4604 and 00365-392) and in part, by the Engineering Research Centers Program NSF Grant CDR-8721512. Much of the nonlinear signal analysis techniques used in this paper were developed under the auspices of the Joint Services Electronics Program, Contract AFOSR F49620-86-C-0045. Appreciation by the authors is extended to Brown & Root International Inc., Shell Oil Company, Sun Exploration and Production Company and Texaco Inc. for supplying the TLP model test data referenced in this paper.

REFERENCES

Barrett, J. R. (1963), "*The Use of Functionals in the analysis of nonlinear physical systems*," Jnl. Electron Control, Vol. 15, pp. 567-615.

Bendat, J. S. and Piersol, A. G. (1971), "*Random Data: Analysis and Measurement Procedures*," New York, Wiley-Interscience.

Bendat, J. S. (1976), "*System Identification from Multiple Input/Output Data*," Jnl. Sound and Vibration, Vol. 49, pp. 293-308.

Dodds, C. J. and Robson, J. D. (1975), "*Partial Coherence in Multivariate random processes*," Jnl. Sound and Vibration, Vol. 42, pp. 243-249.

Kim, K. I., Powers, E. J., Ritz, Ch. P., Miksad, R. W., and Fischer, F. J. (1987), "*Modeling of the Nonlinear Drift Oscillations of Moored Vessels Subject to NonGaussian Random Sea Wave Excitation*," IEEE Jnl. of Oceanic Engineering, OE-12,

pp. 568-575.

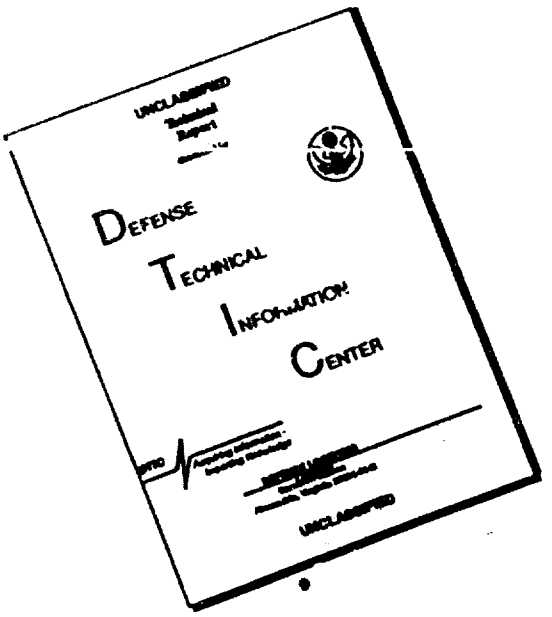
Kim, S. B., Powers, E. J., Miksad, R. W.,Fischer, F. J., and Hong, J. Y. (1989), "*Nonlinear System Coherence Analysis of the Surge Response of Tension Leg Platforms Subject to Non-Gaussian Irregular Seas*," Eighth Int. Conf. on OMAE, Vol. 1, pp. 167-173.

Jenkins, J. M. and Watts, D. G. (1968), "*Spectral Analysis and Its Applications*," San Francisco: Holden-Day.

Powers, E. J., and Miksad, R. W. (1987), "*Polyspectral Measurement and Analysis of Nonlinear Wave Interactions*," Proc. of the Symposium on Nonlinear Wave Interactions in Fluids, Boston, Dec. 13-18, pp. 9-16.

Wegman, E. J., Schwartz, S. C., and Thomas, J. B., ed. (1989), "*Topics in Non-Gaussian Signal Processing*," Springer-Verlag.

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.