

AD-A166 576

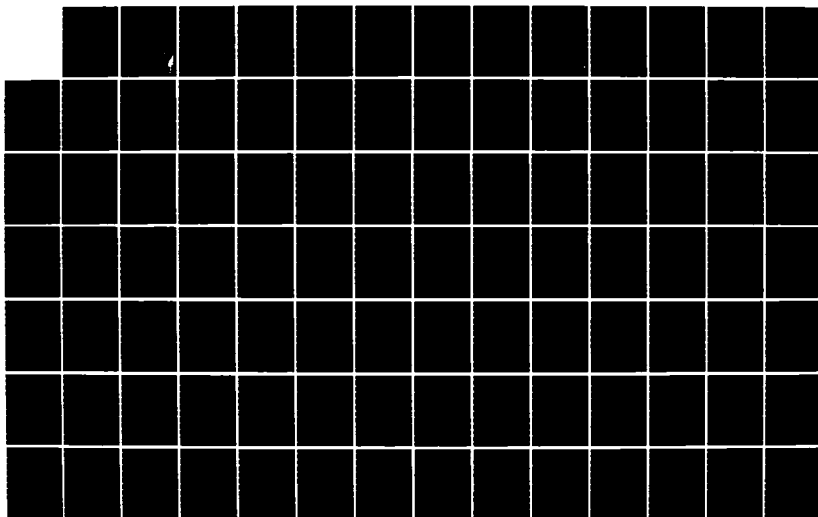
RESEARCH ON INTEGRATED NEAR-MILLIMETER-WAVE RECEIVER
(U) CALIFORNIA UNIV BERKELEY ELECTRONICS RESEARCH LAB
S E SCHWARZ 01 OCT 82 DAAK70-80-C-0134

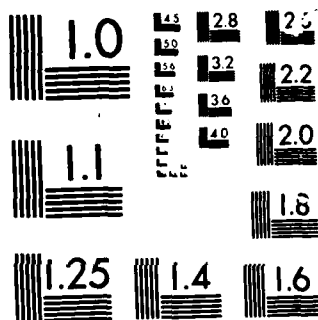
1/2

UNCLASSIFIED

F/G 9/5

NL





MICROCOPY

CHART

AD-A166 576

②
1 Oct 1982
Final Report

Research on Integrated Near-
Millimeter-Wave Receiver

U.S. Army Mobility Equipment Research
and Development Command

Contract DAAK70-80-C-0134

S. E. Schwarz

Department of Electrical Engineering and Computer Science
and the Electronics Research Laboratory
University of California, Berkeley, California, 94720

DTIC
ELECTE
APR 14 1986
S D A

100-111-0074
This document has been approved
for public release and sale; its
distribution is unlimited.

86

4

10

DISCLAIMER NOTICE

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

Abstract

This report deals with problems in the design of monolithic millimeter-wave circuits. Emphasis is placed on: (A) design of dielectric ridge waveguides for millimeter-wave integrated circuits, and (B) a monolithic Schottky diode imaging array at 69 GHz.

The general problem of the design of a specific all-dielectric waveguide, the dielectric ridge guide, is considered. A planar metallic V-coupler is designed to couple energy efficiently between the waveguide and lumped circuit elements. The principal method used is simulation in the range of 2-7 GHz. With good compromise design, typical coupling loss between waveguide and a small device is about 1.4 dB, exclusive of dielectric loss and ohmic loss in the coupler. The use of dielectric ridge guide is attractive especially at the higher end of the millimeter-wave spectrum where metallic loss is excessive.

Monolithic Schottky diode imaging arrays with integral antennas have been built and tested at 69 GHz. An immersion-lens optical system is used for imaging. The measured video responsivity is 534 V/W, from which an optical coupling loss of 7 dB is deduced. The video NEP is $1.45 \times 10^{-11} \text{ W}/\sqrt{\text{Hz}}$. In heterodyne detection experiments, SSB conversion loss of 19.4 dB and DSB NEP of $7 \times 10^{-19} \text{ W/Hz}$ were measured. Although the performance of each elemental receiver is inferior to that of state-of-the-art discrete receiver, imaging times for the array will be less than those required for scanning a single antenna.

QUALITY
INSPECTED
3

By Distribution	Availability Codes Avail and/or Special	Dist A123
--------------------	---	--

TABLE OF CONTENTS

	Page Nos.
CHAPTER 1. MONOLITHIC MILLIMETER-WAVE TECHNOLOGY	1
1.1. Introduction	1
1.2. Dielectric Ridge Waveguides for M ³ IC	3
1.3. Monolithic Schottky Diode Imaging Arrays	8
CHAPTER 2. THEORY OF DIELECTRIC RIDGE WAVEGUIDES	17
2.1. Introduction	17
2.2. Dielectric Slab Waveguides	18
2.2.1. Guided Modes	18
2.2.2. Radiation and Evanescent Modes	25
2.2.3. Orthogonality Relations	26
2.3. Reflection from a Step Discontinuity in a Slab Guide	28
2.3.1. Exact Formulation	31
2.3.2. Approximate Solutions	37
2.4. Dielectric Ridge Waveguides	41
2.4.1. Guided Modes	41
2.4.2. Propagation Constant and Cutoff Frequencies of Higer Order Modes	45
2.5. Radiation from Waveguide Bends	46
2.5.1. Introduction	46
2.5.2. Mode-Matching Analysis	49
CHAPTER 3. DESIGN OF DIELECTRIC RIDGE WAVEGUIDE	56
3.1. The Effective Dielectric Constant Method	56
3.2. Design of the Waveguide	61
3.3. Design of the V-Coupler	63
3.4. Conclusion	70
CHAPTER 4. THEORY OF SCHOTTKY DIODE IMAGING ARRAYS	81
4.1. Imaging Antenna Arrays	81
4.1.1. Imaging Optics	81
4.1.2. The Modified Bow-Tie Antenna	85
4.1.3. Diffraction-Limited Imaging	89

Table of Contents (continued)

	Page Nos.
4.2. Schottky Contacts	92
4.2.1. Barrier Height	93
4.2.2. Junction Capacitance and Junction Breakdown	96
4.2.3. Current Transport Processes	98
4.2.4. Noise Mechanisms	103
4.3. Video Detection	106
4.3.1. Theory of Video Detection	106
4.3.2. Noise-Equivalent Power	110
4.4. Heterodyne Detection	111
4.4.1. Theory of Mixing	112
4.4.2. Attenuator Noise Model	118
4.4.3. System Noise Temperature and NEP	120
CHAPTER 5. EXPERIMENTS WITH SCHOTTKY DIODE IMAGING ARRAYS	129
5.1. Device Fabrication Technologies	129
5.1.1. Schottky Contacts	137
5.1.2. Ohmic Contacts	138
5.1.3. Proton Implantation	140
5.2. Video Detection Experiments	140
5.2.1. Voltage Responsivity	141
5.2.2. Noise-Equivalent Power	146
5.3. Heterodyne Detection Experiments	148
5.3.1. Conversion Loss	152
5.3.2. NEP	155
5.4. Optical System Performance	160
5.4.1. Optical Coupling Loss	160
5.4.2. Electromagnetic Coupling Loss	161
5.4.3. Crosstalk	164
5.5. Conclusion	164
CHAPTER 6. FUTURE DIRECTIONS	171

CHAPTER 1

MONOLITHIC MILLIMETER-WAVE TECHNOLOGY

1.1. Introduction

J. C. Maxwell [1] wrote down his famous equations and predicted the existence of electromagnetic waves in 1864; experimental demonstrations were provided by H. R. Hertz [2] in 1888. Since then the frontiers of the electromagnetic spectrum have expanded continuously. The technology of electromagnetic waves is by now well established from DC to microwave frequencies and in the visible range. Millimeter-waves represent one of the last frontiers in the electromagnetic spectrum. The millimeter-wave frequency range is officially defined to be from 40 to 300 GHz (IEEE, 1976). The major driving forces behind developments in this frequency range are scientific applications, military applications, and the ever-increasing demand for bandwidth in communications. Some of the most interesting scientific applications are in radio astronomy, remote sensing, biological imaging and detection of cancerous tumors. Typical examples of aerospace applications are in missile guidance, radar and electronic warfare. Examples of applications in communications are provided by short-range data communications and satellite communications. (For specific system applications and references, see [3]). The list of potential applications for millimeter-waves is not limited to those just mentioned; and continues to grow as the technology advances.

Traditional millimeter-wave integrated circuits are based on a hybrid approach. Recently, because of advances in GaAs fabrication technology, monolithic millimeter-wave IC (M^3IC) technologies are becoming serious contenders in system applications [4]. By monolithic, we mean an approach wherein all active and passive circuit elements and interconnections are built into the bulk, or onto the surface, of a semi-insulating substrate

by some deposition scheme [5]. The monolithic approach offers the following advantages over the conventional hybrid approach: (1) low cost; (2) improved reliability and reproducibility; (3) small size and weight; (4) broadband performance; and (5) multifunction performance on a chip (higher level of integration). These advantages are mainly derived from elimination of bonding wires and batch processing of the circuit chip. Wire bonds have always been a serious factor in reliability and reproducibility. The parasitics introduced by bonding wire usually provide the bandwidth limitations of the circuit. The advantage of batch processing to cost, size and complexity of the chip is self-evident. These advantages become more apparent as the frequency of interest is pushed higher. As wavelength decreases, the device count per chip and chip count per wafer become larger and hence make batch processing more economical. The monolithic approach does have its potential problem areas, such as: (1) circuit tuning and trouble-shooting, and (2) the difficulty of integrating high power sources (e.g., IMPATT's). Despite these potential problems, the monolithic approach is still the most logical path for the future; these problems simply make it more challenging.

Although the advantages of the monolithic approach are evident, work along this direction has just begun (e.g., [6], [7]). Most work to date has been based on metallic transmission lines, and the finished circuit chips are placed in metallic waveguide enclosures [6]. At millimeter-wave frequencies, metallic waveguides are expensive because of the tight tolerances for machining. The cost, size, and weight advantages of the monolithic approach can hence not be fully realized. It seems that the present state of affairs is only an intermediate step toward truly monolithic millimeter-wave Integrated Circuits.

My own work at Berkeley has included two aspects of monolithic integration of millimeter-wave technology. One is the use of dielectric ridge waveguides as the transmission medium for the M^3 IC [8]. The second one is a demonstration of an imaging array of Schottky diodes in monolithic form [9], [10].

1.2. Dielectric Ridge Waveguides for M^3 IC

Dielectric waveguides are potentially useful alternatives to metallic guides at high frequencies, where metallic conduction loss become excessive. Those guides that are made entirely of dielectric (and do not use metal image planes) are of course free of metallic loss, except for losses in small metallic coupling structures. Their loss advantage over metallic guides tends to increase with increasing frequency. For the all-dielectric guides, an upper bound of the power attenuation coefficient is the bulk attenuation coefficient $\alpha = \sqrt{\mu/\epsilon}$, where σ , μ , and ϵ are the conductivity, magnetic permeability, and electric permittivity of the dielectric, respectively. Assuming that the conductivity of the dielectric is independent of frequency f , α is also independent of frequency and the loss per guide wavelength decreases as $1/f$. For metallic waveguides, on the other hand, two effects cause the loss per guide wavelength to increase with frequency; the dimensions become smaller in proportion to $1/f$, and the surface resistance increases in proportion to $\sqrt{f}$. Thus, typically, the ratio of the loss per guide wavelength for a dielectric guide to that of a metallic guide is on the order of $(f_1/f)^{3/2}$, where f_1 is a crossover frequency determined by the details of the guides in question. (Let the metallic guide be a 50- Ω microstrip made of copper on a BeO substrate with substrate thickness $h = \lambda_0/20$, where λ_0 is the free-space wavelength, and let the ratio of the copper thickness to the

substrate thickness be 1:10. Using the above-mentioned upper bound to approximate the loss of a $1000\text{-}\Omega\cdot\text{cm}$ silicon dielectric waveguide, we find a crossover frequency of 67 GHz.) Various kinds of dielectric waveguides, such as rectangular waveguide [11], [12], image guide [13], [14], [15], strip guide [16], [17], inverted strip guide [18], [19], trapped image guide [20], and dielectric ridge guide [21] have been proposed and analyzed. Waveguide components, such as directional couplers [13], ring resonator filters [22], [23], ferrite components [22], distributed Bragg reflectors [24], grating filters [25], [26], electronic phase shifters [27], frequency scanning antennas [28], beam-steerable antennas [29], and tapered waveguide antennas [30] have been investigated. Systems based on these devices have also become realities [22], [31]-[33]. All these components and systems were based on rectangular dielectric waveguide or image guides, except for the tapered waveguide antenna, which is derived from the dielectric ridge guide. When the material of the dielectric guide is a semiconductor, it becomes possible to fabricate semiconductor devices in the same piece of material [13], [27]. In the millimeter-wave regime, dimensions of quasi-single-mode (c.f. section 2.4.1) waveguides are convenient for fabrication by photolithography.

In the first part of the thesis, the design of one particular type of all-dielectric waveguide, the dielectric ridge guide (DRG) is considered. The DRG is particularly attractive because it lends itself very well to monolithic integration. a) The shape can be conveniently etched out of semiconductor wafers. The web isolates the guided field from other components on the wafer while providing structural support. b) Ohmic loss (such as that associated with dielectric image guide) is avoided. c) A convenient method exists for coupling radiation into the guide. This can be done by tapering the end of the guide into a dielectric

antenna, Fig. 1.1. Patterns obtained in this way are single-lobed and nearly symmetrical, with gain on the order of 10 dB [30]. The most promising technique for fabrication appears to be anisotropic etching of the semiconductor [7]. With this technique, the cross section of the guide is not rectangular, but rather, trapezoidal.

In the M^3IC , there can be several devices, active or passive, distributed or lumped, interconnected by dielectric waveguide lumped elements, such as diodes, a way must be found to couple energy efficiently between the waveguide and the device. At lower frequencies, this can be done by inserting a tapered section of waveguide into a hollow metallic waveguide, where the device can then be mounted on a post. In a M^3IC , where metal waveguide is to be avoided, it is convenient to use the planar V-coupler of Rutledge et al. [30]. This consists of a planar metallic V-shaped structure deposited on the back (planar) surface of the dielectric wafer, where its field interacts with those of the guide. The device to be coupled is connected at the vertex of the V. The fabrication steps of a section of a hypothetical M^3IC are shown in Fig. 1.2..

A principal objective in the design of the DRG and V-couplers is efficient coupling between the guide and the small device at the vertex of the V. It is also desirable to reduce radiation from the guide. Although little radiation occurs in unperturbed straight sections, energy will be lost into slab modes of the support web whenever there is a discontinuity or bend. Further considerations are structural strength and ease of fabrication. The support web should therefore not become too thin. Finally, we shall require quasi-single-mode propagation, in the sense that at the operating frequency, only the non-leaky wave and the lowest order leaky wave propagates.

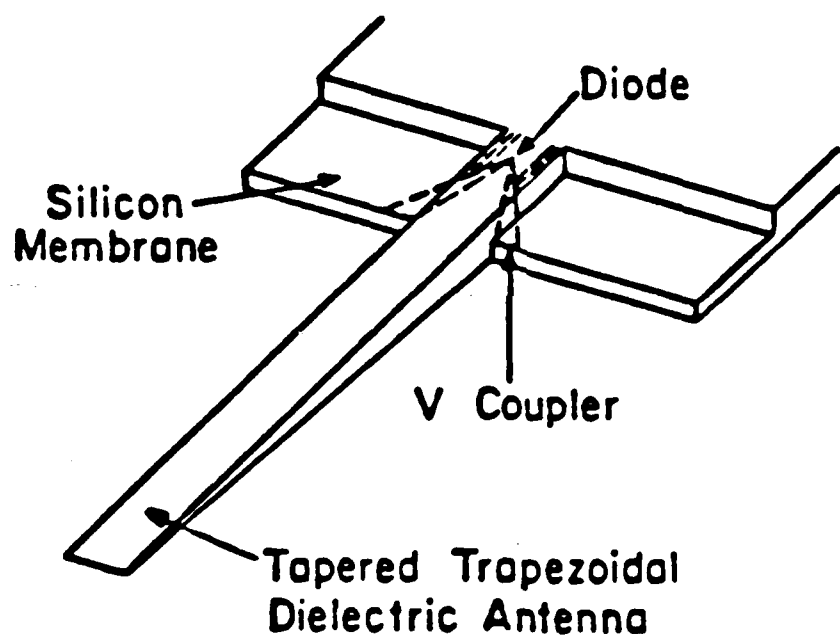


Fig. 1.1

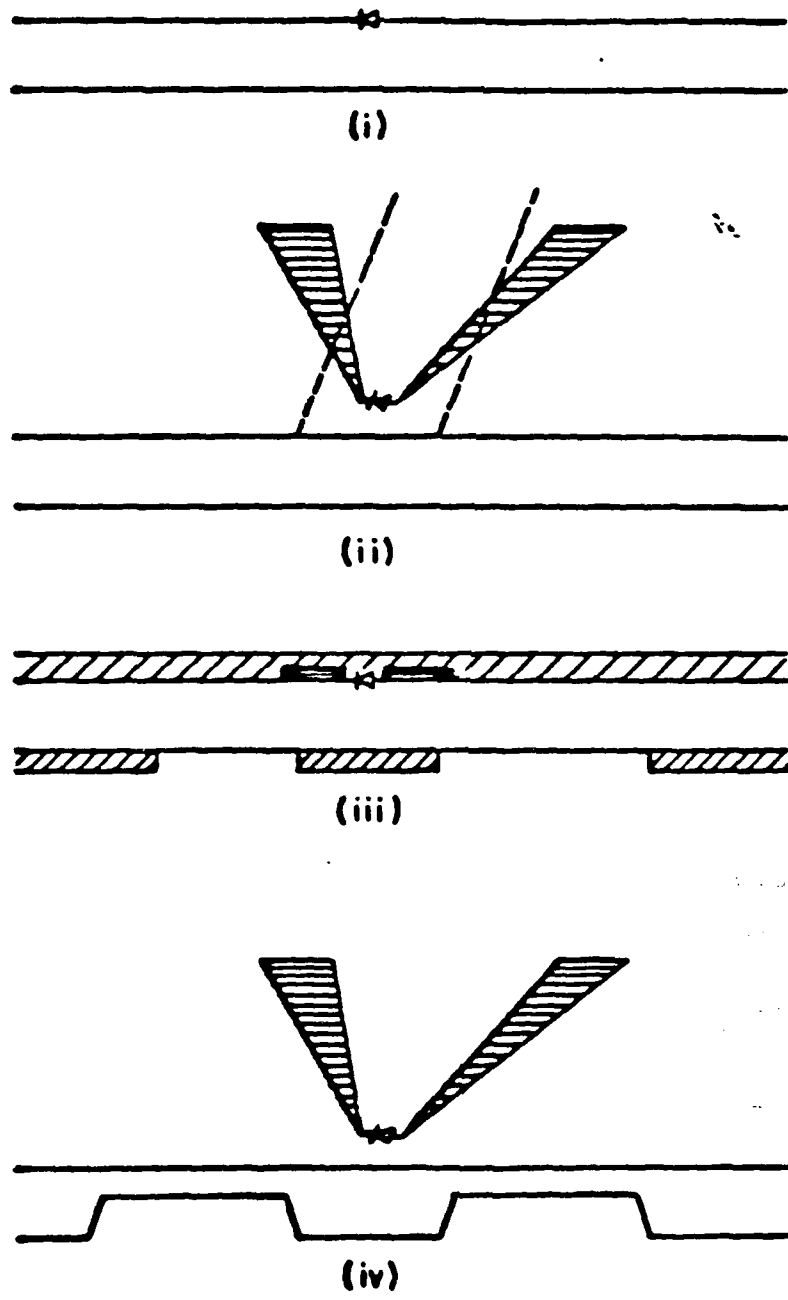


Fig. 1.2

In chapter two we examine the theory of the dielectric ridge waveguide from the unified point of view of the mode-matching method. This provides the necessary understanding and physical insight for the design of the DRG, to be taken up in chapter three.

1.3. Monolithic Schottky Diode Imaging Arrays

An important application of millimeter waves is imaging; that is, mapping the radiation of a distributed source. At the present time, most millimeter-wave imaging systems use a single detector and some form of mechanical scanning, rather than obtaining a complete picture at one time as is commonly done with light [34], [35]. For many applications, however, mechanical scanning is too slow. For example, in plasma diagnostics, events of interest often occur on a time scale of microseconds. In astronomy and atmospheric studies where the signals are usually very weak and therefore require a long integration time, an extremely long time would be required to map out an image. For these applications, an array of detectors is needed rather than a single scanned detector. In light of recent advances in the development of new planar millimeter-wave detectors such as Schottky diodes [36], [37], SIS junctions [38], [39], and microbolometers [40], several researchers have pointed out that monolithic imaging array would be significant [36], [41]-[43].

The success of a monolithic imaging array depends on the development of suitable optics and individual antennas for the detectors. A hybrid array of dual-dipole antennas on quartz with beam-lead Schottky diode was reported by Parrish et al. [44]. Recently, an imaging antenna array with the "reverse-microscope" optical configuration and a modified bow-tie antenna design was reported by Rutledge et al. [45], [46]. This

system is particularly attractive because it offers nearly diffraction-limited imaging and is suitable for monolithic integration. The versatility of this imaging scheme has been demonstrated in a variety of applications, such as a far-infrared imaging antenna arrays [46], [47], tracking antenna arrays [48], and a polarization-sensitive imaging array [49]. Microbolometers were used as detectors in the above applications.

Because of recent advances in GaAs IC fabrication technology, planar Schottky diodes with cutoff frequencies higher than millimeter-wave frequencies have been fabricated by several laboratories [37]. Noise-equivalent powers (NEP) of the order of 10^{-11} W/Hz^{1/2} for video detection and of 10^{-20} W/Hz for heterodyne detection have been obtained. Although an improved microbolometer [50] compares very favorably with Schottky diodes as a video detector at millimeter-wave frequencies, its potential for use as a heterodyne detector is limited, because the responsivity of the bolometer starts to roll off at modulation frequencies around 100 KHz. The construction of a monolithic GaAs Schottky diode imaging array is, therefore, significant. In chapter four of this thesis, theory pertinent to the monolithic Schottky diode imaging array is discussed. The fabrication of the diode array and the experimental results of imaging array are described in chapter five.

References

- [1] J. C. Maxwell, A Treatise on Electricity and Magnetism, 2 vol., Oxford, 1873.
- [2] H. R. Hertz, Sitzb. Berl. Akad. Wiss., Feb. 2, 1888; Wiedem. Ann., vol. 34, p. 551, 1888; English translation in his Electric Waves, p. 107, MacMillan, London, 1893.
- [3] J. C. Wiltse, "Introduction and overview of millimeter waves." in Infrared and Millimeter Waves, vol. 4, K. G. Button ed., pp. 1-21, Academic Press, New York, 1981.
- [4] A. Chu, W. E. Courtney, and L. J. Mahoney, "Monolithic circuits for millimeter-wave systems." Microwave Journal, vol. 26, pp. 28-48, 1983.
- [5] R. A. Pucel, "Design consideration for monolithic microwave circuits," IEEE Microwave Theory Tech., vol. MTT-29, pp. 513-534, 1981.
- [6] D. W. Maki, J. M. Schellenberg, H. Yamasaki, and L. C. T. Lin, "A 69 GHz monolithic FET oscillator,"; L. T. Yuan, "A W-band monolithic GaAs crossbar mixer,"; P. Bauhahn, T. Contolatis, J. Abrokwhah, C. Chao, and C. Seashore, "94 GHz planar GaAs monolithic balanced mixer," in IEEE MTT-S International Microwave Symposium Digest, 1984.
- [7] C. Yao, S. E. Schwarz, and B. J. Blumenstock, "Monolithic integration of a dielectric millimeter-wave antenna and mixer diode: an embryonic millimetro-wave IC," IEEE Trans. Microwave Theory Tech., vol. MTT-30, pp. 1241-1247, 1982.
- [8] T. Wang and S. E. Schwarz, "Design of dielectric ridge waveguides for millimeter-wave integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-31, pp. 128-134, 1983.

- [9] Z. Rav-Roy, C. N. Zah, U. Shreter, D. B. Rutledge, T. Wang, S. E. Schwarz, and T. F. Juech, "Monolithic Schottky diode imaging arrays at 94 GHz," Eighth International Conference on Infrared and Millimeter-Waves, Miami, December 1983.
- [10] T. Wang, S. E. Schwarz, Z. Rav-Noy, C-E. Zah, U. Shreter, D. B. Rutledge, and T. F. Juech, "Monolithic mixer arrays for millimeter-wave imaging," to be published.
- [11] E. A. J. Marcatili, "Dielectric rectangular waveguide and directional coupler for integrated optics," Bell Syst. Tech. J., vol. 48, pp. 2071-2102, 1969.
- [12] J. E. Goell, "A circular-harmonic computer analysis of rectangular dielectric waveguides," Bell Syst. Tech. J., vol. 48, pp. 2133-2160, 1969.
- [13] R. M. Knox and P. P. Toullos, "Integrated circuits for millimeter through optical frequency range," in Proc. Symp. Submillimeter Waves, pp. 497-516, Polytechnic Press, Brooklyn, New York, 1970.
- [14] P. P. Toullos and R. M. Knox, "Image line integrated circuits for system applications at millimeter wavelengths," U.S. Army Electronics Command, Final Rep., Dept. no. ECOM-73-0217-F, 1974.
- [15] K. Solbach and I. Wolff, "The electromagnetic fields and the phase constants of dielectric image lines," IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 266-274, 1978.
- [16] T. Itoh and R. Mittra, "New waveguide structures for millimeter-wave integrated circuits," in Int. Microwave Symp. Dig., 1975, pp. 277-279.
- [17] W. V. McLevige, T. Itoh, and R. Mittra, "New waveguide structures for millimeter wave and optical integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-23, pp. 788-794, 1975.

- [18] T. Itoh, "Inverted strip dielectric waveguide for millimeter wave integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-24, pp. 821-827, 1976.
- [19] R. Mittra, Y-L. Hou, and V. Jamnejad, "Analysis of open dielectric waveguides using mode-matching technique and variational methods," IEEE Trans. Microwave Theory Tech., vol. MTT-28, pp. 36-43, 1980.
- [20] T. Itoh and B. Adelseck, "Trapped image guide for millimeter-wave circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-28, pp. 1433-1436, 1980.
- [21] J. E. Goell, "Rib waveguide for integrated optical circuits," Applied Optics, vol. 12, pp. 2797-2798, 1973.
- [22] B. J. Levin and J. E. Kietzer, "Hybrid millimeter-wave integrated circuits," U.S. Army Electronics Command, Final Report No. ECOM-74-0577-F, 1975.
- [23] T. Itanami and S. Shindo, "Channel dropping filter for millimeter-wave integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 759-764, 1978.
- [24] B-S. Song and T. Itoh, "Distributed Bragg reflection dielectric waveguide oscillators," IEEE Trans. Microwave Theory Tech., vol. MTT-27, pp. 1019-1022, 1979.
- [25] G. L. Matthaei, "A note concerning modes in dielectric waveguide gratings for filter applications," IEEE Trans. Microwave Theory Tech., vol. MTT-31, pp. 309-312, 1983.
- [26] G. L. Matthaei, D. C. Park, Y. M. Kim, and D. L. Johnson, "A study of the filter properties of single and parallel-coupled dielectric-waveguide gratings," IEEE Trans. Microwave Theory Tech., vol. MTT-31, pp. 825-835, 1983.

- [27] H. Jacobs and M. M. Chrepta, "Electronic phase shifter for millimeter-wave semiconductor dielectric integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-22, pp. 411-417, 1974.
- [28] K. L. Kohn, R. T. Horn, H. Jacobs, and E. Freibergs, "Silicon waveguide frequency scanning linear array antenna," IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 764-773, 1978.
- [29] R. E. Horn, H. Jacobs, E. Freibergs, and K. L. Kohn, "Electronic modulated beam-steerable silicon waveguide array antenna," IEEE Trans. Microwave Theory Tech., vol. MTT-28, pp. 647-653, 1980.
- [30] D. B. Rutledge, S. E. Schwarz, T-L. Hwang, D. J. Angelakos, K. K. Mei, and S. Yokota, "Antenna and waveguides for far-infrared integrated circuits," IEEE J. Quantum Electronics, vol. QE-16, pp. 508-516, 1980.
- [31] R. M. Knox, K. J. Brandt, J. E. Kietzer, and R. C. Beavin, "Microwave dielectric waveguide system having high data rate capability," Proc. of Government Microcircuits Applications Conference, Boulder, CO, June 1974.
- [32] J. A. Paul and Y-W. Chang, "Millimeter-wave image-guide integrated passive devices," IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 751-754, 1978.
- [33] J. A. Paul and P. C. H. Yen, "Millimeter-wave passive components and six-port network analyzer in dielectric waveguide," IEEE Trans. Microwave Theory Tech., vol. MTT-29, pp. 948-953, 1981.
- [34] D. T. Hodges, F. B. Foote, E. E. Reber, and R. L. Schellenbaum, "Near-millimeter wave radiometric imaging," in Conf. Digest, Fourth Int. Conf. Infrared Millimeter-wave and their Appl., pp. 51-52, 1979.

- [35] J. Waldman, H. R. Fetterman, P. E. Duffy, T. G. Bryant, and P. E. Tannenwald, "Submillimeter model measurements and their applications to millimeter radar systems," Conf. Digest, Fourth Int. Conf. Infrared Millimeter Waves and their Appl., pp. 49-50, 1979.
- [36] B. J. Clifton, R. A. Murphy, and G. D. Alley, "Integrated monolithic mixers on GaAs for millimeter and submillimeter wave applications," Conf. Digest, Fourth Int. Conf. Infrared Millimeter Waves and their Appl. pp. 84-86, 1979.
- [37] B. J. Clifton, G. D. Alley, R. A. Murphy, and I. H. Mroczkowski, "High-performance quasi-optical GaAs monolithic mixer at 110 GHz," IEEE Trans. Electron Dev., vol. ED-28, pp. 155-157, 1981.
- [38] M. W. McGrath, P. L. Richards, A. D. Smith, H. van Kemper, R. A. Batchelor, D. E. Prober, and P. Santhanam, "Large gain, negative resistance, and oscillations in superconducting quasi-particle heterodyne mixers," Appl. Phys. Lett., vol. 39, pp. 655-658, 1981.
- [39] G. J. Dolan, T. G. Phillips, and D. P. Woody, "Low noise, 115 GHz mixing in superconducting oxide barrier junctions," Appl. Phys. Lett., vol. 34, pp. 347-349, 1979.
- [40] T. L. Hwang, S. E. Schwarz, and D. B. Rutledge, "Micro-bolometers for infrared detection," Appl. Phys. Lett., vol. 34, pp. 773-776, 1979.
- [41] N. C. Luhmann, Jr., "Instrumentations and techniques for plasma diagnostics: an overview," in Infrared and Millimeter-Waves, vol. 2, K. J. Button, Ed., Academic Press, New York, 1979.
- [42] G. A. Gordon, R. L. Hartmann, and P. W. Kruse, "Imaging-mode operation of active NMMW systems," in Infrared and Millimeter-Waves,

- vol. 4, K. J. Button Ed., Academic Press, New York, 1981.
- [43] J. M. Schuchardt, J. M. Newton, T. P. Morton, and J. A. Gagliano, "The coming of MM-wave forward looking imaging radiometers," *Microwave J.*, pp. 45-62, 1981.
- [44] P. T. Parrish, T. C. L. G. Sollner, R. H. Mathews, H. R. Fetterman, C. D. Parker, P. E. Tannenwald, and A. G. Cardasmenos, "Millimeter wave technology," *Proc. SPIE*, vol. 337, pp. 49-52, 1982.
- [45] D. B. Rutledge and M. S. Muha, "Imaging antenna arrays," *IEEE Trans. Antennas Prop.*, vol. AP-30, pp. 535-540, 1982.
- [46] D. P. Neikirk, D. B. Rutledge, and M. S. Muha, "Far-infrared imaging antenna arrays," *Appl. Phys. Lett.*, vol. 40, pp. 203-205, 1982.
- [47] D. P. Neikirk, P. P. Tong, and D. B. Rutledge, "Imaging antenna array at 119 μm ," *Appl. Phys. Lett.*, vol. 41, pp. 329-331, 1982.
- [48] P. P. Tong, D. P. Neikirk, D. Psaltis, D. B. Rutledge, K. Wagner, and P. E. Young, "Tracking antenna arrays for near-millimeter waves," *IEEE Trans. Antenna Prop.*, vol. AP-31, pp. 512-515, 1983.
- [49] P. P. Tong, D. B. Rutledge, and D. P. Neikirk, "Polarization-sensitive imaging arrays," *Symp. Digest, 1984 IEEE MTT-S Int. Microwave Symposium*, pp. 542-544, 1984.
- [50] D. P. Neikirk and D. B. Rutledge, "Air-bridge microbolometer for far-infrared detection," *Appl. Phys. Lett.*, vol. 44, pp. 153-155, 1984.

Figure Captions

- 1.1. Monolithic integration of a dielectric millimeter-wave antenna and mixer diode (after Yao, et al. [7]).
- 1.2. The fabrication steps of a section of a hypothetical M^3IC .
 - (i) a semiconductor device, such as diode is fabricated on the wafer first.
 - (ii) The V-coupler is then fabricated by evaporation.
 - (iii) The device side of the wafer is protected, and the waveguide is defined by photoresist. After anisotropic etching and stripping of the photoresist, the circuit is finished, as shown in (iv).

CHAPTER 2

THEORY OF DIELECTRIC RIDGE WAVEGUIDES

2.1. Introduction

The literature relating to dielectric waveguides is extensive. Most of it is concerned with numerical calculations of particular characteristics (e.g., propagation constant, radiation loss, etc.) for particular waveguide structures. Numerical calculations are certainly necessary in order to obtain detailed information; however, in the design of dielectric waveguide, as in any design problem, tradeoffs of different characteristics have to be made. A clear physical picture of what is happening is absolutely essential if complex interrelations between different characteristics are to be understood and intelligent tradeoffs are to be chosen. Therefore, for design of our particular dielectric ridge waveguide we used the analytical approximations as much as possible. Subsequent to the overall design of the waveguide, anyone interested in fine-tuning the design can then use numerical methods can be used to obtain the detailed information.

In this chapter, we seek to develop a solid understanding of the dielectric ridge waveguide to pave the way for the design to be taken up in the next chapter. The unifying theme of the chapter is the mode-matching technique and its lowest order approximation.

As we shall see in the following, the dielectric ridge waveguide can be seen as formed by a dielectric slab waveguide of thickness b sandwiched between dielectric slab guides of thickness t . The guided modes can be thought as multiple reflection of the slab modes taking place at the two side walls of the ridge guide. We therefore treat the dielectric slab waveguide briefly first, in order to define various quantities and

functions of interest and to bring out the important properties of the slab modes. The reflection of a slab mode by a step discontinuity is discussed next. The mode-matching technique, using the slab modes, is used to formulate the problem. Our interest here, however, is not to carry out the exact solution. In our analytical approach we are more interested in the lowest-order approximation; its nature and limitations. Equipped with an understanding of the reflection problem of the step discontinuity, we then discuss the properties of the dielectric ridge waveguide itself. Finally, the radiation loss from waveguide bend is treated. The mode-matching technique and its lowest-order approximation is synthesized with Marcattis method to obtain an approximate solution for the case of the dielectric ridge waveguide.

2.2. Dielectric Slab Waveguides

The properties of the modes of dielectric slab waveguide are now well documented (e.g., [1], [2]). The purpose of this section is to define and list the field components and variables of the mode function and equations that we shall need in the later part of the chapter.

The dielectric slab waveguide can support three type of modes: (1) guided modes, (2) radiation modes, and (3) evanescent modes. Guided modes and radiation modes are propagating modes, while the evanescent modes are attenuating away from their source. A brief review of guided modes will be presented first. We shall then discuss qualitatively the radiation and the evanescent modes. The orthogonality relations among the modes will be presented last.

2.2.1. Guided Modes

A dielectric slab guide of index of refraction n , thickness d , surrounded by a medium of index of refraction n_0 , sandwiched between two

perfectly-conducting planes is shown in Fig. 2.1. The slab guide supports transverse electric (TE) modes with zero longitudinal electric field ($E_z = 0$) and transverse magnetic (TM) modes with zero longitudinal magnetic field ($H_z = 0$). Assume the mode functions traveling in z-direction depend on z and time t as $e^{-j\beta z} e^{j\omega t}$, where β is the propagation constant and ω is the angular frequency. Because there is no confinement in the x-direction, the electric field $\underline{E}$ satisfies the wave equation

$$\left[\frac{\partial^2}{\partial y^2} + (\epsilon_r k_0^2 - \beta^2) \right] \underline{E} = 0 \quad (2.1)$$

where $\epsilon_r = n_0^2$ for $|y| \geq \frac{d}{2}$, $\epsilon_r = n^2$ for $|y| \leq \frac{d}{2}$, and k_0 is the free space propagation constant.

$$\text{Let } \underline{E} = \underline{E} e^{-j\beta z} e^{j\omega t} \text{ and } \underline{H} = \underline{H} e^{-j\beta z} e^{j\omega t}.$$

A TE mode can be specified by the x-component of its $\underline{E}$ field. The field components are

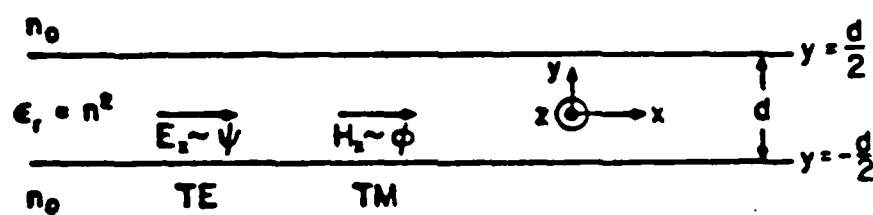
$$\begin{aligned} E_x &= \psi(y) & H_x &= 0 \\ E_y &= 0 & H_y &= \frac{\beta}{\omega \mu_0} \psi(y) \\ E_z &= 0 & H_z &= \frac{-1}{j\omega \mu_0} \frac{\partial \psi(y)}{\partial y} \end{aligned} \quad (2.2)$$

where μ_0 is the free space magnetic permeability. For even TE modes, $\psi(y)$ can be written as (assuming, for the moment, the perfect-conducting planes are at infinity)

$$\psi(y) = \begin{cases} \cos \frac{k_y y}{2} e^{-\alpha(y - \frac{d}{2})} & y > \frac{d}{2} \\ \cos k_y y & |y| \leq \frac{d}{2} \\ \cos \frac{k_y d}{2} e^{\alpha(y + \frac{d}{2})} & y < -\frac{d}{2} \end{cases} \quad (2.3)$$

Perfectly-Conducting plane

//////////////////////////////////// $y = h_1$



//////////////////////////////////// $y = -h_2$
 Perfectly-Conducting plane

Fig. 2.1

where $k_y = [n^2 k_0^2 - \beta^2]^{1/2}$, and $\alpha = [\beta^2 - n_0^2 k_0^2]^{1/2}$. By matching boundary conditions, we have the characteristic equation for the even TE modes

$$k_y d = 2m\pi + 2 \tan^{-1}\left(\frac{\alpha}{k_y}\right) \quad (2.4)$$

where $m = 0, 1, 2, \dots$. For odd TE modes,

$$\psi(y) = \begin{cases} \sin \frac{k_y d}{2} e^{-\alpha(y - \frac{d}{2})} & y > \frac{d}{2} \\ \sin k_y y & |y| \leq \frac{d}{2} \\ -\sin \frac{k_y d}{2} e^{\alpha(y + \frac{d}{2})} & y < -\frac{d}{2} \end{cases} \quad (2.5)$$

The corresponding characteristic equation is

$$k_y d = 2m\pi - 2 \tan^{-1}\left(\frac{k_y}{\alpha}\right) \quad (2.6)$$

where $m = 1, 2, 3, \dots$. Equations (2.4) and (2.6) can be combined, by $\tan^{-1}(x) + \tan^{-1}(1/x) = \frac{\pi}{2}$, to obtain

$$k_y d = m\pi - 2 \tan^{-1}\left[\frac{k_y}{\alpha}\right] \quad (2.7)$$

where $m = 1, 2, 3, \dots$, corresponding to the order of the mode (TE_m mode), with odd values of m for even modes and vice versa.

The power per unit width P carried by a slab TE mode can be calculated by

$$\begin{aligned} P &= \frac{1}{2} \int_{-\infty}^{\infty} (\underline{E} \times \underline{H}^*)_z \, dy \\ &= \frac{1}{2} \frac{\beta}{\omega \mu_0} \langle \psi | \psi \rangle = \frac{\beta(d+2/\alpha)}{4\omega \mu_0} \end{aligned} \quad (2.8)$$

where $\langle g|f \rangle = \int_{-\infty}^{\infty} f g^* dy$. For later reference, we note

$$\langle \psi|\psi \rangle = \frac{1}{2} (d+2/\alpha) \quad (2.9)$$

Similarly, TM modes can be specified by their H_x . The field components are

$$\begin{aligned} E_x &= 0 & H_x &= \phi(y) \\ E_y &= \frac{-\beta}{\omega \epsilon_r \epsilon_0} \phi(y) & H_y &= 0 \\ E_z &= \frac{-1}{j\omega \epsilon_r \epsilon_0} \frac{\partial \phi(y)}{\partial y} & H_z &= 0 \end{aligned} \quad (2.10)$$

The mode functions ϕ of even and odd TM modes can be written exactly the same as (2.3) and (2.5), respectively. The characteristic equation TM_m mode is

$$k_y d = m\pi - \tan^{-1} \left[\left(\frac{n_0}{n} \right)^2 \frac{k_y}{\alpha} \right] \quad (2.11)$$

where $m = 1, 2, 3, \dots$ and odd values of m are for even modes, and vice versa. Figure 2.2 shows the normalized propagation constant ($\beta/k_0 \equiv$ effective index of refraction, see below) for TE_1 , TE_2 , TM_1 , and TM_2 modes as a function of normalized frequency (d/λ_0) for $\epsilon_r = 12$. Notice that, because of high ϵ_r , TM modes are much more loosely confined than the TE modes.

The power per unit width carried by a TM mode is

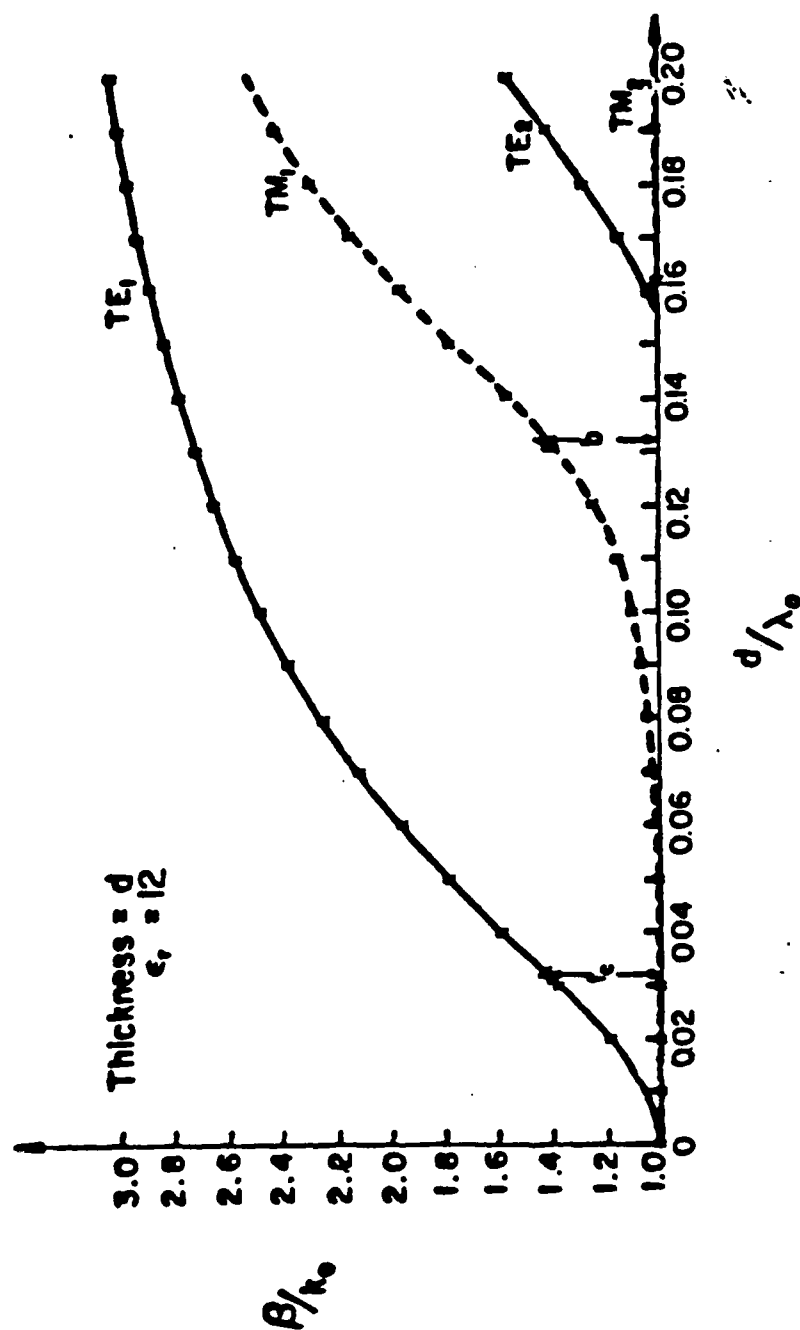


Fig. 2.2

$$\begin{aligned}
 P &= \frac{\beta}{2\omega\epsilon_0} \langle \phi | \frac{1}{\epsilon_r} | \phi \rangle \\
 &= \frac{\beta}{4\omega\epsilon_0 n^2} \left[d + 2 \frac{n_0^2 n^2}{\alpha} \frac{k_y^2 + \alpha^2}{n_0^2 k_y^2 + n^4 \alpha^2} \right]
 \end{aligned} \tag{2.12}$$

and

$$\langle \phi | \frac{1}{\epsilon_r} | \phi \rangle = \frac{1}{2n^2} \left[d + 2 \frac{n_0^2 n^2}{\alpha} \frac{k_y^2 + \alpha^2}{n_0^2 k_y^2 + n^4 \alpha^2} \right] \tag{2.13}$$

Impedance is an important concept for transmission lines and plane waves. Although it does not have the same meaning as in plane waves and transmission lines, we can still define the impedance for the slab modes. The impedance of a TE mode propagating in the z-direction is defined as

$$Z_z' = \frac{E_x}{H_y} = \frac{\omega\mu_0}{\beta} \tag{2.14}$$

by (2.2). It is a characteristic of the slab mode and does not depend on the transverse coordinate. Similarly, the impedance for z-direction propagation for a TM mode is

$$Z_z'' = \frac{E_y}{-H_x} = \frac{\beta}{\omega\epsilon_r\epsilon_0} \tag{2.15}$$

In contrast to that for TE modes, Z_z'' defined in this manner is a function of position because of $\epsilon_r(y)$.

Another quantity of interest is the effective index of refraction. For the ith TE or TM mode, the effective index of refraction is

$$n_i = \frac{\beta_i}{k_0} \tag{2.16}$$

It is related to the effective dielectric constant of the mode by [3]

$$\epsilon_i = n_i^2 = \left(\frac{\beta_i}{k_0} \right)^2 \tag{2.17}$$

2.2.2. Radiation and Evanescent Modes

The radiation modes can be visualized to be generated by infinite plane wave sources placed symmetrically at $y = \pm\infty$. The wave is incident on the slab with an angle ν with respect to the y -direction. The propagation constant in the y -direction, K , of the incident wave in the surrounding medium is related to ν and k_0 by $K = k_0 \cos \nu$. Similarly, the propagation constant of the mode β is given by $\beta = k_0 \sin \nu$. The even TE radiation modes are generated from sources whose $\underline{E}$ fields are polarized in x -direction and are oscillating in phase. The odd TE radiation modes are generated from the same sources oscillating out of phase with each other. The TM radiation modes are obtained when the sources have the $\underline{H}$ field polarized in x -direction.

According to this picture, the mode function of an even TE mode can be written as

$$E_x = \psi(y) = \begin{cases} \frac{\cos(\frac{k_y d}{2})}{\cos(\frac{Kd}{2} + \theta_e)} \cos(Ky + \theta_e) & |y| > \frac{d}{2} \\ \cos(k_y y) & |y| \leq \frac{d}{2} \end{cases} \quad (2.18)$$

where $k_y = [n^2 k_0^2 - \beta^2]^{1/2}$ and $K = [n_0^2 k_0^2 - \beta^2]^{1/2}$. The angle θ_e is determined by matching boundary conditions

$$k_y \tan \frac{k_y d}{2} = K \tan(\frac{Kd}{2} + \theta_e) \quad (2.19)$$

Because of the adjustable parameter θ_e , the spectra of k_y and K are continuous. The other field components can be obtained by (2.2). For odd TE radiation modes, $\psi(y)$ can be written as

$$E_x = \psi(y) = \begin{cases} \frac{\sin(\frac{k_y d}{2})}{\sin(\frac{d}{2} + \theta_0)} \frac{y}{|y|} \sin(K|y| + \theta_0) & |y| > \frac{d}{2} \\ \sin k_y y & |y| \leq \frac{d}{2} \end{cases} \quad (2.20)$$

And θ_0 is determined by the equation

$$k_y \cot(\frac{k_y d}{2}) = K \cot(\frac{Kd}{2} + \theta_0) \quad (2.21)$$

The fact that the field distribution outside the slab is a standing wave, containing both incoming and outgoing waves, may seem troublesome; how can radiation generated by sources inside the waveguide (e.g., waveguide imperfections) contain incoming traveling waves? It can be shown, however, that if the radiation is excited by internal sources it always excites infinitely many radiation modes, which superimpose in such a way that the incoming parts of the standing waves are eliminated by destructive interference [1].

Because $k_y = [n^2 k_0^2 - \beta^2]^{1/2}$ and $K = [n_0^2 k_0^2 - \beta^2]^{1/2}$, the range of k_y and K are restricted if K is real. Obviously, the field distribution caused by fine waveguide imperfections cannot be represented by the radiation and the guided modes alone. The evanescent modes are the modes with imaginary propagation constant $i|\beta|$. In terms of our picture of radiation modes, this corresponds to an imaginary incident angle $i|\nu|$. The mode functions are still given by (2.18)-(2.21). The evanescent modes are, therefore an extension of the radiation modes.

2.2.3. Orthogonality Relations

As we have seen, the spectra of radiation and evanescent modes are

continuous. In order to avoid handling the continuous spectrum, it is customary to discretize the spectrum by placing perfectly-conducting walls above and below the slab guide, Fig. 2.1, [4]. It is well known that a partially-filled parallel plate waveguide supports a finite number of surface waves (the guided modes), a finite number of discrete higher modes, some of which are propagating (the radiation modes) and the remainder are below cutoff (the evanescent modes) [5]. It is assumed that the perfectly-conducting planes are far away from the slab so as to negligibly influence the properties of the surface modes.

The mode functions of a partially-filled parallel-plate waveguide are governed by the Sturm-Liouville eigenvalue problem [6]. Equation (2.1) can be rearranged to give

$$\left[\frac{d}{dy} p(y) \frac{d}{dy} + q(y) \right] \begin{Bmatrix} \psi_n(y) \\ \phi_n(y) \end{Bmatrix} = \beta_n^2 w(y) \begin{Bmatrix} \psi_n(y) \\ \phi_n(y) \end{Bmatrix} \quad (2.22)$$

subject to the boundary conditions

$$\psi_n(-h_2) = \psi_n(h_1) = 0 \quad \text{for TE modes} \quad (2.23)$$

$$\left. \frac{\partial \phi_n}{\partial y} \right|_{-h_2} = \left. \frac{\partial \phi_n}{\partial y} \right|_{h_1} = 0 \quad \text{for TM modes}$$

and $p(y)$, $q(y)$ and $w(y)$ are known functions defined by

$$p(y) = w(y) = \begin{cases} 1 & \text{TE modes} \\ 1/\epsilon_r(y) & \text{TM modes} \end{cases} \quad (2.24)$$

$$q(y) = \begin{cases} k_0^2 \epsilon_r(y) & \text{TE modes} \\ k_0^2 & \text{TM modes} \end{cases} \quad (2.25)$$

The Sturm-Liouville eigenvalue problem defined by (2.22)-(2.25) is Hermitian, because of the perfectly-conducting bounding plates. Therefore, all the eigenvalues β_n^2 are real (both positive and negative) and all eigenfunctions can be chosen to be real. Furthermore, the mode functions of the same type (TE or TM) are mutually orthogonal. With proper normalization, they can be chosen to satisfy the orthogonality relation

$$\langle \psi_m | w(y) | \psi_n \rangle = \langle \psi_m | \psi_n \rangle = \delta_{mn} \quad \text{TE modes} \quad (2.26)$$

and

$$\langle \phi_m | w(y) | \phi_n \rangle = \langle \phi_m | \frac{1}{\epsilon_r(y)} | \phi_n \rangle = \delta_{mn} \quad \text{TM modes} \quad (2.27)$$

for every m and n . A relationship between the mode functions of different types may be obtained, by manipulating (2.22)-(2.25) for both TE and TM modes, as

$$\beta_m^2 \langle \psi_m | \frac{1}{\epsilon_r(y)} | \frac{\partial \phi_n}{\partial y} \rangle + (\beta'_n)^2 \langle \phi_n | \frac{1}{\epsilon_r(y)} | \frac{\partial \psi_m}{\partial y} \rangle = 0 \quad (2.28)$$

where β and β' are the propagation constants for TE and TM modes, respectively. This applies to all three classes of modes. This equation is required to ensure that the TE and TM modes are mutually orthogonal in power, even though the mode functions of one set may not be orthogonal to those of the other set [6].

2.3. Reflection from a Step Discontinuity in a Slab Waveguide

A slab waveguide step discontinuity at the plane of $x = 0$ is shown in Fig. 2.3. Consider the scattering problem depicted schematically in Fig. 2.4. A TE_1 mode for the slab guide with thickness b is incident on the step discontinuity at an angle θ'_1 to the z -direction. In the

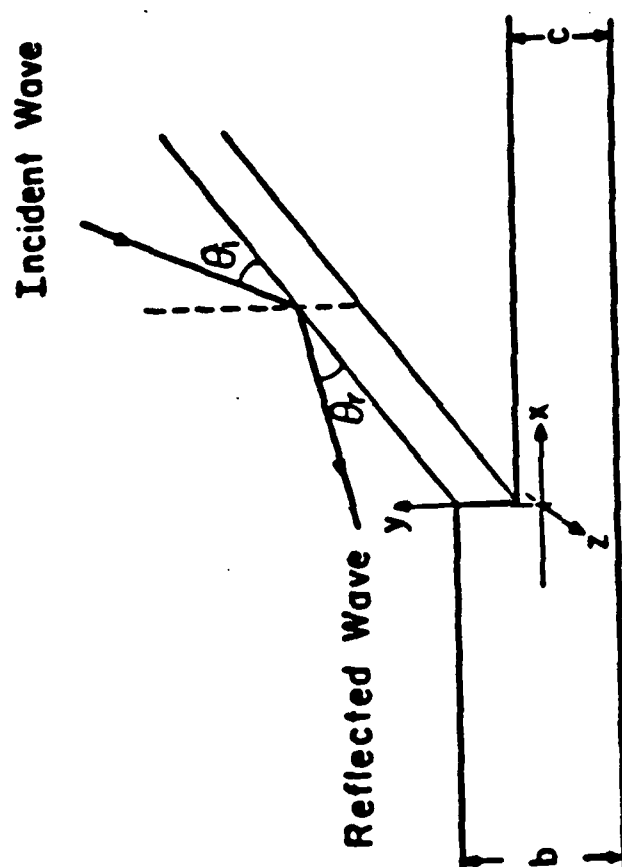


Fig. 2.3

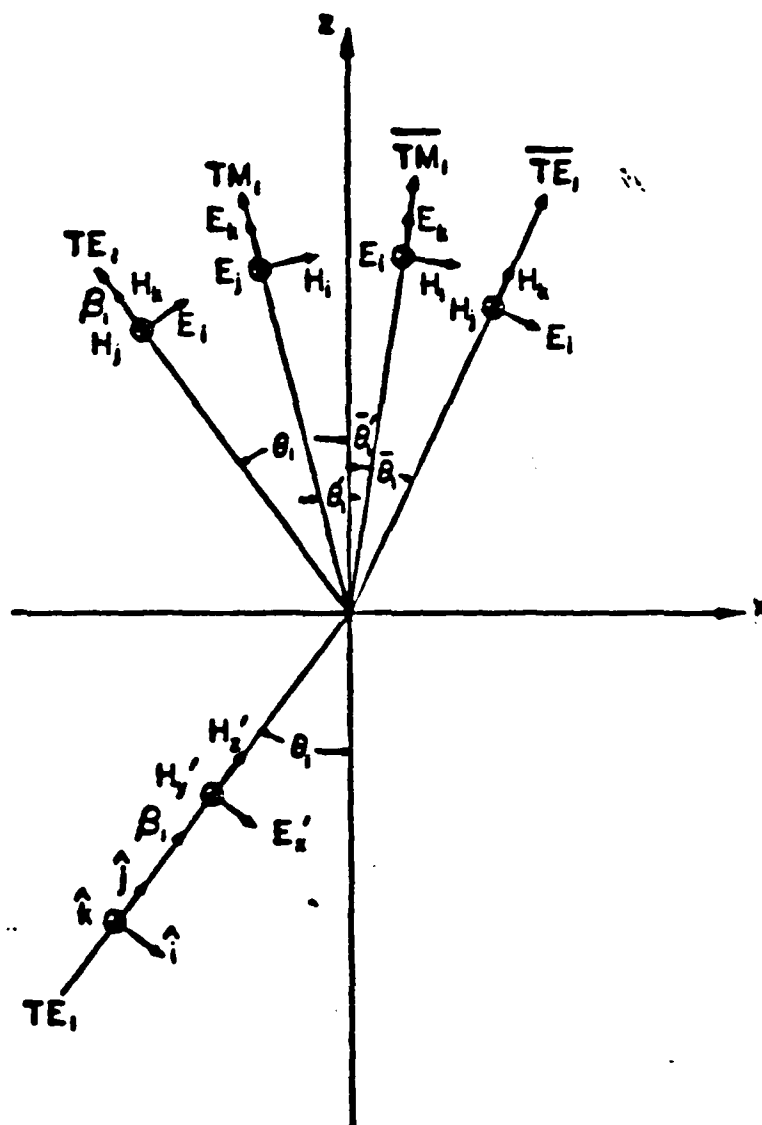


Fig. 4

first part of this section, an formulation of the problem based upon an exact mode-matching technique [5] is presented. In order to gain insight into the nature of the problem, the lowest order approximations are examined in the second part of the section. The connection between the lowest order approximation and the effective dielectric constant method is pointed out also.

The notation used in the rest of the chapter requires some explanation. We use a single prime to denote quantities for TE modes and a double prime for TM modes. For modes on either side of the discontinuity, we use the bar to denote the quantities in the region of slab thickness c ($x \geq 0$) and the unbarred quantities for the region of slab thickness b ($x \leq 0$). For example, β'_i is the propagation constant of the TE_i mode in the region for $x \leq 0$ and $\bar{\beta}''_i$ is the propagation constant for the TM_i mode in the region for $x \geq 0$.

2.3.1. Exact Formulation

The guided modes, the radiation modes and the evanescent modes of TE and TM modes together from a complete set of functions for the fields of the slab waveguide, thus the scattering problem can be solved by expressing the fields in each of the two region in terms of their respective complete mode set.

The field components for the incident wave, expressed in terms of its intrinsic coordinate $\hat{i}$, $\hat{j}$, and $\hat{k}$ (Fig. 2.4) are given by (2.2) with $\hat{x} \rightarrow \hat{i}$, $\hat{y} \rightarrow \hat{j}$, and $\hat{z} \rightarrow \hat{k}$. In terms of the structure coordinate system (x, y, z) , the fields have five components.

$$\begin{aligned}
 E_x &= E_i \cos \theta'_i & H_x &= H_k \sin \theta'_i \\
 E_y &= E_j = 0 & H_y &= H_j \\
 E_z &= -E_i \sin \theta'_i & H_z &= H_k \cos \theta'_i
 \end{aligned}
 \tag{2.29}$$

The propagation constants in the x- and z-direction are

$$k'_{x_1} = \beta'_1 \sin \theta'_1 \quad \text{and} \quad k'_{z_1} = \beta'_1 \cos \theta'_1 \quad (2.30)$$

and

$$(k'_{x_1})^2 + (k'_{z_1})^2 = (\beta'_1)^2 \quad (2.31)$$

The reflected wave is a linear superposition of the slab TE and TM mode. The field components of the reflected TE_m mode are

$$\begin{aligned} E_x &= E_i \cos \theta'_m & H_x &= -H_k \sin \theta'_m \\ E_y &= 0 & H_y &= H_j \\ E_z &= E_i \sin \theta'_m & H_z &= H_k \cos \theta'_m \end{aligned} \quad (2.32)$$

Also

$$k'_{x_m} = \beta'_m \sin \theta'_m \quad k'_{z_1} = \beta'_m \cos \theta'_m \quad (2.33)$$

and

$$(k'_{x_m})^2 + (k'_{z_m})^2 = (\beta'_m)^2 \quad (2.34)$$

Similarly, for the reflected TM_n modes, we have

$$\begin{aligned} E_x &= -E_k \sin \theta''_n & H_x &= H_i \cos \theta''_n \\ E_y &= E_j & H_y &= 0 \\ E_z &= E_k \cos \theta''_n & H_z &= H_i \sin \theta''_n \end{aligned} \quad (2.35)$$

The propagation constants in the x- and z-direction are given similarly,

$$k''_{x_n} = \beta''_n \sin \theta''_n \quad k''_{z_n} = \beta''_n \cos \theta''_n \quad (2.36)$$

and

$$(k''_{x_n})^2 + (k''_{z_n})^2 = (\beta''_n)^2 \quad (2.37)$$

In order to match boundary condition for arbitrary z , all the reflected modes have the same propagation constant in z -direction,

$$k_z = k'_{z_m} = k''_{z_n} = \beta'_i \cos \theta'_i \quad (2.38)$$

It is obvious that if θ'_i is small enough, the angles θ'_m and θ''_n will have to become imaginary in order to satisfy (2.38). In that case, the sign of the k_x 's have to be chosen so that the reflected wave attenuates away from the discontinuity. It is convenient to express various sines and cosines in terms of the slab mode propagation constants β 's and k_z ,

$$\cos \theta_i = \frac{k_z}{\beta_i} \quad \sin \theta_i = \frac{[\beta_i^2 - k_z^2]^{1/2}}{\beta_i} \quad (2.39)$$

where the unprimed quantities stand for both TE and TM modes. When the angle θ_i become imaginary,

$$\sin \theta_i = \frac{[\beta_i^2 - k_z^2]^{1/2}}{\beta_i} = \frac{j[k_z^2 - \beta_i^2]^{1/2}}{\beta_i} \quad x \leq 0 \quad (2.40)$$

The transmitted waves can be defined similarly. If the transmission angle $\bar{\theta}_i$ becomes imaginary, (2.40) become

$$\sin \bar{\theta}_i = \frac{-j[k_z^2 - (\bar{\beta}_i)^2]^{1/2}}{\bar{\beta}_i} \quad x \geq 0 \quad (2.41)$$

Table 2.1 lists the field components for the incident, the reflected, and the transmitted waves.

We are now in a position to match the boundary condition at the step discontinuity. Let

	Reflected		Transmitted	
	TE _m	TM _n	TE _m	TM _n
E _x	$\frac{k_z}{\beta_m'} E_1$	$-\frac{[(\beta_n'')^2 - k_z^2]^{1/2}}{\beta_n''} E_k$	$\frac{k_z}{\beta_n} E_1$	$\frac{[(\beta_n'')^2 - k_z^2]^{1/2}}{\beta_n''} E_k$
E _y	0	E _j	0	E _j
E _z	$\frac{[(\beta_m')^2 - k_z^2]^{1/2}}{\beta_m'} E_1$	$\frac{k_z}{\beta_n''} E_k$	$-\frac{[(\beta_m')^2 - k_z^2]^{1/2}}{\beta_m'} E_1$	$\frac{k_z}{\beta_n''} E_k$
H _x	$\frac{[(\beta_m')^2 - k_z^2]^{1/2}}{\beta_m'} H_k$	$-\frac{[(\beta_m')^2 - k_z^2]^{1/2}}{\beta_m'} H_k$	$\frac{[(\beta_m')^2 - k_z^2]^{1/2}}{\beta_m'} H_k$	$\frac{k_z}{\beta_n''} H_1$
H _y	H _j	0	H _j	0

Table 2.1

r_1 = reflection coefficient of TE_1 mode

s_j = reflection coefficient of TM_j mode

u_ℓ = transmission coefficient of TE_ℓ mode

v_m = transmission coefficient of TM_m mode

Here, as in the waveguide application, we are primarily interested in the case for which all the reflected and transmitted angles are imaginary except for the reflected TE_1 mode. By matching the boundary condition for E_y , we have

$$-\sum_{j=1}^{\infty} s_j \frac{\beta_j''}{\omega \epsilon_r \epsilon_0} \phi_j = -\sum_{m=1}^{\infty} v_m \frac{\bar{\beta}_m''}{\omega \epsilon_r \epsilon_0} \bar{\phi}_m \quad (2.42)$$

Continuity of E_z gives

$$\begin{aligned} & -\frac{[(\beta_1')^2 - k_z^2]^{1/2}}{\beta_1'} \psi_1 + \sum_{i=1}^{\infty} r_i \frac{j[k_z^2 - (\beta_i')^2]^{1/2}}{\beta_i'} \psi_i - \sum_{j=1}^{\infty} s_j \frac{1}{j\omega \epsilon_r \epsilon_0} \frac{k_z}{\beta_j''} \frac{\partial \phi_j}{\partial y} \\ & = -\sum_{\ell=1}^{\infty} u_\ell \frac{-j[k_z^2 - (\beta_\ell')^2]^{1/2}}{\beta_\ell'} \bar{\psi}_\ell - \sum_{m=1}^{\infty} v_m \frac{1}{j\omega \epsilon_r \epsilon_0} \frac{k_z}{\beta_m''} \frac{\partial \bar{\phi}_m}{\partial y} \end{aligned} \quad (2.43)$$

Continuity of H_y requires

$$\frac{\beta_1'}{\omega \mu_0} \psi_1 + \sum_{i=1}^{\infty} r_i \frac{\beta_i'}{\omega \mu_0} \psi_i = \sum_{\ell=1}^{\infty} u_\ell \frac{\bar{\beta}_\ell'}{\omega \mu_0} \bar{\psi}_\ell \quad (2.44)$$

Finally, the boundary condition for H_z is

$$\begin{aligned} & \frac{1}{j\omega \mu_0} \frac{k_z}{\beta_1'} \frac{\partial \psi_1}{\partial y} + \sum_{i=1}^{\infty} r_i \frac{1}{j\omega \mu_0} \frac{k_z}{\beta_i'} \frac{\partial \psi_i}{\partial y} + \sum_{j=1}^{\infty} s_j \frac{j[k_z^2 - (\beta_j'')^2]^{1/2}}{\beta_j''} \phi_j \\ & = \sum_{\ell=1}^{\infty} u_\ell \frac{1}{j\omega \mu_0} \frac{k_z}{\beta_\ell'} \frac{\partial \bar{\psi}_\ell}{\partial y} - \sum_{m=1}^{\infty} v_m \frac{-j[k_z^2 - (\beta_m'')^2]^{1/2}}{\beta_m''} \bar{\phi}_m \end{aligned} \quad (2.45)$$

Equations (2.42)-(2.45) can be simplified by using the orthogonality relations (2.26)-(2.32). Multiplying (2.42) by ϕ_j^1 and integrating, we have, assuming the eigenfunctions are properly normalized.

$$s_j \beta_j^u = \sum_{m=1}^{\infty} v_m \overline{\beta_m^u} \langle \phi_j | \frac{1}{\epsilon_r} | \overline{\phi_m} \rangle \quad (2.46)$$

Similarly, $\langle \psi_i | (2.43) \rangle$ gives

$$\begin{aligned} & - \frac{[(\beta_1^1)^2 - k_z^2]^{1/2}}{\beta_1^1} \delta_{1i} + r_i \frac{j[k_z^2 - (\beta_1^1)^2]^{1/2}}{\beta_1^1} - \sum_{j=1}^{\infty} s_j \frac{1}{j\omega\epsilon_0} \frac{k_z}{\beta_j^u} \langle \psi_i | \frac{1}{\epsilon_r} | \frac{\partial \phi_j}{\partial y} \rangle \\ & = - \sum_{l=1}^{\infty} u_l \frac{-j[k_z^2 - (\beta_l^1)^2]^{1/2}}{\beta_l^1} \langle \psi_i | \overline{\psi_l} \rangle - \sum_{m=1}^{\infty} v_m \frac{1}{j\omega\epsilon_0} \frac{k_z}{\beta_m^u} \langle \psi_i | \frac{1}{\epsilon_r} | \frac{\partial \phi_m}{\partial y} \rangle \quad (2.47) \end{aligned}$$

$\langle \psi_i | (2.44) \rangle$ gives

$$\beta_1^1 \delta_{1i} + r_i \beta_1^1 = \sum_{l=1}^{\infty} u_l \overline{\beta_l^1} \langle \psi_i | \overline{\psi_l} \rangle \quad (2.48)$$

and $\langle \phi_j | \frac{1}{\epsilon_r} | (2.45) \rangle$ gives

$$\begin{aligned} & \frac{1}{j\omega\mu_0} \frac{k_z}{\beta_1^1} \langle \phi_j | \frac{1}{\epsilon_r} | \frac{\partial \psi_1}{\partial y} \rangle + \sum_{l=1}^{\infty} r_l \frac{1}{j\omega\mu_0} \frac{k_z}{\beta_l^1} \langle \phi_j | \frac{1}{\epsilon_r} | \frac{\partial \psi_l}{\partial y} \rangle + s_j \frac{j[k_z^2 - (\beta_j^1)^2]^{1/2}}{\beta_j^1} \\ & = \sum_{l=1}^{\infty} u_l \frac{1}{j\omega\mu_0} \frac{k_z}{\beta_l^1} \langle \phi_j | \frac{1}{\epsilon_r} | \frac{\partial \overline{\psi_l}}{\partial y} \rangle - \sum_{m=1}^{\infty} v_m \frac{-j[k_z^2 - (\beta_m^u)^2]^{1/2}}{\beta_m^u} \langle \phi_j | \frac{1}{\epsilon_r} | \overline{\phi_m} \rangle \quad (2.49) \end{aligned}$$

The fact that all the radiation and evanescent modes are necessary in order to match boundary may seem to suggest power lost to radiation. Because we are considering the case for which all the reflection and transmission angles are imaginary, it is easy to see that all the higher order power excited at the discontinuity is completely reactive and none of it is radiated.

2.3.2. Approximate Solution

We consider the lowest order approximation to the reflection coefficient of the TE_1 mode first. Two modes are included in the approximation, the TE_1 mode for the two regions. Next, we shall consider the correction to the lowest order approximation due to the TM_1 modes.

For the lowest order approximation, (2.47) and (2.48) become

$$\begin{aligned}
 & -\frac{[(\beta_1')^2 - k_z^2]^{1/2}}{\beta_1'} + r_1^{(1)} \frac{j[k_z^2 - (\beta_1')^2]^{1/2}}{\beta_1'} \\
 & = -u_1^{(1)} \frac{-j[k_z^2 - (\beta_1')^2]^{1/2}}{\beta_1'} \langle \psi_1 | \bar{\psi}_1 \rangle
 \end{aligned} \tag{2.50}$$

and

$$\beta_1' + r_1^{(1)} \beta_1' = u_1^{(1)} \bar{\beta}_1' \langle \psi_1 | \psi_1 \rangle \tag{2.51}$$

respectively. Solving for the reflection coefficient $r_1^{(1)}$, we obtain

$$\begin{aligned}
 r_1^{(1)} &= \frac{\frac{[(\beta_1')^2 - k_z^2]^{1/2}}{\beta_1'^2} - \frac{-j[k_z^2 - (\beta_1')^2]^{1/2}}{(\beta_1')^2}}{\frac{[(\beta_1')^2 - k_z^2]^{1/2}}{\beta_1'^2} + \frac{-j[k_z^2 - (\beta_1')^2]^{1/2}}{(\beta_1')^2}} \\
 &= e^{j2\delta^{(1)}}
 \end{aligned} \tag{2.52}$$

where

$$\begin{aligned}
 \delta^{(1)} &= \tan^{-1} \left[\left(\frac{\beta_1'}{\beta_1} \right)^2 \left[\frac{k_z^2 - (\beta_1')^2}{(\beta_1')^2 - k_z^2} \right]^{1/2} \right] \\
 &= \tan^{-1} \left[\left(\frac{n_1'}{n_1} \right)^2 \left[\frac{k_z^2 - (\overline{n_1'} k_0)^2}{(n_1' k_0)^2 - k_z^2} \right]^{1/2} \right]
 \end{aligned} \tag{2.53}$$

The definition of the effective index of refraction (2.16) has been used to cast (2.53) into a form reminiscent to the $\tan^{-1}$ term in (2.11).

Consider another scattering problem, that of two semi-infinite media of index of refraction n_1' and $\overline{n_1'}$ for $x \leq 0$ and $x > 0$, respectively. The reflection coefficient of a TM wave incident from $x < 0$ with angle θ_1' to the interface is exactly the same as (2.53). This is the essence of the effective dielectric constant method [3]. The lowest order approximation to the exact mode-matching method is, therefore equivalent to the effective dielectric constant method. Equation (2.5.2) can also be casted into the form

$$\begin{aligned}
 r_1^{(1)} &= \frac{z_k' \sin \theta_1' - \overline{z_k'} \sin \overline{\theta_1'}}{z_k' \sin \theta_1' + \overline{z_k'} \sin \overline{\theta_1'}} \\
 &= \frac{z_x' - \overline{z_x'}}{z_x' + \overline{z_x'}}
 \end{aligned} \tag{2.54}$$

where z_k is the impedance of the slab mode in the propagation direction (2.14), and z_x is the impedance of the slab mode looking in the x-direction. This is exactly the result one would get by applying the transmission line analogy to the problem [7]. Now, we have two different points of view for looking at the lowest order approximation.

The transmission coefficient $u_j^{(1)}$ can be obtained similarly.

$$u_j^{(1)} = \frac{\beta_j^i}{\beta_j^r} \frac{2z_x'}{z_x' + \bar{z}_x^r} \frac{1}{\langle \psi_j | \bar{\psi}_j \rangle} \quad (2.55)$$

Although power is conserved when $\bar{\theta}_j^r$ is imaginary, it is not conserved if $\bar{\theta}_j^r$ is real. If, instead of (2.46) and (2.48), we take $\langle \bar{\phi}_j | (2.42) \rangle$ and $\langle \bar{\psi}_j | (2.44) \rangle$. It has been shown that power is always conserved in all orders of approximation for arbitrary incident mode and angle [6]. The corresponding lowest order reflection coefficient becomes

$$R_j^{(1)} = \frac{z_x' - |\langle \bar{\psi}_j | \psi_j \rangle|^2 \bar{z}_x^r}{z_x' + |\langle \bar{\psi}_j | \psi_j \rangle|^2 \bar{z}_x^r} \quad (2.56)$$

$$\equiv e^{j2\Delta^{(1)}}$$

where

$$\Delta^{(1)} = \tan^{-1} \left[H \left(\frac{n_j^i}{n_j^r} \right)^2 \frac{k_z^2 - (\bar{n}_j^r k_0)^2}{(n_j^i k_0)^2 - k_z^2} \right]^{1/2} \quad (2.57)$$

and $H = |\langle \bar{\psi}_j | \psi_j \rangle|^2$.

The transmission coefficient becomes

$$u_j^{(1)} = \frac{\beta_j^i}{\beta_j^r} \frac{2z_x'}{z_x' + \bar{z}_x^r} \langle \bar{\psi}_j | \psi_j \rangle \quad (2.58)$$

Comparing (2.53) and (2.57), the factor H therefore serves as a measure of how good the approximation is. The closer to unity it is the better is the approximation. Physically, this corresponds to smaller difference in b and c and hence smaller discontinuity.

Equations (2.46)-(2.49) lend themselves to successive approximation. We can derive the correction to the reflection coefficient $r_1^{(1)}$ due to the TM_1 slab mode by writing

$$\begin{aligned} r_1^{(2)} &= r_1^{(1)} + \delta r_1 \\ u_1^{(2)} &= u_1^{(1)} + \delta u_1 \end{aligned} \quad (2.59)$$

Equations (2.46)-(2.49) become

$$s_1 \beta_1'' = v_1 \langle \phi_1 | \frac{1}{\epsilon_r} | \bar{\phi}_1 \rangle \quad (2.60)$$

$$\begin{aligned} (-1 + r_1^{(2)}) \frac{[(\beta_1')^2 - k_z^2]^{1/2}}{\beta_1'} &= s_1 \frac{1}{j\omega\epsilon_0} \frac{k_z}{1} \langle \psi_1 | \frac{1}{\epsilon_r} | \frac{\partial \phi_1}{\partial y} \rangle \\ &= -u_1^{(2)} \frac{-j[k_z^2 - (\beta_1')^2]^{1/2}}{\beta_1'} \langle \psi_1 | \bar{\psi}_1 \rangle - v_1 \frac{1}{j\omega\epsilon_0} \frac{k_z}{\beta_1''} \langle \psi_1 | \frac{1}{\epsilon_r} | \frac{\partial \bar{\phi}_1}{\partial y} \rangle \end{aligned} \quad (2.61)$$

$$(1 + r_1^{(2)}) \beta_1' = u_1 \bar{\beta}_1' \langle \psi_1 | \bar{\psi}_1 \rangle \quad (2.62)$$

and

$$\begin{aligned} (1 + r_1^{(2)}) \frac{1}{j\omega\mu_0} \frac{k_z}{\beta_1'} \langle \phi_1 | \frac{1}{\epsilon_r} | \frac{\partial \psi_1}{\partial y} \rangle + s_1 \frac{j[k_z^2 - (\beta_1'')^2]^{1/2}}{\beta_1''} \\ = u_1 \frac{1}{j\omega\mu_0} \frac{k_z}{\beta_1'} \langle \phi_1 | \frac{1}{\epsilon_r} | \frac{\partial \bar{\psi}_1}{\partial y} \rangle - v_1 \frac{-j[k_z^2 - (\beta_1'')^2]^{1/2}}{\beta_1''} \langle \phi_1 | \frac{1}{\epsilon_r} | \phi_1 \rangle \end{aligned} \quad (2.63)$$

respectively.

After long and tedious manipulations, we have

$$\delta r_1 = \frac{-(1 + r_1^{(1)})}{[1 + \frac{BC}{AD}]} \quad (2.64)$$

where

$$A = \frac{1}{j\omega\mu_0} \left[\frac{k_z}{\beta_1^2} \langle \phi_1 | \frac{1}{\epsilon_r} | \frac{\partial \psi_1}{\partial y} \rangle - \frac{\beta_1^2}{\beta_1^2} \frac{1}{\langle \psi_1 | \psi_1 \rangle} \frac{k_z}{\beta_1^2} \langle \phi_1 | \frac{1}{\epsilon_r} | \frac{\partial \overline{\psi_1}}{\partial y} \rangle \right] \quad (2.65)$$

$$B = \frac{j[k_z^2 - (\beta_1^2)^2]^{1/2}}{\beta_1^2} + \frac{j}{\beta_1^2} \frac{-j[k_z^2 - (\beta_1^2)^2]^{1/2}}{\beta_1^2} \quad (2.66)$$

$$C = \frac{[(\beta_1^2)^2 - k_z^2]^{1/2}}{\beta_1^2} + \frac{\beta_1^2}{\beta_1^2} \frac{-j[k_z^2 - (\beta_1^2)^2]^{1/2}}{\beta_1^2} \quad (2.67)$$

and

$$D = \frac{1}{j\omega\epsilon_0} \left[\frac{k_z}{\beta_1^2} \langle \psi_1 | \frac{1}{\epsilon_r} | \frac{\partial \phi_1}{\partial y} \rangle - \frac{j}{\beta_1^2} \frac{1}{\langle \phi_1 | \frac{1}{\epsilon_r} | \phi_1 \rangle} \frac{k_z}{\beta_1^2} \langle \psi_1 | \frac{1}{\epsilon_r} | \frac{\partial \overline{\phi_1}}{\partial y} \rangle \right] \quad (2.68)$$

Because ψ_1 and ϕ_1 are both symmetric functions of y , A and D are expected to be small. It is also interesting to notice that TE and TM modes coupled through the terms A and D . Although the incident TE mode has only five field components, the reflected wave has all six of the field components.

2.4. Dielectric Ridge Waveguide

Our purpose in this section is to use the approximation and physical understanding developed in the last section to discuss the properties of the dielectric ridge waveguide [8]. The general properties of the guided modes are discussed first. From the waveguide design point of view, the most important properties of the waveguide are the propagation constant and the cutoff frequencies of the higher order modes. These are taken up in the second part of the section.

2.4.1. Guided Modes

The dielectric ridge waveguide consists of a central guiding region of width a and height b supported by a web of thickness t , as shown in Fig. 2.5. The cross section of the ridge waveguide is seen to consist

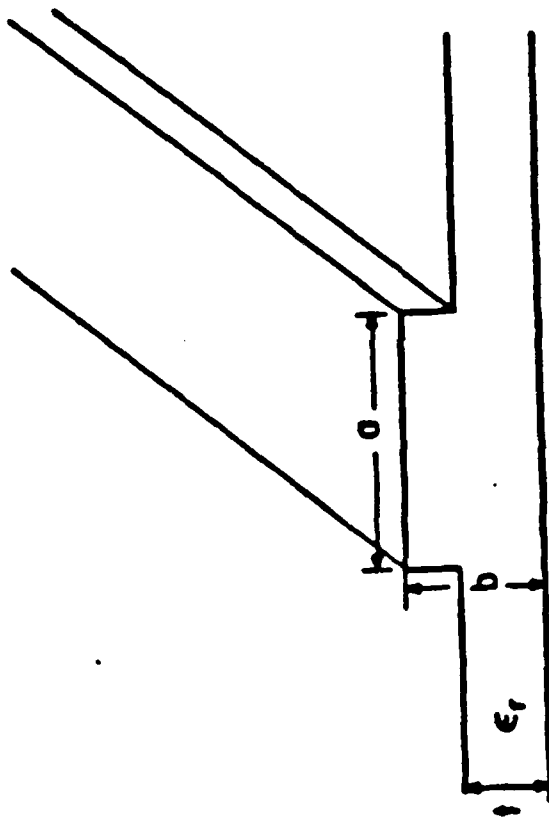


Fig. 2.5

of two dielectric step junctions of the type just discussed connected by a length of uniform slab waveguide. The guiding process of the guided mode can be viewed as multiple reflections taken place at the two step discontinuities forming the waveguide side walls. The exact mode-matching technique discussed in the last section can be applied to solve the waveguide problem. The fields of the guided mode in the core region and in the web region can be expanded in terms of the slab modes of their respective region. The guiding characteristics of the waveguide are completely determined by a single parameter, i.e., the longitudinal propagation constant k_z . The boundary value problem has been solved by the transverse resonance condition [6].

Because of the TE-TM mode coupling occurring at the side walls of the waveguide, the guided modes of the ridge waveguide have all six field components. The guided modes, however, can still be classified according to their dominant electric field direction. As is clear from the discussion in the last section, the TE_q slab mode derived guided mode has the dominant electric field component in the x-direction. By convention, this type of mode is designated as the E_{pq}^x mode where p and q are the number of field extrema in the x- and y-direction, respectively. Similarly, E_{pq}^y modes are derived from the TM_q slab mode and have their dominant electric field in the y-direction.

The dielectric ridge waveguide can support two fundamental modes, the E_{11}^x and the E_{11}^y modes. These two modes have no cutoff frequency and are always propagating. As was pointed out by Peng and Oliner [10], all modes of the ridge waveguide are leaky except for the E_{11}^x mode. This is caused by the TE-TM mode coupling. Let's consider the E_{11}^y mode as an example. As was pointed out earlier, the E_{11}^y mode is mainly derived from the slab TM_1 mode. The propagation constant of the guided mode k_z is intermediate

between $n_1'' k_0$ and $\bar{n}_1'' k_0$. As is obvious from Fig. 2.2, at a given frequency, $\bar{n}_1'' k_0$ of the TE_1 slab mode in the web region is greater than $n_1'' k_0$ if t ($t < b$) is greater than t_c . Because $\bar{k}_x^T$, the propagation constant in the x-direction of the slab mode in the web region, is related to $\bar{n}_1'' k_0$, and k_z is smaller than $n_1'' k_0$, $\bar{k}_x^T$ is therefore real;

$$\begin{aligned} (\bar{k}_x^T)^2 &= (\bar{n}_1'' k_0)^2 - k_z^2 \\ &\geq [(\bar{n}_1'')^2 - (n_1'')^2] k_0^2 > 0 \end{aligned}$$

A real $\bar{k}_x^T$ means power will leak into the TE_1 slab mode of the web region and propagate away. A similar argument can be applied to all the higher order modes of the ridge waveguide. The leakage phenomenon has been demonstrated experimentally [11]. Because of the leakage phenomenon, the operating mode of the dielectric ridge waveguide is chosen to be the E_{11}^x mode. In order to avoid the leaky higher order modes, it is important to be able to predict the cutoff frequencies of the higher order modes. The operating regime of the waveguide where only the two fundamental modes are propagating is called the quasi-single mode regime.

The field components of the E_{11}^x mode, to lowest order of approximation, are listed below for reference

$$\begin{aligned} E_x &= \frac{k_z}{n_1'' k_0} \psi_1(y) \cos k_x x & H_x &= \frac{1}{\omega \mu_0} \frac{k_x}{n_1'' k_0} \frac{\partial \psi}{\partial y} \sin k_x x \\ E_y &= 0 & H_y &= \frac{n_1'' k_0}{\omega \mu_0} \psi(y) \cos k_x x \\ E_z &= \frac{j k_x}{n_1'' k_0} \psi(y) \sin k_x x & H_z &= -\frac{1}{j \omega \mu_0} \frac{k_z}{n_1'' k_0} \frac{\partial \psi}{\partial y} \cos k_x x \end{aligned} \quad (2.69)$$

for $|x| \leq \frac{a}{2}$. The fields for $|x| > \frac{a}{2}$ are obvious and will not be listed here.

2.4.2. Propagation Constant and Cutoff Frequencies of Higher Order

Modes

Knowing the reflection coefficient $r_1^{(1)}$ (2.52) and (2.53), we can calculate the propagation constant k_z of the E_{11}^x mode. By transverse resonance condition, under the lowest order approximation, the round trip phase shift in the x-direction in the core region is given by

$$e^{-j2k'_x a} e^{j4\delta^{(1)}} = e^{-j2(m-1)\pi} \quad (2.70)$$

where $k'_x = [(n_1' k_0)^2 - k_z^2]^{1/2}$, and $m = 1, 2, 3, \dots$. The characteristic equation for the E_{1m}^x mode is therefore

$$k_x a - 2\delta^{(1)} = (m-1)\pi \quad (2.71)$$

This can be transformed to a form identical to (2.11); we have

$$k_x a = m\pi - 2 \tan^{-1} \left[\left(\frac{n_1'}{n_1} \right)^2 \frac{(n_1' k_0)^2 - k_z^2}{k_z^2 - (n_1' k_0)^2} \right]^{1/2} \quad (2.72)$$

This is the well-known effective dielectric constant result. Its validity depends, of course, on the accuracy of the reflection phase shift $\delta^{(1)}$. The waveguide width a also affects the accuracy of (2.72) as follows. All the higher order slab modes decay exponentially away from the side walls as $e^{-\bar{\alpha}x}$. If $e^{-\bar{\alpha}a}$ is still significant (i.e., the tail of the higher slab mode generated at one side wall is still significant at the other side wall), the reflection on the side wall can no longer be treated as independent; as was implicitly assumed in (2.70).

The cutoff frequency f_{21}^x of the E_{21}^x mode can be estimated by (2.72). For E_{21}^x mode, $k'_x a$ is greater than π ; $k_x = \frac{\pi}{a}$ at the cutoff frequency. Equation (2.72) becomes

$$\frac{\pi}{2} = \tan^{-1} \left[\left(\frac{n_1'}{n_1} \right)^2 \frac{(n_1' k_0)^2 - k_z^2}{k_z^2 - (n_1' k_0)^2} \right]^{1/2}$$

That is

$$k_z^2 = (\overline{n_1'} k_0)^2 - k_x^2$$

Therefore, at the cutoff frequency, we have

$$f_{21}^x : a_c = \frac{\lambda_0}{2[(n_1')^2 - (\overline{n_1'})^2]^{1/2}} \quad (2.73)$$

where λ_0 is the free space wavelength.

Similarly, the cutoff frequency f_{12}^x of the E_{12}^x mode occurs when

$$f_{12}^x : b_c = \frac{\lambda_0}{2[n^2 - n_0^2]^{1/2}} \quad (2.74)$$

This is an exact relation. Because E_{12}^x mode is derived from the slab TE_2 mode of the core region, this mode cannot become propagating before the TE_2 slab mode does.

2.5. Radiation Loss from Waveguide Bends

In this section, we first discuss the general properties of radiation loss from waveguide bends. A synthesis of the mode-matching technique and Mareatili's [12] theory is presented next.

2.5.1. Introduction

Figure 2.6 shows a dielectric ridge waveguide bend having radius of curvature R . There are two components to the bend loss; the dissipative loss and the mode conversion loss. The dissipative loss, resulting from either radiation loss or coupling to a high loss undesired mode, is

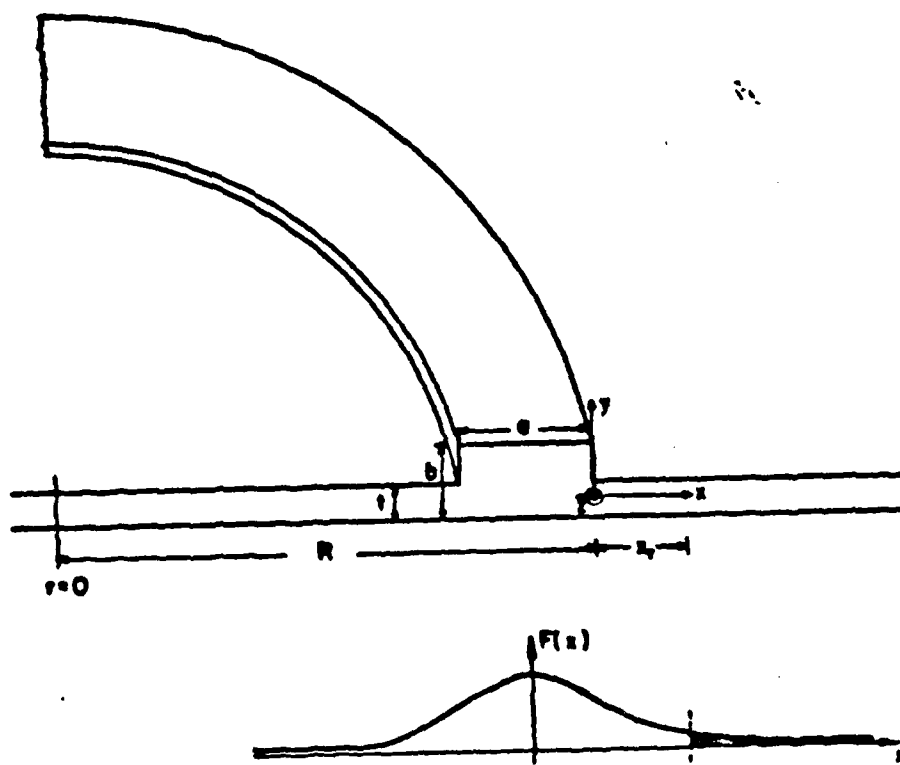


Fig. 2.6

characterized by an attenuation coefficient α_r . The mode conversion loss for the straight waveguide mode is incurred on entering or leaving the curved region. It has been shown that for single-mode open waveguide whose transverse fields extend indefinitely into the surrounding medium, the bend loss is dominated by radiation loss in the regime of tolerable total loss [13]. We shall, therefore, concentrate on radiation loss.

Qualitatively, the origin of radiation loss can be understood as follows. Assume that the transverse field distribution $F(x)$ is virtually the same in the curved region as in a straight guide. Since the fields in the core region travels at a velocity close to the unperturbed value. At some transverse distance x_r , the fields (the hatched portion) will be traveling at a velocity greater than that of the phase velocity of the slab mode in that region. This portion of energy will be coupled to the slab mode and being radiated away. It was argued in [13] that this loss occurs in a longitudinal distance equal to the collimated-beam length of a Gaussian beam associated with the field distribution $F(x)$. The result of this analysis was shown to agree qualitatively with Marcetili's field analysis.

In Marcetili's paper [12], a dielectric rod waveguide was analyzed. It was assumed that the fields were strongly guided so that the boundary conditions for k_x and k_y were determined independently. In a curved waveguide, the radial wave functions can be expressed in terms of Bessel functions. The Watson's first term approximation to the Bessel functions was used to find an analytical approximation. The result showed that the radiation loss per radian $\alpha_r R$ is

$$\alpha_r R = M R e^{-R/R_0} \quad (2.75)$$

where M is a constant independent of R .

The function $\alpha_r R$ has an inflection point at $R = 2R_0$. The tangent at the $R = 2R_0$ goes to zero at $R = 4R_0$. It was argued that the radiation loss per radian became negligible for R greater than $4R_0$. It was then reasonable to define a critical radius of curvature R_c for the waveguide bend to be

$$R_c \equiv 4R_0 \quad (2.76)$$

Although Marcatili's original analysis was done for a dielectric rod waveguide and for strongly guided modes, it can be extended to the dielectric ridge waveguide and for weakly guided modes by using the mode-matching technique.

2.5.2. Mode-Matching Analysis

Consider an E_{p1}^x mode propagating in the azimuthal direction with dependence $e^{jv\theta - j\omega t}$. The simplest field components have the form

$$E_x = \frac{\omega\mu_0}{k_x^2} \frac{v}{x} J_v([(nk_0)^2 k_y^2]^{1/2} x) \psi_1(y) \quad H_x = \frac{1}{k_x^2} \frac{\partial J_v}{\partial x} \frac{\partial \psi_1}{\partial y}$$

$$E_y = 0 \quad H_y = J_v(x) \psi_1(y) \quad (2.77)$$

$$E_z = \frac{j\omega\mu_0}{k_x^2} \frac{\partial J_v}{\partial x} \psi_1(y) \quad H_z = \frac{jv}{k_x^2 x} J_v(x) \frac{\partial \psi_1}{\partial y}$$

where $n^2 k_0^2 - k_y^2 = n_0^2 k_0^2 + \alpha^2 = (n_1' k_0)^2$, and J_v 's are the Bessel functions.

It is easy to show that the function ψ_1 is nothing but the TE_1 slab mode function. Since the slab mode functions form a complete set of functions, it is natural to extend the mode-matching technique to the present problem. We shall not, however, try to write down the exact formulation. Instead, we shall use the effective dielectric constant method directly and apply Marcatili's approximation in

order to find an analytic approximation. The accuracy of this approach, of course, is subjected to the same restrictions as in the waveguide case.

We follow the notation and choices of the radial wave functions as in [12]. We have

$$\begin{aligned}
 J_\nu[(\bar{n}_1^T k_0)(R+x)] & \quad 0 < x \leq R-a \quad \text{region 2} \\
 J_\nu = J_\nu[(n_1^T k_0)(R+x) + \psi] & \quad (R-a) \leq x \leq R \quad \text{region 1} \\
 H_\nu^{(2)}[(\bar{n}_1^T k_0)(R+x)] & \quad x > R \quad \text{region 3}
 \end{aligned} \tag{2.78}$$

Since n_1^T and $\bar{n}_1^T$ are known, the variable is the order ν of the Bessel functions. Matching boundary at $x = R-a$ and $x = R$ yields

$$\frac{J_\nu(\rho_{13})}{J_\nu'(\rho_{13})} = \frac{\rho_3}{\rho_{13}} \frac{H_\nu^{(2)}(\rho_3)}{H_\nu^{(2)'}(\rho_3)} \tag{2.79}$$

and

$$\frac{J_\nu(\rho_{12})}{J_\nu'(\rho_{12})} = \frac{\rho_2}{\rho_{12}} \frac{J_\nu(\rho_2)}{J_\nu'(\rho_2)} \tag{2.80}$$

where

$$\begin{aligned}
 \rho_{13} &= n_1^T k_0 R + \psi, \quad \rho_{12} = n_1^T k_0 (R-a) + \psi \\
 \rho_3 &= \bar{n}_1^T R \quad \text{and} \quad \rho_2 = \bar{n}_1^T k_0 (R-a)
 \end{aligned} \tag{2.81}$$

Note that, up to this point, the form of the equations are identical to Marcatili's if the factor $\bar{n}_1^T k_0 = [(n_3 k_0)^2 - (k_y')^2]^{1/2}$, is replaced by

$$\bar{n}_1^T k_0 \rightarrow [(n_3 k_0)^2 - (k_y')^2]^{1/2}$$

where k'_y is that corresponding to the core region, and $n_3 = n$ in our case.

Following Marcatili, the Bessel functions are approximated by their Watson's first term approximation. These approximations are valid if

$$\frac{v^2}{\rho^2 - v^2} \ll 1$$

However, this is not strictly true for the case of interest. The characteristic equations (2.79) (2.80) now become

$$\begin{aligned} & \frac{1}{3v^2} [(\rho_{13}^2 - v^2)^{3/2} - (\rho_{12}^2 - v^2)^{3/2}] \\ &= p\pi - \tan^{-1} \left[\left(\frac{n_1}{n_1} \right)^2 \frac{\rho_{13}^2 - v^2}{v^2 - \rho_3^2} \right]^{1/2} \left[1 + j \exp \left(- \frac{2}{3} \frac{(v^2 - \rho_3^2)^{3/2}}{v^2} \right) \right] \\ & \quad - \tan^{-1} \left[\left(\frac{n_1}{n_1} \right)^2 \frac{\rho_{12}^2 - v^2}{v^2 - \rho_3^2} \right]^{1/2} \end{aligned} \quad (2.82)$$

This can be simplified by expanding in terms of $1/R$ to yield

$$\begin{aligned} k_x a = p\pi - \tan^{-1} \left[\left(\frac{n_1}{n_1} \right)^2 \frac{k_x}{\gamma} \left[1 + j \exp \left(- \frac{2}{3} \frac{R \gamma^2}{k_z^3} \right) \right] \right] \\ - \tan^{-1} \left[\left(\frac{n_1}{n_1} \right)^2 \frac{k_x}{\gamma} \left[1 - \frac{a}{R} \left[\left(\frac{n_1 k_0}{k_x} \right)^2 + \left(\frac{n_1 k_0}{\gamma} \right)^2 \right] \right] \right] \end{aligned} \quad (2.83)$$

where $\gamma = [k_z^2 - (n_1 k_0)^2]^{1/2}$, and $v = k_z R$. Let k_{x0} be the solution when R approaches infinity; i.e., k_{x0} satisfies

$$k_{x0} a = p\pi - 2 \tan^{-1} \left[\left(\frac{n_1}{n_1} \right)^2 \frac{k_{x0}}{\gamma_0} \right]$$

as is expected. The perturbation to k_x , k_x and γ due to finite (but large) R can be obtained by varying k_x , x_z and γ . Let $k_x = k_{x0} (1 + \delta)$, $\gamma = \gamma_0 (1 + \delta')$ and $k_z = k_{z0} (1 + \delta'')$. The quantities δ , δ' and δ'' are related by

$$\delta' = - \left(\frac{k_{x0}}{\gamma_0} \right)^2 \delta \quad (2.84)$$

and

$$\delta'' = - \left(\frac{k_{x0}}{k_{z0}} \right)^2 \delta \quad (2.85)$$

for $k_x^2 + k_z^2 = (n_1' k_0)^2$ and $k_z^2 - \gamma^2 = (\bar{n}_1' k_0)^2$. by retaining only the first order terms, we have

$$\begin{aligned} \delta &= \frac{\left(\frac{n_1'}{\bar{n}_1'} \right)^2 \left(\frac{k_{x0}}{\gamma_0} \right)^2 \left[\frac{a}{R} \left[\left(\frac{n_1' k_0}{k_{x0}} \right)^2 + \left(\frac{\bar{n}_1' k_0}{\gamma_0} \right)^2 \right] - j \exp \left[- \frac{2}{3} \frac{R \gamma_0^3}{k_{z0}^2} \right] \right]}{\{ k_{x0}^2 a \left[1 + \left[\left(\frac{\bar{n}_1'}{n_1'} \right)^4 \left(\frac{k_{x0}}{\gamma_0} \right)^2 \right] + 2 \left[\left(\frac{\bar{n}_1'}{n_1'} \right)^2 \left(\frac{k_{x0}}{\gamma_0} \right) \left[1 + \left(\frac{k_{x0}}{\gamma_0} \right)^2 \right] \right] \}} \\ &\equiv \frac{2c}{a k_{x0}} - j \frac{k_{z0} \alpha_r}{k_{x0}^2} \end{aligned} \quad (2.86)$$

where we have defined the constants c and α_r , following Marcatili. It follows from (2.85) that α_r is the attenuation coefficient of the guided wave due to radiation loss and c^2 is the mode conversion loss that would exist at a junction between a straight guide and a curved one.

The critical radius of curvature R_c is, by (2.86)

$$R_c = 6 \frac{k_{z0}^2}{\gamma_0^3} \quad (2.87)$$

It has the same form as Marcatili's, except that k_{z0} and γ_0 are now calculated from the effective dielectric constant method. The multiplication constant M is, of course, different from Marcatili's.

References

- [1] D. Marcuse, Theory of Dielectric Optical Waveguides, Academic Press, New York, 1974.
- [2] H. Kogelnik, "Theory of dielectric waveguides," in Integrated Optics, T. Tamir ed., Springer-Verlag, New York, 1975.
- [3] R. M. Knox and P. P. Toullos, "Integrated circuits for millimeter wave through optical frequency range," in Proc. Symposium on Submillimeter Waves, Polytechnic Press, Brooklyn, New York, pp. 497-516, 1970.
- [4] K. Solbach and I. Wolff, "The electromagnetic fields and phase constants of dielectric image lines," IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 266-274, 1978.
- [5] W. Schlosser and H. G. Unger, "Partially filled waveguides and surface waveguides of rectangular cross section," in Advance in Microwaves, L. Young, ed., Academic Press, New York, vol. 1, pp. 319-387, 1966.
- [6] S-T. Peng and A. A. Oliner, "Guidance and leakage properties of a class of open dielectric waveguides: part I - mathematical formulations," IEEE Trans. Microwave Theory Tech., vol. MTT-29, pp. 843-855, 1981.
- [7] S. Ramo, J. R. Whinnery, and T. Van Duzer, Fields and Waves in Communication Electronics, John Wiley, New York, 1965.
- [8] T. Wang and S. E. Schwarz, "Design of dielectric ridge waveguides for millimeter-wave integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-31, pp. 128-134, 1982.
- [9] A. A. Oliner, S-T. Peng, T-I. Hsu, and A. Sanchez, "Guidance and leakage properties of a class of open dielectric waveguides: part II

- new physical effects," IEEE Trans. Microwave Theory Tech., vol. MTT-29, pp. 855-869, 1981.
- [10] S-T. Peng and A. A. Oliner, "Leakage and resonance effects on strip waveguides for integrated optics," Trans. Electronics and Commun. Eng. Jap., vol. E61, pp. 151-154, 1978.
- [11] K. Ogusu, S. Kawakami, and S. Nishida, "Optical strip waveguide: an analysis," *Applied Optics*, vol. 18, pp. 908-914, 1979.
- [12] E. A. J. Marcatili, "Bends in optical dielectric guides," Bell Syst. Tech. J., vol. 48, pp. 2103-2132, 1969.
- [13] E. A. J. Marcatili and S. E. Miller, "Improved relations describing directional control in electromagnetic wave guidance," Bell Syst. Tech. J., vol. 48, pp. 2161-2188.

Figure Captions

- 2.1. The dielectric slab guide. The perfectly-conducting planes are included to facilitate the discussion of radiation and evanescent modes; they are far removed from the dielectric slab and have no influence on the guided modes.
- 2.2. The dispersion relations for the lowest four slab modes for slab guide of $\epsilon_r = 12$.
- 2.3. Reflection from a step discontinuity in a slab guide. The incident angle θ_i is θ_i' when the incident wave is the TE_1 mode.
- 2.4. Schematic representation of a TE_1 slab mode scattered by a step discontinuity along z-axis. The field components of each of the incident and scattered slab modes are also indicated.
- 2.5. A dielectric ridge waveguide (DRG) of width a , height b , and web thickness t . It is seen that the DRG is formed by two step discontinuities separated by a distance a .
- 2.6. A curved section of the DRG. The radius of curvature R is defined from the center to the outer side of the waveguide. Also shown is a schematic representation of field distribution. For a ridge field distribution, the shaded part of the field is traveling at a speed faster than that of the lowest slab mode and hence will be radiated.

CHAPTER 3

DESIGN OF DIELECTRIC RIDGE WAVEGUIDE

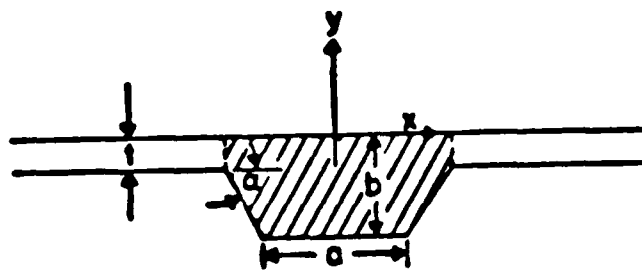
The principal design considerations are: (a) quasi-single mode propagation; (b) low radiation loss in waveguide bends; (c) low coupling loss between waveguide and the lumped circuit element located at the vertex of the V-coupler; and (d) adequate physical strength.

Before we embark on the design of the waveguide itself, we have to verify the accuracy of the lowest order approximation of the mode-matching technique (i.e., the effective dielectric constant method). We then proceed to the design of the waveguide and the V-coupler. Finally, the performance of the design is summarized [1].

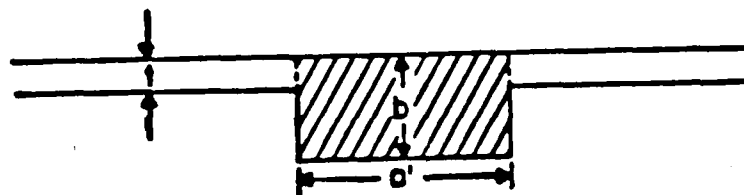
3.1. The Effective Dielectric Constant Method

There are two reasons to verify the accuracy of the effective dielectric constant [EDC] method (2.72). Firstly, although the EDC method is widely used, experimental confirmations have been obtained either for multimode waveguides or for guides in which the differences of dielectric constant between the guide and its surroundings were small [2], [3], and [4]. Secondly, the cross-section for the dielectric ridge waveguide fabricated on semiconductor (e.g., silicon) by anisotropic etching is trapezoidal, instead of rectangular, Fig. 3.1(a). Because in the quasi-single mode regime, the waveguide dimension is small compared with the free space wavelength, it is possible to represent the trapezoidal core region (the shaded region in Fig. 3.1) by an effective rectangular core region of the same height and core area whose width is a' , as shown in Fig. 3.1(b).

We have carried out a simulation experiment in the frequency range 2-7 GHz to verify the EDC method in the quasi-single mode regime. Emerson



(a)



(b)

Fig. 3.1

and Cuming HiK material with $\epsilon_r = 12$ is used to simulate a silicon waveguide at millimeter wavelengths. The experimental setup is shown in Fig. 3.2. Power from the sweep oscillator is coupled into the DRG by a V-coupler. A shorting plane at the far end of the waveguide establishes a standing wave. The reflected wave is coupled back into the coaxial cable by the V-coupler. The magnitude and phase of the reflected wave are then measured by the network analyzer. When a small conducting wire is placed near the waveguide, the reflected wave is disturbed, except when the wire is located at one of the nodes of the field parallel to the probe. Thus, the guide wavelength is measured by locating the nodes.

In Fig. 3.3, the measured propagation constant is plotted against normalized guide height b/λ_0 for guides with $a/b = 2$, for the two cases $t/b = 0.2$ and 0.4 . Also shown in the figure are the dispersion curves predicted by the EDC method. We see that the propagation constant is predicted quite well by the EDC method for $t/b = 0.4$. In this case, the error is less than 4 percent over the range of measurement. However, for $t/b = 0.2$, errors become more significant.

The next higher order mode for the waveguides under consideration is the E_{21}^x mode (2.73). For the waveguide with $t/b = 0.4$, at the cutoff frequency f_{21}^x , b is approximately $0.113 \lambda_0$; for a waveguide with $t/b = 0.2$, b is approximately $0.09 \lambda_0$ at the cutoff of E_{21}^x mode. As was mentioned in section 2.4.2, the accuracy of the EDC method depends on whether the reflection at each side wall can be treated as independent. This can be estimated as follows. According to Fig. 2.2 the effective index of refraction n_y^e is practically equal to one for $b/\lambda_0 = 0.1$. [The reason we pick $b/\lambda_0 = 0.10$ is that the overall design of the waveguide is useful over 10% Bandwidth below f_{21}^x . See sec. 3.3.] Therefore, K_{y1}^e can be approximated by $[n^2 - 1]^{1/2} = 20.84$. The decay length of the TM_1 slab

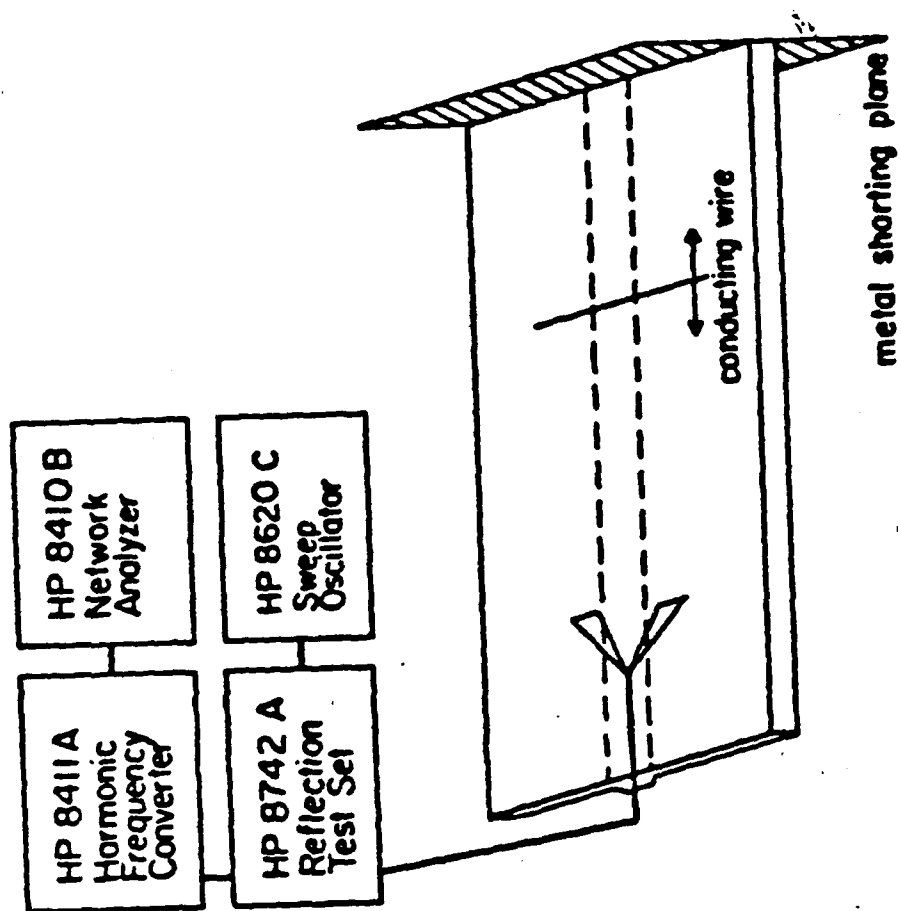
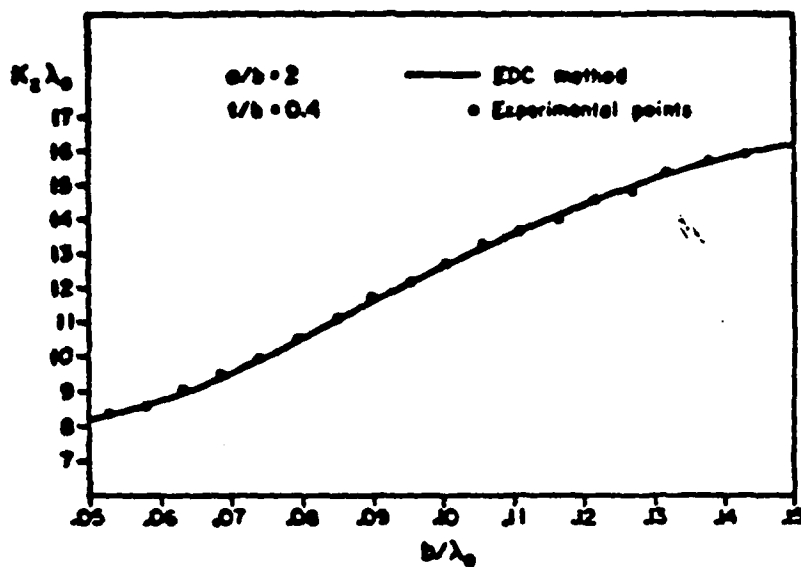
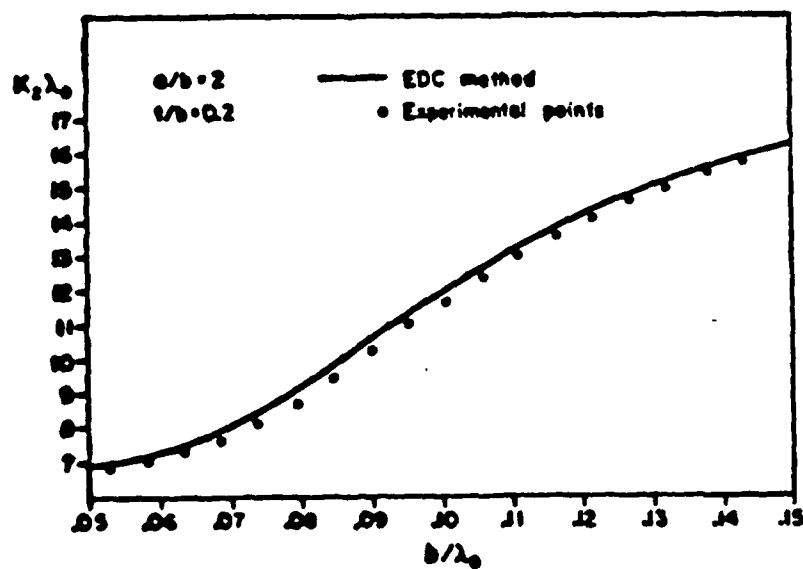


Fig. 3.2



(a)



(b)

Fig. 3.3

mode away from the side way is given by $1/\alpha''$, where $-(\alpha'')^2 + (k_{y1}'')^2 + k_z^2 = n^2 k_0^2$. For a waveguide with $t/b = 0.4$ at frequency of $b/\lambda_0 = 0.10$, we have the decay length $1/\alpha'' \approx 0.09 \lambda_0$. The decay factor $e^{-\alpha'' a}$ come out about only 7%. Since the TM_1 slab mode component is small, compared with the TE_1 slab mode (from symmetry), to start with, the reflections at two side walls can be treated as independent. The EDC method is expected to be accurate in this situation, as was verified experimentally (Fig. 3.3(a)). The agreement at lower frequencies is due to a smaller discontinuity (in terms of free space wavelength), and hence the other modes are less significant. For the waveguide with $t/b = 0.2$ at a frequency $b/\lambda_0 = 0.08$, the decay length of the TM_1 slab mode $1/\alpha''$ is $0.166 \lambda_0$. Together with the fact that the TM_1 slab mode component is larger since the discontinuity is larger, it is not surprising that the EDC method is not accurate for this case. For waveguides with larger a/b ratio, we expect the EDC to be accurate for the same reason. For a waveguide with $t/b = 0.4$ and $a/b = 4.0$, at the cutoff frequency, b/λ_0 is approximately $0.07 \lambda_0$. If the frequency is $b/\lambda_0 = 0.065 \lambda_0$, $1/\alpha''$ is about $0.12 \lambda_0$. The decay factor $e^{-\alpha'' x}$ is only 0.02. This is another reason that the EDC method is more accurate for waveguides with larger aspect ratio a/b . In the frequency range of quasi-single mode operation the core area $a'b$ is about the same for waveguides with different aspect ratio. For waveguides with larger a/b , the waveguide must operate at a smaller normalized frequency b/λ_0 . Since the discontinuity becomes smaller in terms of free space wavelength, single-mode approximation is expected to be better.

3.2. Design of the Waveguide

Having verified the accuracy of the EDC method for cases of interest,

we can now use it to design the waveguide. There are several considerations that will affect our choice of waveguide geometry: a) all the higher order modes are to be cut off; b) reasonably small radii of curvature should give tolerable radiation loss; c) coupling loss between waveguide and V-coupler should be acceptably small; and d) the web thickness should be thick enough to provide adequate physical strength. The first of these considerations has already been dealt with. We shall discuss b), c), and d) in the following paragraphs.

Radiation loss from waveguide bends was discussed in sec. 2.5. It has been shown that the dependence of radiation loss on the radius of curvature R is of exponential nature, (2.86). When R is larger than a critical radius of curvature R_c , the radiation loss decreases slowly with increasing R . On the other hand, when R is smaller than R_c , radiation loss increases very rapidly as R decreases. The critical radius of curvature is given by (2.87). The exponential nature of the dependence of radiation loss on R and the validity of (2.87) has been verified in [5]. (Actually the critical radius of curvature verified in [5] is $2 k_z^2 / \gamma_0^3$. The numerical constant in (2.87) is of course somewhat arbitrary. We adopt the factor of 6 used by Marcatili [12] as being more conservative.)

The critical radius of curvature R_c and the radiation loss per radiant for $R = R_c$ at that frequency were calculated, for waveguides with $a/b = 4$ for several values of t/b ; the results are shown in Fig. 3.4 and Fig. 3.5, respectively. Over the 10% Bandwidth below f_{21}^x , the critical radius of curvature are generally between $2.5 \lambda_0$ to $3.0 \lambda_0$, with increasing R_c for larger values of t/b . Within the same frequency range, the radiation loss per radiant $\alpha_r R_c$ is generally between 0.13 dB to 0.25 dB.

The choice of the waveguide web to height ratio t/b is based on two considerations. On the one hand, we like to have t/b as large as possible for the sake of mechanical strength. On the other hand, as t/b becomes larger, both the critical radius of curvature R_c and radiation loss per radiant $\alpha_r R_c$ increase. Based on Fig. 3.4 and Fig. 3.5, it is reasonable to choose t/b to be 0.4.

The only variable left for the DRG design is the ratio a/b . In order for the higher order modes to be cut off, the product ab must not be too large. Thus, smaller a implies larger b , and since t/b has been fixed, larger t as well. Thus, greater mechanical strength is obtained with a smaller value of a/b . For the web to be thicker than $75 \mu\text{m}$, at 100 GHz, a/b has to be smaller than 4.5. As a/b becomes smaller, however, the fields are less confined in the x -direction and this leads to more radiation loss from a curved section with the critical radius of curvature. Thus, choice of a/b amounts to a compromise between mechanical strength and reduction of radiation loss from curved sections. In order not to unduly increase the radiation loss, $a/b \geq 2$ has been chosen. In order to choose a value of a/b within the range $2 \leq a/b \leq 4.5$, we must consider the influence of this parameter on the coupling loss between the guide and V-coupler.

3.3. Design of the V-Coupler

A V-coupler with inner angle θ , angular arm width $\Delta\theta$, and length l is shown in Fig. 3.6. The V-coupler is characterized by a driving-point impedance Z and coupling loss L . The design value of Z will be 100Ω . The objective of the design of the V-coupler and the DRG (with $t/b = 0.4$) is to find values of θ , $\Delta\theta$, l , and a/b that result in minimum coupling loss.

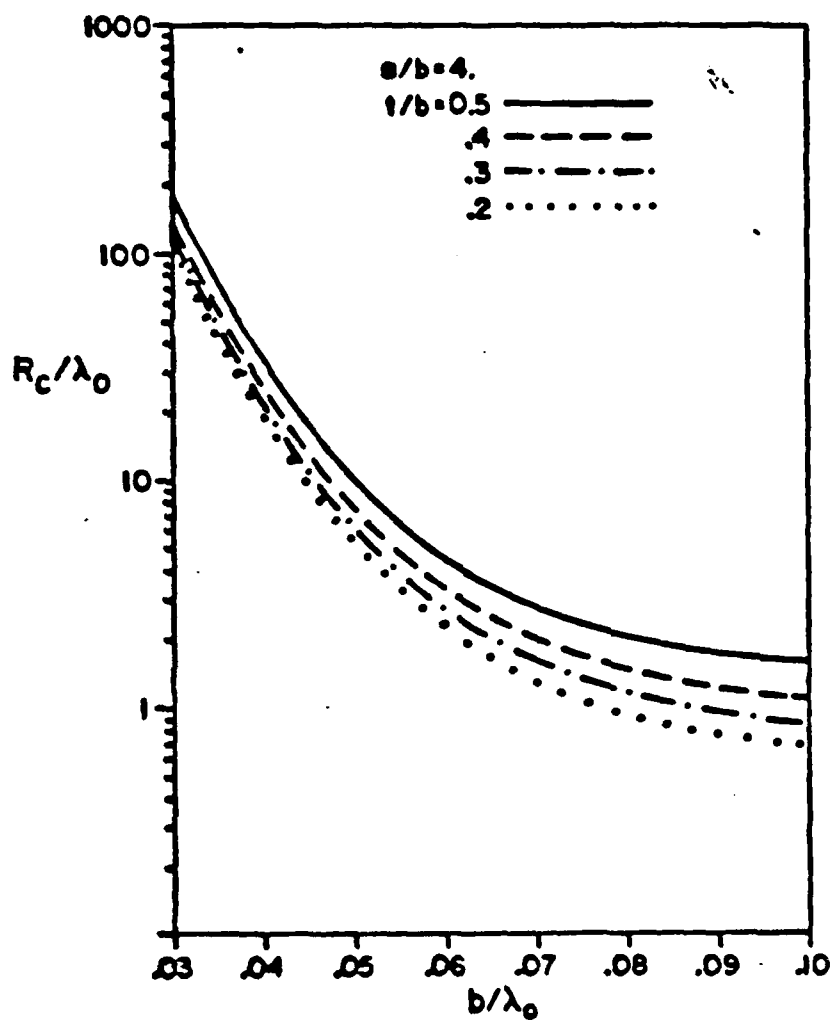


Fig. 3.4

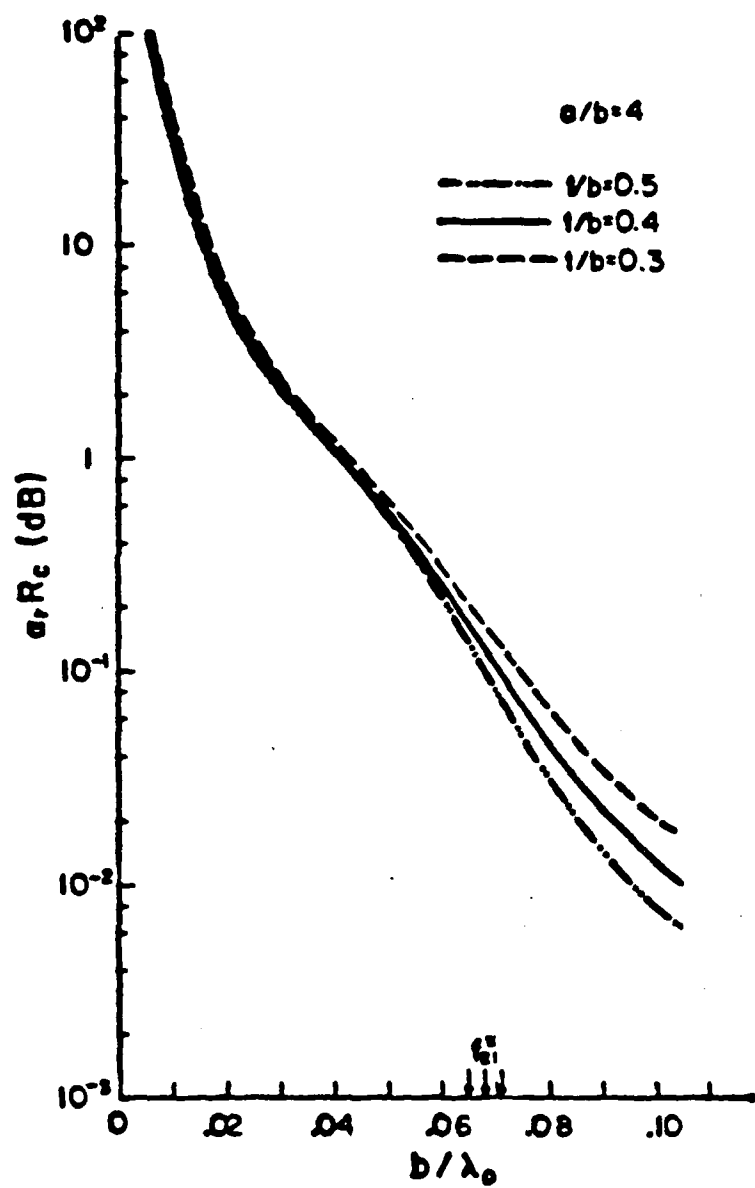


Fig. 3.5

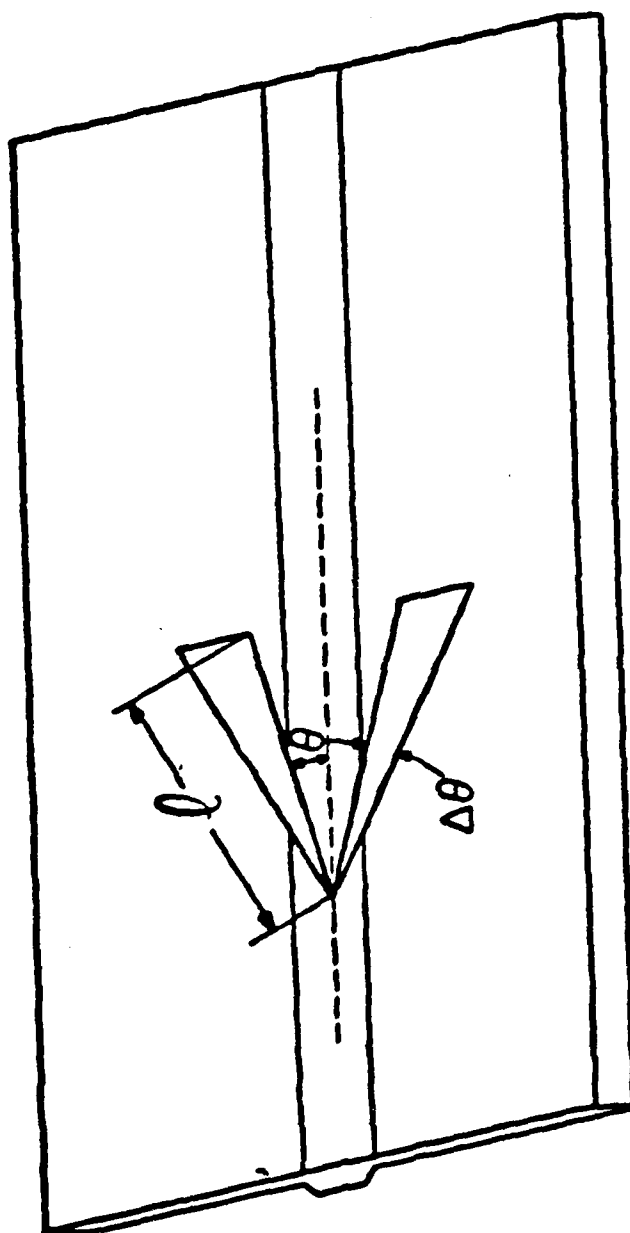


Fig. 3.6

In earlier work [6], [7], an approximate formula for the V-coupler's driving point impedance was used

$$Z = \frac{377 \Omega}{\sqrt{\frac{\epsilon_r + 1}{2}}} \frac{K(k)}{K(\sqrt{1-k^2})} \quad (3.1)$$

where K is the complete elliptic integral of the first kind and

$$k = \frac{\tan \frac{\theta}{2}}{\tan \frac{\theta + \Delta\theta}{2}} \quad (3.2)$$

This formula is based on conformal mapping of coplanar strips in free space. The correction factor

$$\sqrt{\frac{\epsilon_r + 1}{2}}$$

is used to take into account the presence of the dielectric waveguide [8]. This formula was shown to give estimates of Z to within 30 percent for rectangular and trapezoidal dielectric waveguides without webs [6], but our simulation experiments indicate that it is not a good approximation for V-couplers on DRG. The operation of the V-coupler involves radiation into free space, coupling to the slab modes, and ohmic loss. A complete description would require a full wave analysis, which is still lacking. The approach we have used is to design by means of simulation in the frequency range 2-6 GHz.

The experimental setup is shown in Fig. 3.7. The V-coupler is cut from 3-mil copper foil. An image plane is used in the simulation in order to avoid a balun between asymmetric coax and the symmetric V-coupler. The image plane reduces the observed driving-point impedance by half.

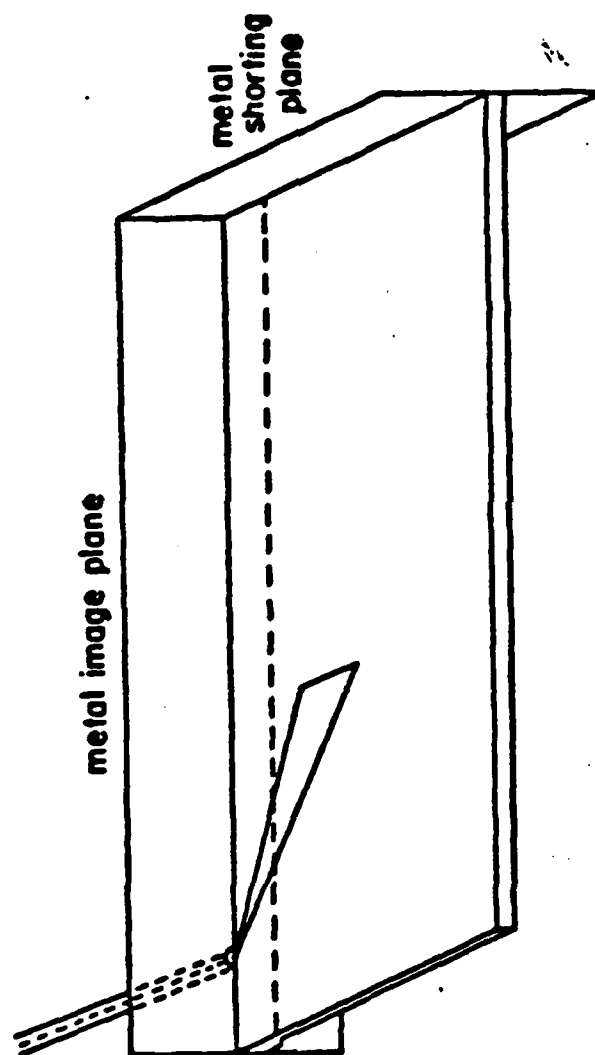


Fig. 3.7

When the waveguide is terminated by a good absorber, the network analyzer measures the driving-point impedance of the V-coupler. Once the driving-point impedance has been adjusted to 50Ω (with image plane; 100Ω without image plane), the network analyzer then measures the two-way coupling loss. (The measurement actually also includes whatever dielectric loss and scattering loss exist in the waveguide. These, however, are expected to be small and, in any case, do not affect minimization of the coupling loss.)

It is found experimentally that the driving-point impedance of the V-coupler does not vary much with the length of its arms, so long as $l \sin \theta \geq 0.7a$. The experimental procedure is to find a set of $\{\theta, \Delta\theta\}$ that gives the correct driving-point impedance for a long V-coupler on a given DRG. We find that typically a driving point reflection coefficient of less than -18 dB can be obtained over a 10 percent bandwidth. Because we are interested primarily in the frequencies just below the cutoff frequency f_{21} of the E_{21}^x mode (the next higher order mode for DRG with $t/b = 0.4$ and $a/b \geq 2$), the design is adjusted to give reflection coefficient <18 dB over approximately the range $0.9 f_{21} < f < f_{21}$.

Although (3.1) is not itself a good approximation, it is found that the set of designs $\{\theta, \Delta\theta\}$ giving the correct driving-point impedance is characterized by a value of the constant k defined in (3.2). This means that once a pair of angles $\theta, \Delta\theta$ is found that gives the desired impedance, other pairs of angles resulting in the same value of k will also give that impedance. Thus, by modifying the constant factor in (3.1), an approximate formula for Z may be obtained. The formula

$$Z = \frac{377 \Omega}{n' + \overline{n}'} \frac{K(k)}{K(\sqrt{1-k^2})} \quad (3.3)$$

is found to be accurate within 15 percent, where $\overline{n}'$ and n' are the

effective indices of refraction at $f = 0.95 f_{21}$.

Once θ and $\Delta\theta$ have been found, the coupling loss is measured as a function of coupler length and frequency. The measured coupled loss usually varies ± 0.2 dB over the frequency range of interest; thus, in comparing designs we actually compare typical values of the loss L . A plot of L as a function of coupler length on the waveguide with $a/b = 4$ for two pairs of angles $\theta, \Delta\theta$ (both of which give the desired driving-point impedance), is shown in Fig. 3.8. The coupling loss as a function of $\Delta\theta$ (with l always adjusted for lowest loss), for V-couplers on DRG with $a/b = 4$, is shown in Fig. 3.9. The corresponding designs are shown in Table 3.1. The best design for $a/b = 4$ is found to be $\theta = 9^\circ$, $\Delta\theta = 5^\circ$, $l = 28b$. In this case, the coupling loss is 1.4 dB. The above procedures can be repeated for DRG's with different a/b , in order to obtain an overall design of V-coupler and DRG that gives minimum L . The minimum coupling loss achievable as a function of a/b (with optimum l , θ , and $\Delta\theta$, and with the desired driving-point impedance) is shown in Fig. 3.10. The designs for the corresponding couplers are given in Table 3.2. The coupling improves substantially as a/b increases from two to four, but improvement slows for $a/b > 4$. As pointed out earlier, increasing a/b tends to decrease the mechanical strength of the waveguide. Considering this tradeoff, $a/b = 4$ appears to be a good choice.

3.4. Conclusion

We have shown that for a good design, a coupling loss on the order of 1.4 dB can be expected between dielectric ridge waveguide and a small device connected at the vertex of a V-coupler. To this must be added dielectric losses and ohmic losses in the V-coupler at the millimeter-wave frequency. For the case of $1000 \Omega\text{cm}$ silicon, one expects dielectric

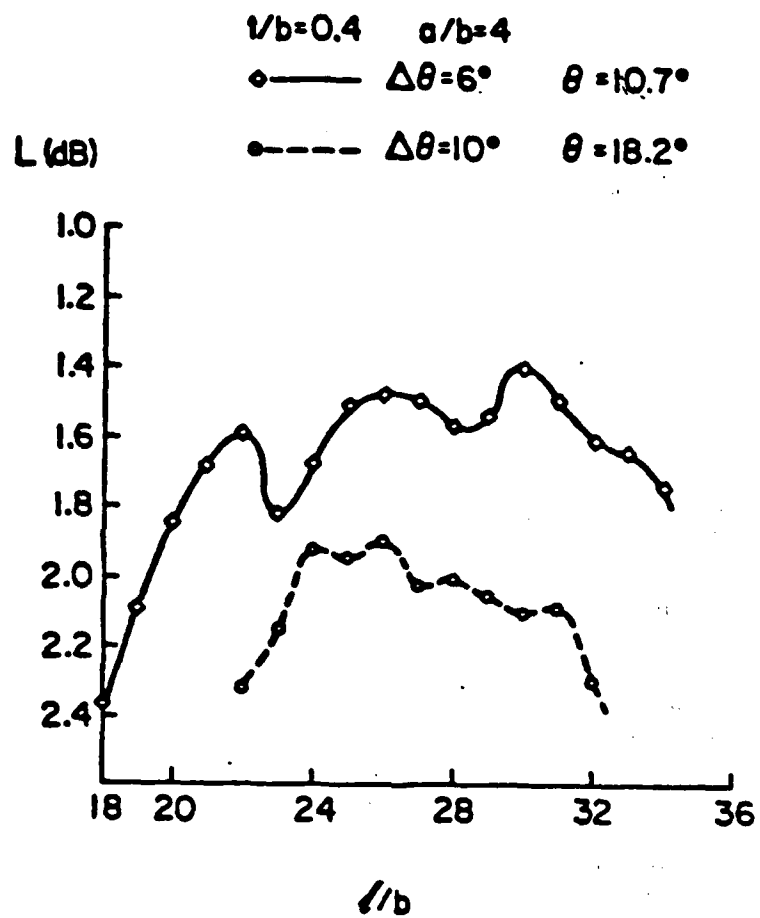


Fig. 3.8

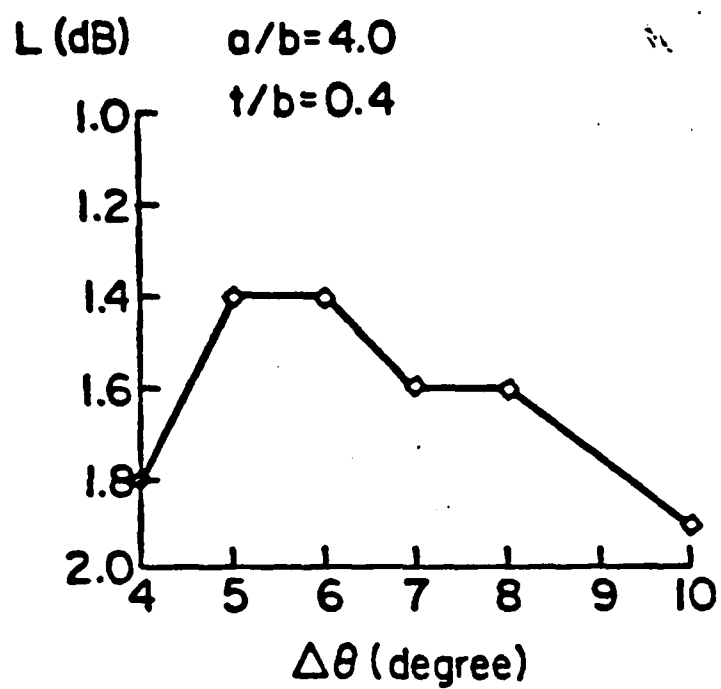


Fig. 3.9

TABLE 3.1

DATA FOR EXPERIMENTAL POINTS SHOWN IN FIG. 3.9

$a/b = 4, \quad t/b = 0.4$			
$\Delta\theta$	θ (deg)	l/b	L (dB)
4	7.1	30	1.8
5	8.9	28	1.4
6	10.7	30	1.4
7	12.5	28	1.6
8	14.4	27	1.6
9	16.2	--	--
10	18.2	26	1.9

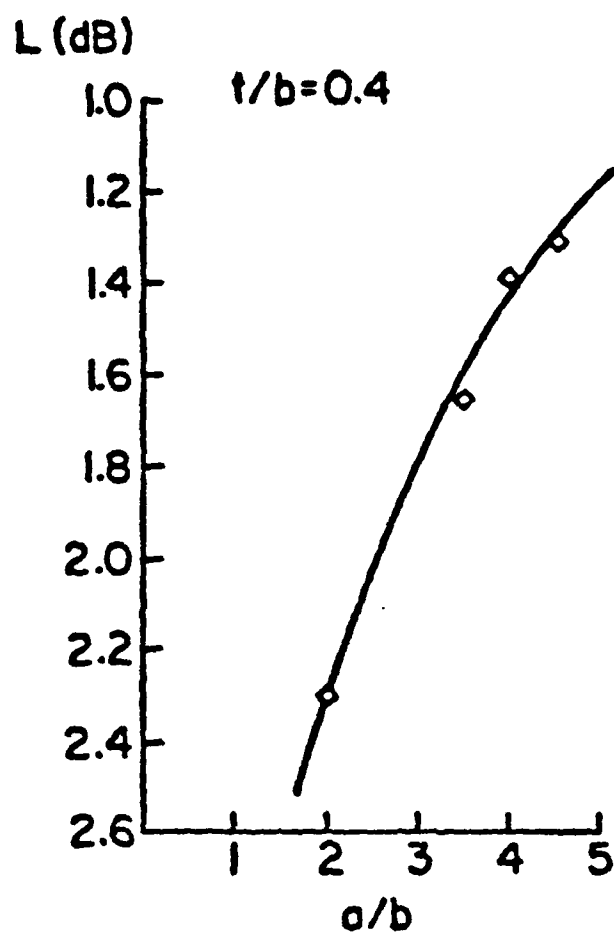


Fig. 3.10

TABLE 3.2

DATA FOR EXPERIMENTAL POINTS SHOWN IN FIG. 3.10

$t/b = 0.4$					
a/b	k	$\Delta\theta$ (deg)	θ (deg)	z/b	L (dB)
2	.707	6°	15°	13	2.3
3.5	.637	5°	8.9°	21	1.6
4	.637	5°	8.9°	28	1.4
4.5	.637	6°	10.7°	29	1.3

loss of the order of 0.5 dB/cm (nearly independent of frequency). In earlier work at 85 GHz, ohmic loss in the V-coupler was found to be about 1.2 dB [9]. Thus, one can expect to couple radiation from free space, through a centimeter of guide, and into a correctly matched device with an overall coupling loss of slightly more than 3 dB.

It is also interesting to compare the unloaded Q of a ring resonator of the DRG and a half-wavelength resonator made of metallic transmission line. The unloaded Q of the ring resonator can be calculated by [10]

$$Q = \frac{k_{z0}}{2\alpha_r} \quad (3.4)$$

where k_{z0} is the unperturbed propagation constant in a straight DRG, and α_r is the attenuation coefficient due to radiation loss. As was noted in Fig. 3.5, $\alpha_r R_c$ in the 10% Bandwidth below f_{21}^x is below 0.25 dB and R_c is roughly equal to 3λ . For a ring resonator of radius of curvature of R_c , the unloaded Q is about 88, which is independent of frequency. Furthermore, the unloaded Q can be made larger by choosing a larger radius of curvature. Since by (2.86), α_r is exponentially proportional to $(-R/R_c)$. For example, the unloaded Q becomes 240 when R is made equal to $2R_c$. The unloaded Q of the half-wavelength resonator made of coplanar waveguide is about 30 at 100 GHz and scales as $1/\sqrt{f}$ because of skin effect [11]. The advantage of the dielectric ridge waveguide over the metallic transmission line is obvious in this respect. The high Q for a ring resonator of small radius of curvature R also makes the channel dropping filter [10] possible. Because the frequency spacing of the ring resonator Δf satisfies the condition

$$\frac{\Delta f}{f} > \frac{1}{Q} \quad (3.5)$$

where

$$\Delta f \approx \frac{c}{2\pi R}$$

(3.6)

For $R = R_c \sim 3 \lambda$, we have

$$\frac{\Delta f}{f} \sim \frac{1}{6} > \frac{1}{Q} \sim \frac{1}{88}$$

The relative low loss and high Q of the dielectric ridge waveguide make it an interesting component for the M^3IC , especially at the higher end of the millimeter-wave spectrum.

References

- [1] T. Wang and S. E. Schwarz, "Design of dielectric ridge waveguides for millimeter-wave integrated circuits," IEEE Trans. Microwave Theory Tech., vol. MTT-31, pp. 128-134, 1983.
- [2] R. M. Knox and P. P. Toullos, "Integrated circuits for millimeter wave through optical frequency range," in Proc. Symposium on Submillimeter Waves, pp. 497-516, Polytechnic Press, Brooklyn, New York, 1970.
- [3] K. Solbach and I. Wolff, "The electromagnetic fields and phase constants of dielectric image lines," IEEE Trans. Microwave Theory Tech., vol. MTT-26, pp. 266-274, 1978.
- [4] J. G. Gallagher, "Mode dispersion of trapezoidal cross-section dielectric optical waveguides by effective-index method," Electron. Lett., vol. 15, pp. 734-735, 1979.
- [5] K. Solbach, "The measurement of radiation losses in dielectric image line bends and the calculation of a minimum acceptable curvature radius," IEEE Microwave Theory Tech., vol. MTT-27, pp. 51-53, 1979.
- [6] D. B. Rutledge, S. E. Schwarz, T-L. Hwang, D. J. Angelakos, K. K. Mei, and S. Yokota, "Antenna and waveguides for far-infrared integrated circuits," IEEE J. Quantum Electronics, vol. QE-16, pp. 508-516, 1980.
- [7] D. B. Rutledge, S. E. Schwarz, and A. T. Adams, "Infrared and submillimeter antennas," Infrared Phys., vol. 18, pp. 713-729, 1978.
- [8] J. Galejs, Antenna in Inhomogeneous Media, Pergamon Press, Oxford, 1969.

- [9] C. Yao, S. E. Schwarz, and B. J. Blumenstock, "Monolithic integration of a dielectric millimeter-wave antenna and mixer diode: an embryonic millimeter-wave IC," IEEE Trans. Microwave Theory Tech., vol. MTT-30, pp. 1241-1247, 1982.
- [10] E. A. J. Marcatili, "Bends in optical dielectric guides," Bell Syst. Tech. J., vol. 48, pp. 2103-2132, 1969.
- [11] N. L. Wang, private communication.

Figure Captions

- 3.1. (a) A DRG fabricated in a semiconductor by anisotropic etching; for silicon, $\alpha = 55^\circ$. (b) Equivalent rectangular DRG used in the EDC method to approximate the trapezoidal DRG.
- 3.2. Experimental setup used in the dispersion measurement.
- 3.3. Comparison of the measured dispersion curves (experimental points) with those predicted by the EDC method (solid line) for DRG with $a/b = 2$. (a) $t/b = 0.4$. (b) $t/b = 0.2$.
- 3.4. Critical radius of curvature R_c from (2.87) for DRG with $a/b = 4$.
- 3.5. Radiation loss per radian for $R = R_c$ at that frequency. The arrows on the horizontal axis indicate the cutoff frequency f_{21}^x of the next higher order mode.
- 3.6. The V-coupler.
- 3.7. Experimental setup used in the V-coupler design. The electronics are the same as in the dispersion measurement.
- 3.8. Typical coupling loss L as a function of coupler length l .
- 3.9. Coupling loss as a function of $\Delta\theta$ (with l always adjusted for lowest loss) for V-couplers on DRG with $a/b = 4$. The values of θ are given in Table 3.1.
- 3.10. The minimum coupling loss achievable as a function of a/b . The values of $\Delta\theta$, θ , and l are given in Table 3.2.

CHAPTER 4

THEORY OF SCHOTTKY DIODE IMAGING ARRAYS

Theoretical aspects of the Schottky diode imaging arrays are presented in this chapter. To fully appreciate the advantages and limitations of the imaging optics used in this work, section 1 is devoted to understanding them physically. Section 2 reviews the fundamentals of metal-semiconductor contacts in order to bring out important design considerations of the diode. Theories of video and heterodyne detection and their noise models are presented in sections 3 and 4, respectively. This sets the stage for the experiments with Schottky diode imaging arrays, to be described in the next chapter.

4.1. Imaging Antenna Arrays

The antenna and imaging optics used in this work are those developed by Rutledge and coworkers [1]-[3]. The imaging optics is the "Reverse-Microscope" optical configuration and the antenna is the modified bow-tie antenna. The image is obtained by plotting the signal received by each antenna.

4.1.1. Imaging Optics

Antennas on dielectric substrates often excite trapped surface waves that reduce the radiation efficiency of the antenna and cause crosstalk between neighboring antennas. This surface-wave crosstalk decays slowly; the power falling off only as $1/r$, characteristic of surface traveling waves. One approach to reduce surface-wave excitation is to make the substrate very thin, but then fabrication becomes difficult.

The other approach to reduce surface-wave excitation is to image through the substrate using a substrate lens, as shown in Fig. 4.1. When the

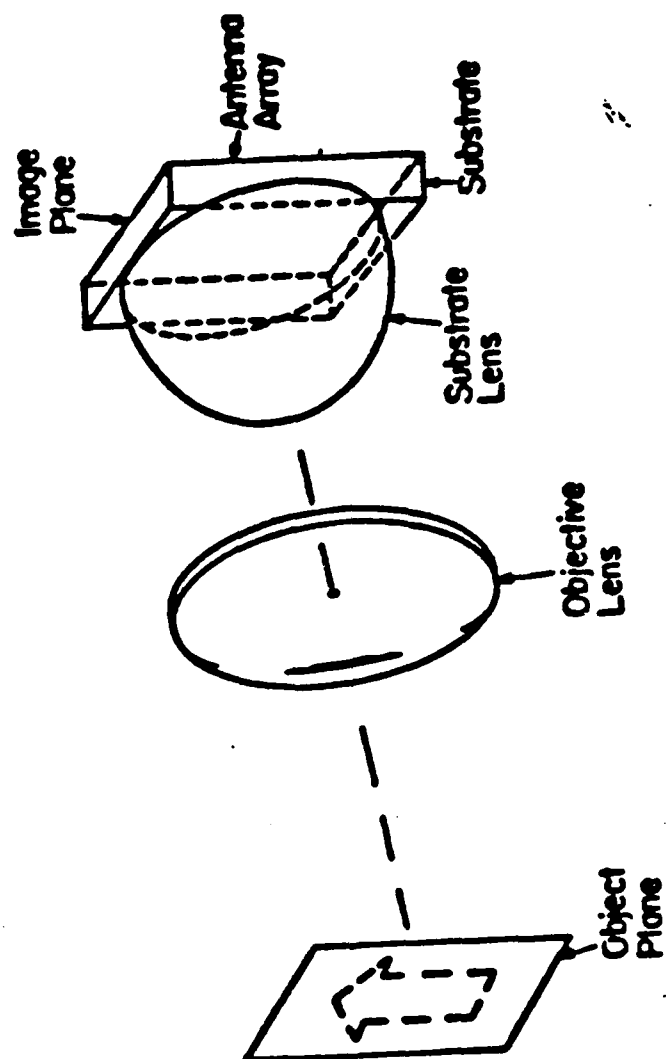


Fig. 4.1

the substrate lens is made of the same material as the substrate, surface-waves can be eliminated completely. Alternatively, the substrate lens material and the thickness of the substrate can be so chosen that, all the surface-modes are below cutoff at the operating frequency [4]. The imaging optics is similar to an oil immersion microscope, except that the position of the object and the image are reversed.

This "reverse microscope" design has several advantages. The trapped surface wave is eliminated because waves transmitted from the antennas are incident at a small angle to the normal of the lens surface. The second advantage is that the optical system can be made nearly aplanatic, with virtually no spherical aberration or coma [5]. The aplanatic arrangement is shown in Fig. 4.2. Finally, this approach integrates well with the modified bow-tie antenna design to be described later, allowing the system to approach diffraction-limited imaging. The disadvantages are those of any optical system with lenses; material and reflection loss.

The material for the substrate lens has to have as little loss as possible. Gallium Arsenide, silicon and fused quartz are good candidates for millimeter-wave applications. GaAs is the ideal choice because, as we shall see, it is the substrate material for the Schottky diode arrays. It is ruled out, however, for bulk GaAs material is very expensive and difficult to machine. Silicon is favored because of its close match of relative dielectric constant to the GaAs substrate; $\epsilon_r = 12$ for silicon and $\epsilon_r = 13$ for GaAs. The substrate thickness at the cutoff of the surface modes is around 850 μm at an operating frequency of 69 GHz. (If fused quartz is used, the GaAs substrate has to be thinned down to less than 100 μm and hence the processing difficulty is increased.) The disadvantage of a silicon substrate lens is higher reflection loss from the lens surface. This, however, can in principle be eliminated through the use of anti-reflection coating.

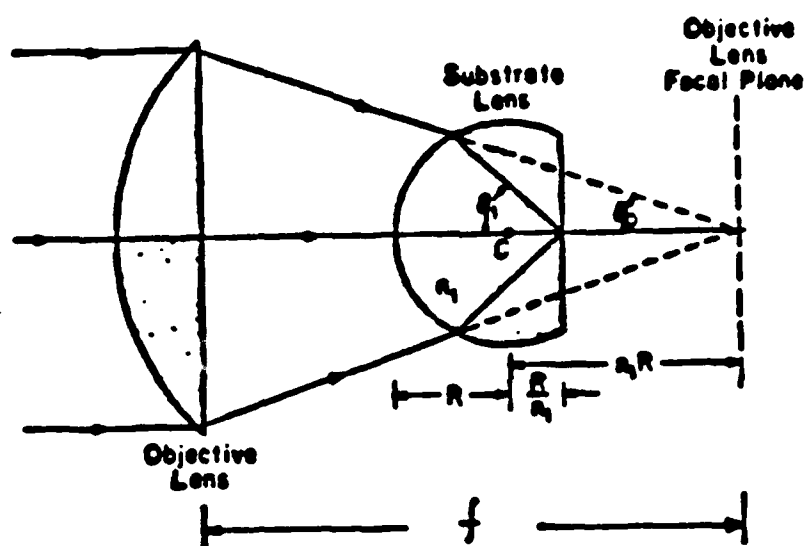


Fig. 4.2

In addition to reflection loss and surface modes, the dielectric constant also affects the antenna feed pattern of the bow-tie antenna. It is not obvious whether silicon or fused quartz is a better choice. Similar work at Caltech uses the quartz substrate lens approach [6].

4.1.2. The Modified Bow-Tie Antenna

At an air-dielectric interface, currents tend to propagate along metals at a velocity that is intermediate between the velocity of light in the air and in the dielectric; the wave is slowed roughly by a factor of $\sqrt{(\epsilon_r+1)}/2$. This means that the current excite evanescent waves in the air. On the other hand, the wave in the dielectric is fast, and energy radiates into the dielectric. By reciprocity theorem, the antenna is more sensitive to radiation coming from the substrate.

The antenna design [2] is shown in Fig. 4.3. The antenna spacing and antenna bow angle are a compromise that give adequate sampling, coupling efficiency, antenna impedance and antenna pattern. The measured antenna patterns are shown in Fig. 4.4 for two dielectric constants. The patterns were measured with substrate lenses of the same material as the substrate itself.

The antenna is fairly broadband. In fact, if the bow were infinite in extent, its impedance would be strictly independent of frequency. In practice, it is observed that the impedance is essentially independent of frequency if each bow arm is more than a free-space wave-length long. The measured frequency-independent impedances as a function of the bow angle θ are plotted in Fig. 4.5 for two dielectric constants. This impedance can be calculated accurately from transmission-line theory, using the static capacitance and inductance of the bow tie. The result is

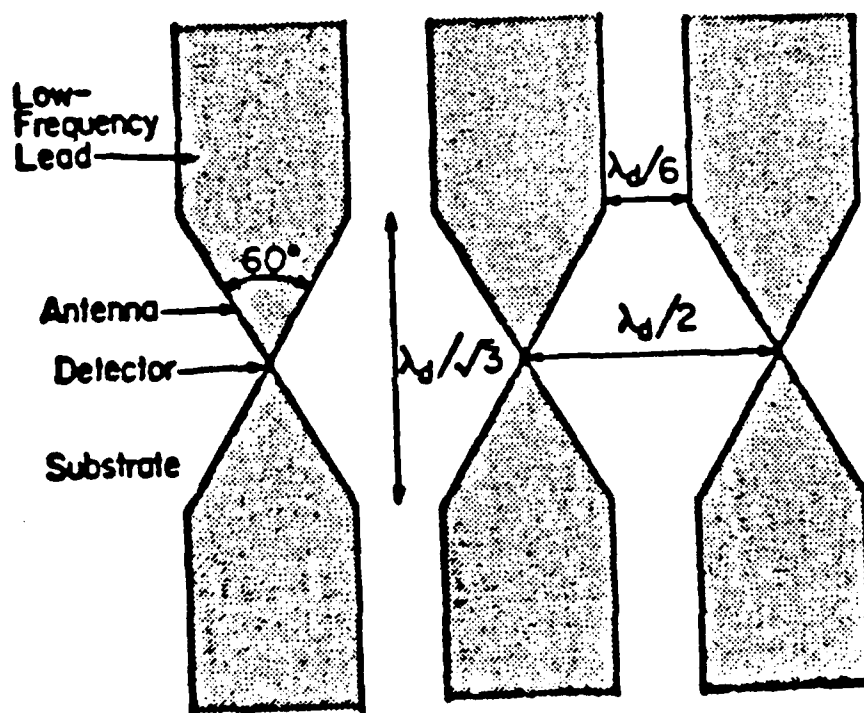
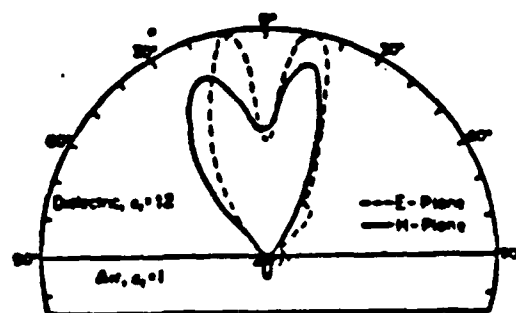
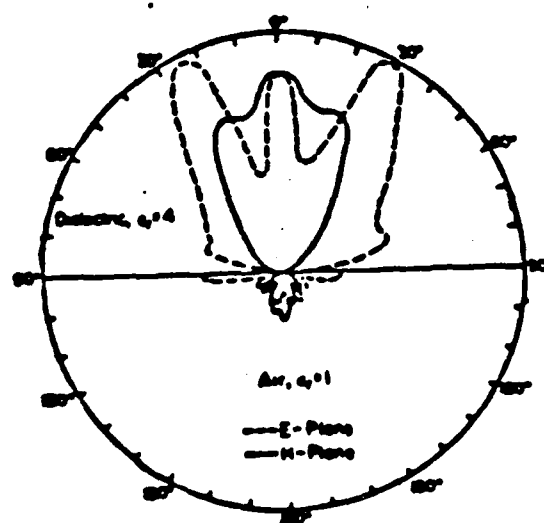


Fig. 4.3



(a)



(b)

Fig. 4.4

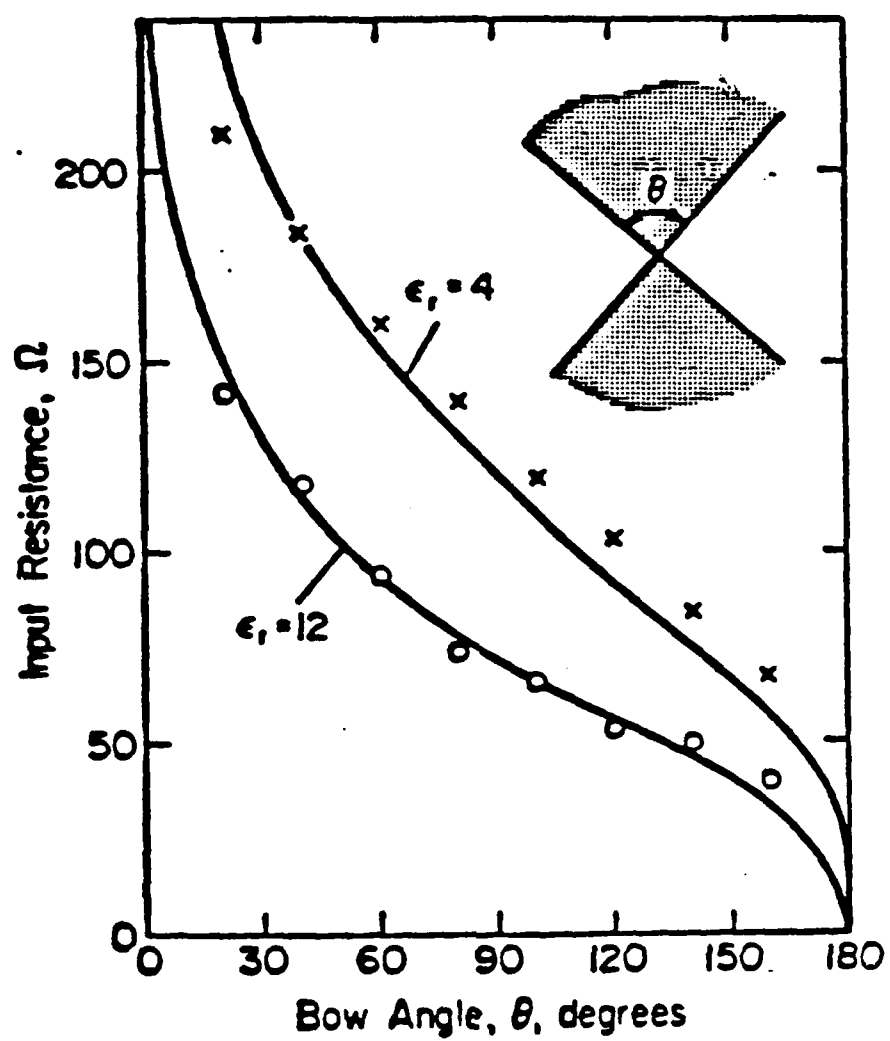


Fig. 4.5

$$z = \frac{\sqrt{\frac{\mu_0}{\epsilon_0}}}{\sqrt{\frac{\epsilon_r + 1}{2}}} \cdot \frac{K(k)}{K(\sqrt{1-k^2})} \quad (4.1)$$

where K is the complete elliptic integral of the first kind and $k = \tan^2(45^\circ - \theta/4)$.

4.1.3. Diffraction-Limited Imaging

An important characteristic of imaging systems is their behavior as low pass filters; that is, no matter how high the spatial frequency components of an object, because of diffraction, an image of the object will contain only components less than some system cutoff frequency f_0 . This is easily understood by the Fourier analysis of the optical system [8].

When a plane wave is incident normally on a simple convex lens of diameter D and focal length f , the cutoff spatial frequency f_0 of field distribution is

$$\begin{aligned} f_0 &= \frac{K_x}{2\pi} \\ &= \frac{K_0}{2\pi} \sin \theta_0 \sim \frac{K_0}{2} \cdot \frac{D/2}{f} \\ &= \frac{D}{2\lambda f} \end{aligned} \quad (4.2)$$

where λ and K are the free space wavelength and wavenumber, respectively. The geometry is depicted by the extended dash line in Fig. 4.2.

The diffraction spot size Δr_0 is defined as the radius of the first dark ring in the Airy pattern. It is related to the spatial frequency cutoff by

$$\Delta r_0 = 0.61 \frac{1}{f_0} \quad (4.3)$$

A convenient number to characterize a imaging system is its f -number, $f^\#$, defined by

$$f^\# \equiv \frac{f}{D} = \frac{1}{2 \sin \theta_0} \quad (4.4)$$

It is especially convenient to use for a complex imaging system, such as the reverse microscope optics.

The diffraction-limited spot size can be expressed in terms of the f -number of the lens

$$\Delta r_0 = 1.22 \lambda f^\# \quad (4.5)$$

Similarly, the spatial frequency cutoff f_0 can be written as

$$f_0 = \frac{1}{2\lambda f^\#} \quad (4.6)$$

In the reverse microscope optics, the dimensions and separations are chosen so that the aplanatic condition is satisfied. Under this condition, a simple relation exists for arbitrary angle θ_0 :

$$n_0 \sin \theta_1 = n_1 \sin \theta_0 \quad (4.7)$$

where $n_0 = 1$ for air, and n_1 is the index of refraction of the substrate lens. The system f -number $f_s^\#$ can be defined in analogy to (4.4),

$$f_s^\# = \frac{1}{2 \sin \theta_1} \quad (4.8)$$

Combining (4.7) and (4.8), we have

$$f_s^\# = f^\# / n_1 \quad (4.9)$$

The system spatial frequency cutoff f_s , and the diffraction-limited spot size of the imaging system Δr_s , can now be obtained

$$\begin{aligned} f_s &= \frac{1}{2(\frac{\lambda}{n_1})f_s^\#} \\ &= \frac{n_1^2}{2\lambda f_s^\#} = n_1^2 f_0 \end{aligned} \quad (4.10)$$

and

$$\begin{aligned} \Delta r_s &= 1.22 \frac{\lambda}{n_1} f_s^\# \\ &= 1.22 \frac{\lambda f_s^\#}{n_1^2} = \frac{\Delta r_0}{n_1^2} \end{aligned} \quad (4.11)$$

For the reverse-microscope optics, the two lens combination has the spatial frequency cutoff increased by a factor of n_1^2 and the diffraction limited spot size decreased by the same factor.

Knowing the spatial frequency cutoff f_s for field distribution in the image plane, we can determine the appropriate sampling interval T_E to reconstruct the image by field detection through the Whittaker-Shannon sampling theorem [8]

$$\begin{aligned} T_E &= \frac{1}{2} \frac{1}{f_s} \\ &= \frac{\lambda f_s^\#}{n_1^2} \end{aligned} \quad (4.12)$$

recalling that $f_s^\#$ is the f -number of the objective lens. For imaging optics with $f_s^\# = 2$ for the objective lens, the field sampling interval $T_E = 669 \mu\text{m}$ at 69 GHz. The antenna spacing actually implemented in the imaging array is $660 \mu\text{m}$. If intensity detection, instead of field detection, is used to reconstruct the image, the appropriate sampling

AD-A166 576

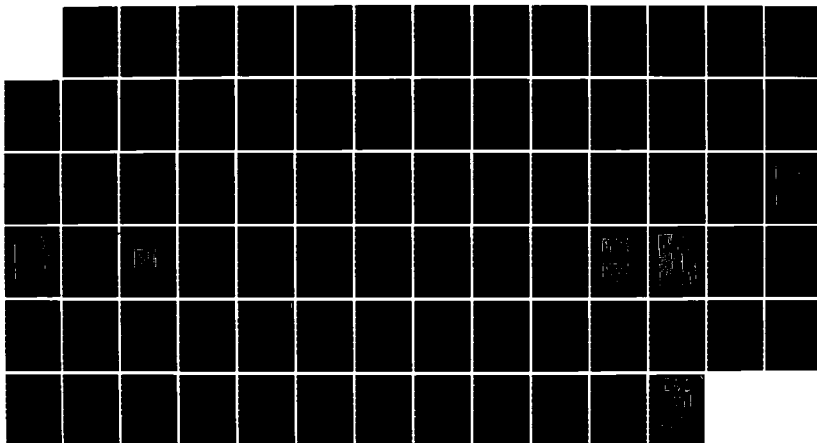
RESEARCH ON INTEGRATED NEAR-MILLIMETER-WAVE RECEIVER
(U) CALIFORNIA UNIV BERKELEY ELECTRONICS RESEARCH LAB
S E SCHWARZ 01 OCT 82 DAAK70-80-C-0134

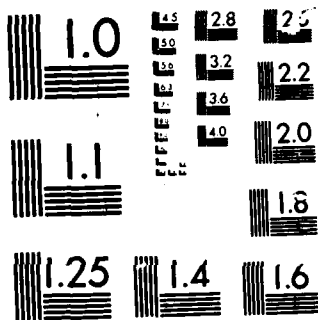
2/2

UNCLASSIFIED

F/G 9/5

NL





MICROCOPY

CHART

interval I_I is half of the T_E

$$T_I = \frac{f^\#}{2n_1^2} \quad (4.13)$$

There is another important consideration for the actual spatial frequency cutoff of the antenna imaging array. In the discussion so far, we have tacitly assumed that the antenna responds equally to radiation incident at all angles. This is certainly not true, as manifested by the antenna pattern. In order to fully realize the potential of the imaging optics, the ideal antenna for this application should have the main lobe of the antenna pattern matched to the angle θ_1 of the imaging optics.

Even if the imaging system is used to measure the power emitted from a point source, such as in radio astronomy, matching of the field pattern is essential to ensure capturing the total incident power while avoiding background noise pickup.

4.2. Schottky Contacts

With regard to the diode design, the most important properties of the diode are the ideality factor, junction capacitance, and junction breakdown voltage. In order to calculate the junction capacitance and breakdown voltage, we have to know the diode barrier height; which is covered in the first part of this section. Junction capacitance and junction breakdown are next briefly reviewed. Current conduction mechanisms are then reviewed in order to understand their effects on the diode ideality factor. Finally, the noise mechanisms of the diode are presented.

4.2.1. Barrier Height

The theory of Schottky barrier formation is still incomplete; the barrier height still cannot be determined from first principles. The purpose of the present subsection is to qualitatively summarize the current theories. An excellent review by Schluter has appeared recently [9].

When a metal and a semiconductor are brought into contact, their individual Fermi levels will adjust to be the same on both sides of the interface. The adjustment occurs on a local microscopic scale (0.1-1 nm range) near the interface through a build up a small interfacial dipole Δ . The bulk semiconductor, which has a well-defined Fermi level, adjusts in turn by long range (100-1000 nm) band bending. The situation is depicted in Fig. 4.6 for an n type semiconductor. The rectifying barrier ϕ_B is the energy necessary to transport a metal electron into the semiconductor.

In the simplest model of barrier height formation, due to Schottky, the metal and semiconductor are assumed to be in equilibrium but no direct interaction or charge flow at the contact and therefore no interface dipole ($\Delta=0$) is formed. As shown in Fig. 4.6 the barrier height ϕ_B is then just the difference between the metal work function ϕ_M and the semiconductor electron affinity χ_s , i.e., $\phi_B = \phi_M - \chi_s$. It was soon recognized that this model was too simple to describe a number of experimental findings. In particular, it was found that the barrier heights for many metals in contact with silicon were practically independent of the metal work function. This lead Bardeen to propose a model in which surface states, located energetically in the semiconductor gap, play a central role. Bardeen recognized that a relatively low density of surface states (about one per 1000 surface atoms) on the free

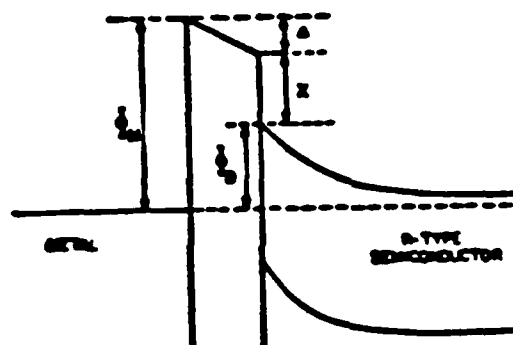


Fig. 4.6

surface would be sufficient to "pin" energetically the Fermi level at the surface and thus to make the work function independent of the doping in the interior as had been observed. In 1965 Heine [10] commented on the nature of these interface states. Bardeen had introduced it in the spirit of surface states on a free semiconductor surface and was not specific as to what changes would occur in the case of an intimate metal-semiconductor contact. Heine pointed out that, strictly speaking, localized states cannot exist at such a junction because of the coupling to a continuum of free-electron-like states on the metal side. He argued, however, that these metal wavefunctions will decay into the semiconductor and their tails will play the role of Bardeen's surface states. The typical decay length was estimated to be of 0.1-1 nm for free-electron-like metals in contact with silicon.

The key assumption of weak interaction, i.e., the neglect of strong chemical bonding across the interface, has been challenged for its general validity. Alternative models stressing the importance of chemical bonds was first put forward by Phillips and coworkers [11], [12]. They proposed a simple but general model in which one or more bulk-like compounds are formed between the semiconductor and the metal which in turn influence the barrier height. A similar thermochemical analysis of Schottky barrier behavior has been done by Brillson [13]. He found a strong transition for barrier height between reactive (heat of formation $\Delta H_f < 0$) and non-reactive ($\Delta H_f > 0$) junctions. While these analyses do not produce any explicit microscopic picture for Fermi level pinning and barrier formation, they clearly indicate the need for a theory able to distinguish between reactive and non-reactive mechanisms which goes beyond the Bardeen-Heine Model.

A particular mechanism of Fermi level pinning has been promoted by Spicer recently [14]. The concept is derived from attempting to interpret consistently a number of different experimental observations.

(a) Photoemission studies (in agreement with most recent theoretical calculations) show no intrinsic surface state in the gap of clean III-V semiconductors.

(b) Direct measurement of the surface Fermi level position indicates that complete Fermi level pinning can be reached with much less than one monolayer metal coverage.

(c) The final pinning position appears to be independent of the type of adatom, provided that reactive metals are used.

The absence of intrinsic surface states and the low coverage saturation rule out the Bardeen-Heine model as an adequate description and suggest that chemical bonding is important. The insensitivity of the Fermi level position to the chemical nature of the adatom led Spicer to conclude that the Fermi level pinning was caused by secondary induced electronic states, e.g., defect level near the surface. A relatively low number of new defects (about one per 100 surface atoms) necessary to pin the Fermi level electrostatically could be created using the heat of condensation of surface adatoms.

It is clear from all these studies that a fundamental understanding on a microscopic basis of the formation of a Schottky barrier has not been reached yet. For the purpose of our device application, it suffices to say that, for a Schottky contact on GaAs, ϕ_B is generally around 0.85 eV.

4.2.2. Junction Capacitance and Junction Breakdown [15]

Once the Schottky barrier height ϕ_B and the doping concentration N_D (hence the Fermi level E_F) in the bulk semiconductor are known, the

depletion width and the junction capacitance can be obtained by solving the Poisson equation in the semiconductor. Under the abrupt approximation that the charge density $\rho = qN_D$ in the space charge region, $0 \leq X \leq W$, and $\rho = 0$, and $dv/dx = 0$ for $X > W$, where W is the depletion width, we obtain

$$W = \left[\frac{2\epsilon}{qN_D} (V_{bi} - V - \frac{KT}{q}) \right]$$

$$|E(x)| = \frac{qN_D}{\epsilon} (W-x) = E_m - \frac{qN_D}{\epsilon} x$$

$$V(x) = \frac{qN_D}{\epsilon} (WX - \frac{1}{2} x^2) - \phi_B$$

where q is the magnitude of electron charge, ϵ is the dielectric constant of the semiconductor, K is the Boltzmann constant, T is the absolute temperature in degree Kelvin, and the term KT/q arises from the contribution of the majority-carrier distribution tail, $V_{bi} = \phi_B - \frac{E_f}{q}$ is the build-in potential for electrons in the semiconductor conduction band, E_m is the maximum field strength which occurs at the metal-semiconductor interface, at $X = 0$.

$$E_m = \left[\frac{2qN_D}{\epsilon} (V_{bi} - V - \frac{KT}{q}) \right]^{1/2}$$

$$= \frac{2(V_{bi} - V - KT/q)}{W}$$

The space charge Q_{sc} per unit area of the semiconductor and the depletion-layer capacitance C per unit area are given by

$$Q_{sc} = qN_D W = \left[2q\epsilon N_D (V_{bi} - V - \frac{KT}{q}) \right]^{1/2}$$

$$C \equiv \frac{|\partial Q_{sc}|}{\partial V} = \left[\frac{q\epsilon N_D}{2(V_{bi} - V - KT/q)} \right]^{1/2} = \frac{\epsilon}{W} \quad (4.14)$$

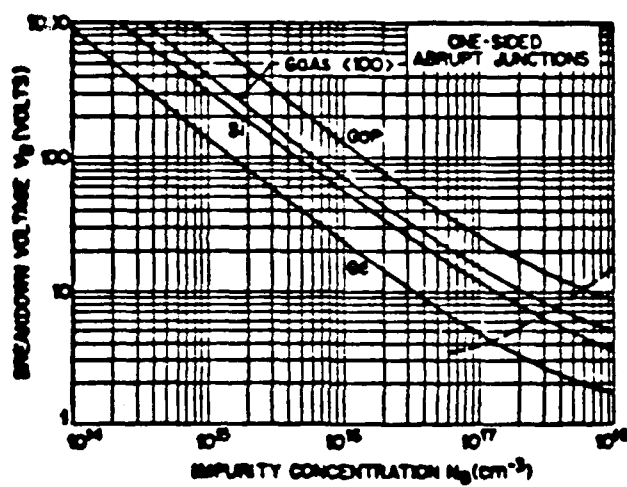


Fig. 4.7

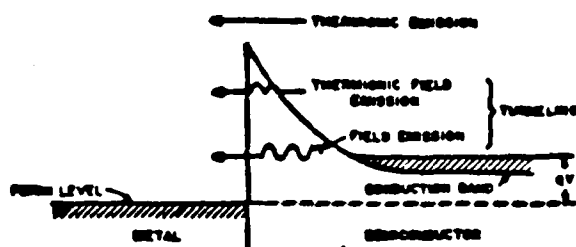


Fig. 4.8

h is Planck constant, and m^* is the effective mass of the semiconductor. For field emission case, i.e., high N_D and/or low T , $\text{Coth}(E_{00}/KT) \rightarrow 1$ and

$$V_0 = E_{00}/q$$

which is temperature independent. For the thermionic emission case, i.e., high T and/or low N_D , and

$$V_0 = KT/q$$

Often, the diode I-V characteristics is written ($I \gg I_s$) as

$$I = I_s \exp\left[\frac{qV}{nKT}\right] \quad (4.18)$$

where n is called the ideality factor. For pure thermionic emission, $n = 1$, and the deviation from unity may be used as a measure of the relative contribution of tunneling to conduction. At room temperature, deviation from purely thermionic emission occurs in GaAs at doping densities as low as $(1-2) \times 10^{17}/\text{cm}^3$. Thus for doping densities higher than this, a value of n greater than unity can be expected.

In the thermionic emission theory, the saturation current density J_s can be related to the semiconductor material properties and the known barrier height. For semiconductors with isotropic effective mass in the lowest minimum of the conduction band such as n-type GaAs, J_s can be written as

$$\begin{aligned} J_s &= \left(\frac{4\pi q m^* K^2}{h^3} \right) T^2 \exp\left[-\frac{q\phi_B}{KT}\right] \\ &= A^* T^2 \exp\left[-\frac{q\phi_B}{KT}\right] \end{aligned} \quad (4.19)$$

where A^* is related to the Richardson constant A for free electron by

$A^*/A = m^*/m_0$, m_0 is the free electron mass, and $A = 120 \text{ A/cm}^2/\text{K}^2$. For multiple-valley semiconductors, the appropriate Richardson constant A^* associated with a single energy minimum is given by

$$\frac{A_1^*}{A} = \frac{1}{m_0} (\ell_1^2 m_1^* m_2^* + \ell_2^2 m_2^* m_x^* + \ell_3^2 m_x^* m_y^*)^{1/2} \quad (4.20)$$

where ℓ_1 , ℓ_2 , and ℓ_3 are the direction cosines of the normal to the emitting plane relative to the principal axes of the ellipsoid, and m_x^* , m_y^* , and m_z^* are the components of the effective mass tensor. For [100] GaAs, A^*/A is 0.068.

4.2.4. Noise Mechanisms

There are several noise sources found in a Schottky diode [23], [41]. Thermal (Johnson) noise is always present even at zero bias. When the diode is conducting, shot noise and flicker (1/f) noise appear. As the current density through the diode increases, four more noise mechanisms appear. They are related to excess fluctuation of electron velocities and number, the formation of microclusters at the junction, and the graininess of the junction caused by relatively small number of dopants in the thin epi-layer. Thermal noise, shot noise, and the flicker noise can be characterized quantitatively, as is done in the following. The excess noise at high current density, however, cannot be quantitatively characterized yet.

Thermal noise is due to the random thermal motion of the electrons and is unaffected by the presence or absence of direct current, since typical electron drift velocities in a conductor are much less than electron thermal velocities. In a resistor R , the equivalent noise voltage source can be written to be

$$\overline{V_{N,th}^2} = 4KTBR \quad (4.21)$$

where $\overline{V_{n,th}^2}$ is the mean square value of the open circuit noise voltage, and B is the bandwidth. The maximum available noise power to a matched load is

$$P_n = KTB$$

Equation (4.21) shows that the noise spectral density is independent of angular frequency ω , and for thermal noise, this is true so long as $\hbar\omega \ll KT$.

Shot noise is due to the fact that electrical charge is not continuous but is carried in discrete amounts equal to the electron charge. The mean square noise current $\overline{I_s^2}$ of a Schottky diode is related to the total current passing through it by

$$\overline{I_s^2} = 2qB(|I_+| + |I_-|) \quad (4.22)$$

where $|I_+|$ and $|I_-|$ are the magnitude of the forward and reverse current, respectively. Again, the noise spectral density is independent of frequency. Under normal operating conditions, the junction current I is

$$I = |I_+| - |I_-| \approx |I_+|$$

Therefore

$$\overline{I_s^2} = 2qIB \quad (4.23)$$

The shot noise current source can be represented by a Thévenin equivalent circuit, with the mean square voltage source given by

$$\begin{aligned}
 \overline{V_s^2} &= \overline{I_s^2} R_j^2 \\
 &= 2qIB \left(\frac{nKT}{qI} \right)^2 \\
 &= 2KT B n R_j
 \end{aligned}
 \tag{4.24}$$

Compared with (4.21), it is half of the noise for a resistor of the same resistance.

Flicker noise is always associated with a flow of direct current. The source of it is not yet well understood, and has been attributed to a combination of factors [23], [24], including surface effects resulting from fluctuation in the occupation of slow surface states, local breakdown, and a shifting of the Fermi level associated with dislocations at the surface. The mean square noise voltage source per unit bandwidth of the diode, $\overline{V_f^2}$, is given by

$$\overline{V_f^2} = 2KT B \left(\frac{f_0}{f} \right)^\alpha R_j
 \tag{4.25}$$

where α is a number close to unity (and hence the alternative name "1/f noise"). Because of the 1/f dependence on frequency, it is the dominant noise mechanism at lower frequencies.

High current density at the junction introduce excess noise through fluctuations in electron number and their velocities. The principal mechanisms producing the excess velocity fluctuations are: (1) generation of hot electrons whose energies exceed significantly the thermal equilibrium value; and (2) local heating of the junction caused by high current densities. The excess fluctuations of the number of electrons are attributed to: (1) intervalley scattering, i.e., hot electrons are injected into different region of the conduction band where they

become nearly immobile because of their high effective mass; and (2) trapping of electrons in the undepleted epilayer and in the vicinity of the junction.

4.3. Video Detection

The theory and equivalent circuit representation of video detection is presented first. Noise generated by the detector itself places a lower limit on the minimum detectable power. We shall combine the theory of video detection and noise mechanisms in the diode to derive the noise equivalent power NEP.

4.3.1. Theory of Video Detection

Video detection consisting of measuring the averaged rectified voltage of the incoming r.f. signal by forward-biasing the diode. Because there is only one small r.f. signal involved, we can analyze the detection processes by a Taylor expansion of the DC I-V characteristics about the DC bias point V_b .

$$I(V_b + \delta V) = I(V_0) + I'(V_0)\delta V + \frac{1}{2!} I''(V_0)(\delta V)^2 + \dots \quad (4.26)$$

where

$$I'(V_0) = \left. \frac{\partial I}{\partial V} \right|_{V_b} \quad \text{and} \quad I''(V_0) = \left. \frac{\partial^2 I}{\partial V^2} \right|_{V_b}$$

and so on. For a Schottky diode with DC I-V characteristics given by (4.18). $I'(V_b)$ and $I''(V_b)$ can be expressed as

$$\begin{aligned} I'(V_b) &= \frac{1}{R_j} \\ &= \frac{I(V_b)}{nKT/q} \end{aligned} \quad (4.27)$$

$$\text{and } I'' = \frac{I(V_b)}{(nKT/q)^2} \quad (4.28)$$

where R_j is the incremental junction resistance. The first term in (4.26) is the DC bias current, the second term averages to zero for any periodic signal. The detected video signal is the term represented by the third term

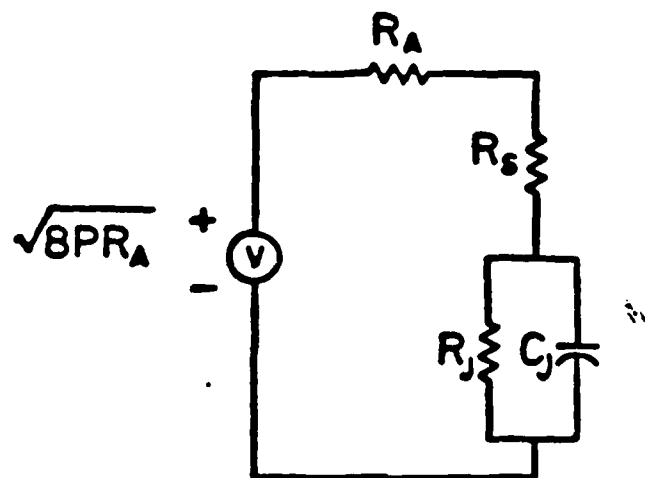
$$i_s(V_b, \delta V) = \frac{1}{2} I''(V_b) (\delta V)^2 \quad (4.29)$$

because $\delta V = V_j \cos(\omega t + \theta)$, where V_j is the r.f. voltage across the junction, we have

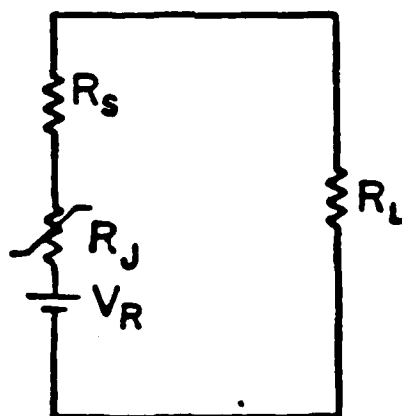
$$\begin{aligned} i_s(V_b, \delta V) &= \frac{1}{2} I''(V_b) \frac{1}{2} |V_j|^2 \\ &= \frac{1}{4} I''(V_b) |V_j|^2 \end{aligned} \quad (4.30)$$

The situation can be modeled by the equivalent circuit model shown in Fig. 4.9. The incoming power P is represented by a voltage source of peak voltage $V_s = \sqrt{8PR_A}$, where R_A is the antenna source impedance. The rectified voltage (or current) is detected across a measuring equipment with load impedance R_L .

The response of a video detector is characterized by a voltage (or current) responsivity, $R_V(R_I)$. It is defined as the ratio of the rectified voltage (current) to the power available to the diode. For an imaging system, however, it is more appropriate to define the responsivities in terms of the power incident on the objective lens of the system, because inevitably there is always some optical coupling loss, $L_{op} = P_{\text{incident}}/P_{\text{received}}$, associated with the imaging optics.



(a)



(b)

Fig. 4.9

The responsivities defined in this way correspond to the actual measured value.

Before we go forward to discuss the theoretical responsivities, it is useful to define a "intrinsic current responsivity" of the diode as the ratio of short-circuited rectified current to the power available to an ideal diode,

$$R_I^0 = \frac{\frac{1}{4} I''(V_b) |V_j|^2}{\frac{1}{2} |V_j|^2 I'(V_b)} = \frac{1}{2} \frac{q}{kT} \quad (4.31)$$

$$= 19.31 \text{ A/W at } 300^\circ\text{K.}$$

By the equivalent circuit model, we can express the theoretical voltage responsivity as

$$R_V = \frac{1}{L_{op}} \cdot \frac{R_j q}{2nKT} \cdot \left[\frac{4R_A R_j X_j^2}{X_j^2 (R_j + R_s + R_A)^2 + R_j^2 (R_s + R_A)^2} \right] \cdot \left[\frac{R_L}{R_s + R_j + R_L} \right] \quad (4.32)$$

where $X_j = 1/\omega C_j(V_b)$ and ω is the angular frequency of the r.f. signal. The junction capacitance as a function of DC bias voltage is given by (4.14). The term in the first bracket represents the mismatch and parasitic loss at the signal frequency. The second term is just the voltage divider for low frequency measurement. The junction incremental resistance R_j decreases with increasing bias, while the junction capacitance C_j increase with increasing bias. The voltage responsivity peaks at a certain bias, beyond which it rolls off due to larger capacitance which shunts the signal across the junction. Below the optimum bias, the observed detected voltage drops because of a larger voltage drop across the junction resistance (the term represented by the second bracket in (4.32)). It is obvious from (4.32) that a diode with small

junction capacitance, ideality factor and series resistance is required for good detector performance. As we shall see in the following, the ideality factor and the series resistance also affect the noise properties of the detector. In order to minimize series resistance, Gallium Arsenide was chosen for its low field electron mobility, which is about 6 times higher than that of silicon.

4.3.2. Noise-Equivalent Power

The presence of noise generated by the diode itself limits the minimum amount of power that can be detected. The noise property of the diode is best characterized by the noise equivalent power, NEP. It is defined as the power required to generate a rectified voltage equal to the noise root-mean-square voltage $\sqrt{V_m^2}$ per root H_z bandwidth measured across the load resistor

$$\sqrt{\frac{V_m^2}{B}} = NEP \cdot R_V$$

That is

$$NEP = \frac{\sqrt{V_m^2}}{R_V \sqrt{B}} [W/H_z^{1/2}] \quad (4.33)$$

Knowing the low frequency equivalent circuit (Fig. 4.9), we can calculate $\sqrt{V_m^2}$ from a known noise voltage source with $V_s = [V_N^2]^{1/2}$ or vice versa. Notice that the NEP is inversely proportional to the voltage responsivity.

In video detection, the optimum responsivity occurs at a relatively low bias current density. The incremental junction resistance R_j is always much greater than the series resistance. Therefore, only the shot noise and the flicker noise need to be considered. So the total noise

voltage source of the diode $\overline{V_N^2}$, can be written as

$$\begin{aligned}\overline{V_N^2} &= \overline{V_s^2} + \overline{V_f^2} \\ &= 2KTB R_j \left[1 + \left(\frac{f_0}{f} \right)^\alpha \right]\end{aligned}\quad (4.34)$$

where f_0 is the cross-over frequency at which the shot noise equals to the flicker noise. By the low frequency equivalent circuit, we have

$$\overline{V_m^2} = \overline{V_N^2} \left[\frac{R_L}{R_s + R_j + R_L} \right]$$

Therefore, the theoretical NEP is

$$\text{NEP} = \frac{\sqrt{2KTR_j} \left[1 + \left(\frac{f_0}{f} \right)^\alpha \right]^{1/2} \left[\frac{R_L}{R_s + R_j + R_L} \right]}{R_v}\quad (4.35)$$

4.4. Heterodyne Detection

When two signals of different frequencies are applied across a Schottky diode, the sum and difference frequencies (the Intermediate Frequency, IF) of the applied signals are generated by virtue of its nonlinearity. The process is called mixing. These new frequencies will further beat with the originally applied signals and with each other to create a rather complex waveform.

Consider a signal $V_s \cos \omega_s t$ and a pump signal (the local oscillator, LO) $V_p \cos \omega_p t$, applied across the Schottky diode whose I-V characteristic is described by (4.18). The quadratic term alone in (4.26) can produce frequency components other than ω_s and ω_p :

$$\begin{aligned}
 i(t) \propto (V_s \cos \omega_s t + V_p \cos \omega_p t)^2 &= \frac{1}{2} (V_s^2 + V_p^2) + V_s V_p \cos(\omega_s - \omega_p)t \\
 &+ V_s V_p \cos(\omega_s + \omega_p)t + \frac{1}{2} V_s^2 \cos \omega_s t \\
 &+ \frac{1}{2} V_p^2 \cos 2\omega_p t
 \end{aligned}$$

The intermediate frequency $\omega_0 = \omega_s - \omega_p$ is usually much lower than the signal frequency ω_s (but sufficiently high to avoid the excess $1/f$ noise), and can be easily amplified and thus detected at a higher signal level. This detection scheme is called heterodyne detection. In the small-signal analysis, the IF term in (4.26) is proportional to the voltage of the signal; this is in contrast to the case of video detection where the rectified current is proportional to the power of the signal. The IF term is also proportional to the pump voltage. It is therefore possible to increase the IF signal and hence the sensitivity of the receiver by increasing the LO power.

One important figure of merit for a mixer is its conversion loss, L , which is defined as the ratio of available power at the signal frequency to the IF power delivered to the IF amplifier. Another important figure of merit is its noise equivalent power NEP, defined as the signal power per unit IF bandwidth required to generate the noise power per unit IF bandwidth measured at the output of the IF amplifier.

4.4.1. Theory of Mixing

The basic principles of mixing using crystal diodes were reviewed in depth by Torrey and Whitmer in 1948 [25]. In their classical frequency domain approach, the assumption of a sinusoidal local oscillator voltage was used, with the implicit assumption that the diode was short-circuited at all harmonics of the LO. The diode is then modeled as a

linear, sinusoidally time-varying resistance, whose Fourier coefficients at the signal, the LO and the image ($\omega_i = \omega_p + \omega_0$) frequencies forms a conversion matrix. The conversion matrix can be used to calculate the conversion loss.

The effects of parasitic series resistance and capacitance (assumed constant) of the mixer diode were investigated using approximate methods by many authors [26]-[29]. A new intuitive approach to mixer analysis was taken by Barber [30], who approximated the diode by a switch whose pulse-duty ratio determined the conversion properties of the mixer.

Dickens et al. [31] modified Barber's approach to include the effects of series resistance and junction capacitance and showed that it agreed with experiments reasonably well. All these analyses, however, required simplifying assumptions about the termination of higher order sidebands and ignored the effects of nonlinear diode capacitance. Saleh [32] in 1971 studied the effects of local oscillator waveforms and higher order sideband terminations on resistive mixers, and demonstrated that these characteristics must be considered if an accurate analysis is to be made.

In the general mixer analysis of Held and Kerr [33], arbitrary embedding impedances are allowed at the harmonics of the LO and at all sideband frequencies, and any diode capacitance law can be assumed. Large-signal nonlinear analysis of the LO waveforms of the pumped diode were solved by a combined time-domain and frequency-domain technique. The intrinsic diode capacitance and conductance are characterized in the time-domain, while the diode series resistance and the embedding impedance are linear and are represented in the frequency domain. After determining the LO waveforms at the diode, the small-signal conversion matrix similar to Torrey and Whitmer's approach that contained all the significant sidebands was constructed for the diode. The conversion loss could then

be determined using the conversion matrix.

The time-domain approach by Heine [34] is an extension of the incremental method of mixer characterization described by Torrey and Whitmer. The method applied only to mixers in which the IF is a small fraction of the LO and signal frequencies. The analysis involves first-order perturbation. Mixer performance is predicted using the scattering matrix of the mixer network which is derived from a set of single-frequency measurements under LO drive at various bias levels. It has been demonstrated that the time-domain approach and the frequency-domain approach are identical in the regime of its validity.

For purposes of estimating conversion loss of the imaging array, we found Dickens [31] treatment most convenient to use. It has been shown that most mixer diodes (adjusted for lowest conversion loss) behave as though the barrier itself were switched ON and OFF at the LO frequency, and that the resistance in the ON state is just that of the limiting series resistance R_s , and the impedance in the OFF state is just that expected of the series resistance R_s in series with the barrier capacitance C_j . Although the barrier capacitance with measured results are obtained if the capacitance value at the DC bias voltage is used. Thus the frequency cutoff $f_{c0} = [2\pi R_s C_j (V_b)]^{-1}$.

Knowing the pulse duty ratio t and the cutoff frequency f_{c0} , the intrinsic mixer conversion loss L can be read from Fig. 4.10(a) for a double sideband mixer. Similarly, the diode impedance seen by the signal frequency (RF) and the IF are given by Figs. 4.10(b) and 4.10(c), respectively. The RF and IF impedances have been computed as quantities normalized to R_0 , where R_0 is the average diode impedance and is approximated by

$$R_0 = R_s/t \quad (4.36)$$

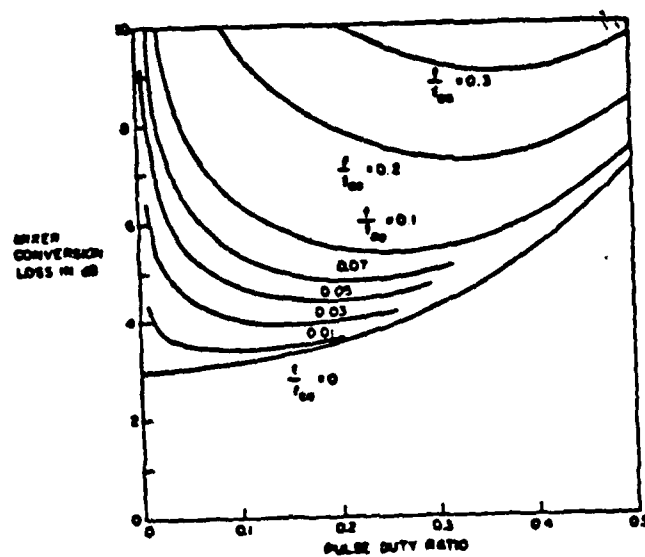


Fig. 4.10(a)

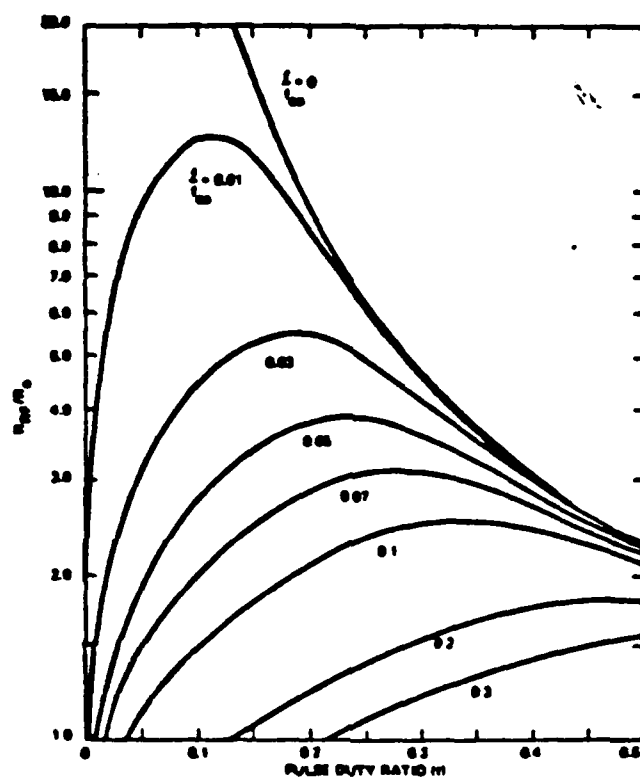


Fig. 4.10(b)

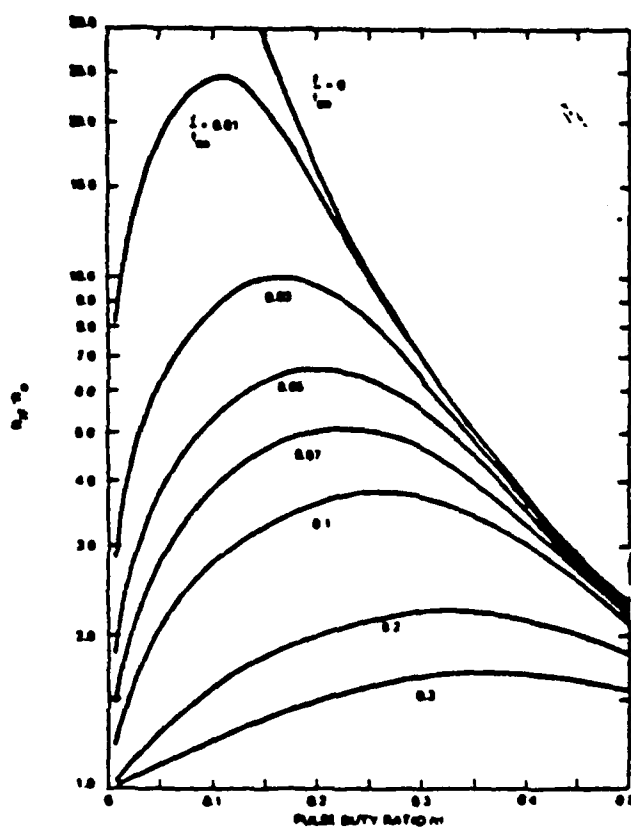


Fig. 4.10(c)

This is the time-averaged diode impedance and thus is the impedance presented to the LO. The results presented in Fig. 4.10 were calculated assuming impedance match at RF and IF. To calculate the mixer conversion loss, therefore, the impedance mismatch losses at RF and IF have to be included.

4.4.2. Attenuator Noise Model

The theory of shot noise in mixers has been examined by several authors over the years [35]-[39]. Recently, Kerr [40] has applied the general mixer theory of Held and Kerr [33] to investigate the noise behavior of ideal mixers. It is shown that an ideal diode, operated as a mixer, is equivalent to a lossy network at a temperature T_{eq} . The equivalent network has ports for all sideband frequencies at which real power flow can occur between the diode and its embedding. I shall first derive the simple noise model for an attenuator, then generalize it to mixers.

Consider an attenuator, at temperature T_{eq} with attenuation coefficient L (defined as $P_{out} = P_{in}/L$), terminated by a matched resistor also at T_{eq} . Under thermal equilibrium, the available noise power P measured at the output of the attenuator is

$$P = KT_{eq}B$$

where B is the noise bandwidth. There are two sources for the measured noise power. One is the noise power contributed by the matched termination, $KT_{eq}B \frac{1}{L}$. The other source is the excess noise power, P_{exc} , generated by the attenuator itself. So the total noise power can be written as

$$\begin{aligned}
 P &= KT_{eq} B \\
 &= KT_{eq} B \left[\frac{1}{L} \right] + P_{exc}
 \end{aligned}
 \tag{4.37}$$

We find that the excess noise power generated by an attenuator in thermal equilibrium is:

$$P_{exc} = KT_{eq} B \left[1 - \frac{1}{L} \right] \tag{4.38}$$

For the mixer, let L_i be the conversion loss for the i th sideband where real power flow occurs. The noise power measured at a matched load can be expressed, similar to (4.38), as

$$P = P_{exc} + \sum_{i=1}^n \frac{KT_{eq} B}{L_i} \tag{4.39}$$

A similar equation for the excess noise power generated by the mixer is found to be

$$P_{exc} = KT_{eq} B \left[1 - \sum_{i=1}^n \frac{1}{L_i} \right] \tag{4.40}$$

The noise in a mixer diode is dominated by shot noise and Johnson noise at moderate current density. We can relate the diode noise temperature T_{eq} to the diode physical temperature T as follows. The mean square noise voltage for the combined Johnson and shot noise can be written as

$$\begin{aligned}
 \overline{V_N^2} &= 4KTBR_s + 2KTbR_j \\
 &\equiv 4KT_{eq} B R_{tot}
 \end{aligned}
 \tag{4.41}$$

where R_s is the series resistance, R_j is the incremental junction

resistance, R_{tot} is the total resistance, and $R_{\text{tot}} = R_s + R_j$. We find that

$$T_{\text{eq}} = T \frac{R_s + \frac{n}{2} R_j}{R_s + R_j} \quad (4.42)$$

In the limit of $(R_s/R_j) \ll 1$, the noise temperature of the mixer is $\frac{n}{2} T$. When $R_s \gg R_j$, the noise is dominated by Johnson noise and T_{eq} approaches T . Under normal conditions, $R_j \gg R_s$; this is another reason to have the diode ideality factor as close to 1 as possible. It should be stressed that T_{eq} includes the effect of Johnson noise and shot noise only. The excess noise occurring at high current density is not accounted for in T_{eq} .

4.4.3. System Noise Temperature and NEP

Assume the signal can enter only through the signal sideband (with conversion Loss L_1) and the image sideband (with conversion loss L_{-1}). The input noise temperature T_n of the diode (Signal Sideband, SSB or Double Sideband, DSB) is defined as the temperature, when the input port is conjugate-matched, necessary to produce the measured excess noise power, assuming the diode is noise-free.

For the single sideband noise temperature, the noise power is assumed to come from the signal sideband only.

$$\begin{aligned} P_{\text{exc}} &= K T_{\text{eq}} B \left[1 - \frac{1}{L_1} - \frac{1}{L_{-1}} \right] \\ &= K T_{M, \text{SSB}} B \frac{1}{L_1} \\ &\equiv K T_M' B \end{aligned} \quad (4.43)$$

we have

$$\begin{aligned} T_{M, \text{SSB}} &= T_{\text{eq}} \left[L_1 - 1 - \frac{L_1}{L_{-1}} \right] \\ &= T_{M-1}' \end{aligned} \quad (4.44)$$

where $T_{M,SSB}$ and T_M' are the SSB input noise temperature and the mixer output noise temperature, respectively.

For double sideband noise temperature, the noise power is assumed to come from both the signal and the image sideband,

$$\begin{aligned} P_{exc} &= KT_{eq}B \left[1 - \frac{1}{L_1} - \frac{1}{L_{-1}} \right] \\ &= KT_{M,DSB} \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right] \end{aligned} \quad (4.45)$$

where

$$T_{M,DSB} = T_{eq} \left[1 - \frac{1}{L_1} - \frac{1}{L_{-1}} \right] \div \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right] \quad (4.46)$$

In a complete detection system, other than the mixer, there is an IF amplifier section. Any amplifier inevitably introduces some excess noise of its own. The noise of an amplifier of amplification G is usually characterized by a noise figure F . This is related to the noise power measured at a matched termination, when the input of the amplifier is also conjugate-matched,

$$\begin{aligned} P &= P_{exc} + KTBG \\ &\equiv FKTBG \end{aligned} \quad (4.47)$$

In analogy to the input mixer noise, we can define the input IF amplifier noise temperature as

$$T_{IF} = T(F-1) \quad (4.48)$$

The system noise temperature, T_s , is defined as the temperature, when the detection system is conjugate-matched, necessary to generate

the measured noise power at the IF output port, assuming the mixers and IF amplifiers are noiseless. The single sideband system noise temperature can be written as

$$\begin{aligned}
 KT_{S,SSB} B \frac{1}{L_1} G &= [(T_A + T_{M,SSB}) \frac{1}{L_1} + T_{IF}] GKB \\
 &= NEP_{SSB} GB \frac{1}{L_1}
 \end{aligned} \tag{4.49}$$

= Measured Noise Power

where SSB noise equivalent power is defined. We have

$$T_{S,SSB} = T_A + T_{M,SSB} + L_1 T_{IF} \tag{4.50}$$

where T_A is the antenna noise temperature. Similarly, the double sideband temperature can be expressed as

$$\begin{aligned}
 KT_{S,DSB} B \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right] G &= [T_A \frac{1}{L_1} + T_A \frac{1}{L_{-1}} + T'_M + T_{IF}] KTB \\
 &= NEP_{DSB} GB \left(\frac{1}{L_1} + \frac{1}{L_{-1}} \right)
 \end{aligned} \tag{4.51}$$

assuming the antenna noise temperature for the image sideband is also T_A .

We have

$$\begin{aligned}
 T_{S,DSB} &= T_A + T'_M \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right] + T_{IF} \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right] \\
 &= T_A + T_{M,DSB} + T_{IF} \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right]
 \end{aligned} \tag{4.52}$$

It is convenient to define a double sideband conversion loss L_{DSB} as

$$L_{DSB} = \left[\frac{1}{L_1} + \frac{1}{L_{-1}} \right]^{-1} \quad (4.53)$$

Then the system noise temperature for both SSB and DSB cases can be expressed identically by

$$T_{S,SSB} = T_A + L_{SSB} [T_M + T_{IF}]$$

DSB DSB

where $L_{SSB} = L_1$.

The NEP can be related to the system noise temperature through

$$NEP_{SSB} = K T_{SSB} \quad (4.55)$$

DSB DSB

References

- [1] D. B. Rutledge and M. S. Muha, "Imaging antenna arrays," IEEE Trans. Antenna Propagat., vol. AP-30, pp. 535-540, 1982.
- [2] D. P. Neikirk, D. B. Rutledge, M. S. Muha, H. Park, and C-X. Yu, "Far-infrared imaging antenna arrays," Appl. Phys. Lett., vol. 40, pp. 203-205, 1982.
- [3] D. P. Neikirk, P. P. Tong, and D. B. Rutledge, "Imaging antenna array at 119 μm ," Appl. Phys. Lett. vol. 41, p. 329, 1982.
- [4] D. Marcuse, Theory of Dielectric Optical Waveguides, Academic Press, New York, 1974.
- [5] F. A. Jenkin and H. E. White, Fundamentals of Optics, Fourth ed., McGraw-Hill, New York, pp. 166-170, 1976.
- [6] Z. Rav-Noy, C-N. Zah, U. Shreter, D. B. Rutledge, T. Wang, S. E. Schwarz, and T. F. Kench, "Monolithic Schottky diode imaging arrays at 94 GHz," Eighth International Conference on Infrared and Millimeter-Waves, Miami, December 1983.
- [7] D. B. Rutledge, D. P. Neikirk, and D. P. Kasilingam, "Integrated-circuit antennas," in Infrared and Millimeter Waves, Chap. 1, vol. 10 Academic Press, New York, 1983.
- [8] J. W. Goodman, Introduction to Fourier Optics, New York, McGraw-Hill, 1968.
- [9] M. Schluter, "Theoretical models of Schottky barriers," Thin Solid Films, vol. 93, pp. 3-9, 1982.
- [10] V. Heine, Phys. Rev. A, vol. 138, p. 1689, 1965.
- [11] J. C. Phillips, J. Vac. Sci. Tech., vol. 11, p. 947, 1974.
- [12] J. M. Andrews and J. C. Phillips, Phys. Rev. Lett., vol. 35, p. 56, 1975.

- [13] L. J. Brillson, Phys. Rev. Lett., vol. 40, p. 260, 1978.
- [14] W. E. Spicer, I. Lindau, P. Skeath, C. Y. Su, and P. Chye, Phys. Rev. Lett., vol. 44, p. 420, 1980.
- [15] S. M. Sze, Physics of Semiconductor Devices, Second Ed., Wiley, New York, 1981.
- [16] N. F. Mott, "Note on the contact between a metal and a insulator or semiconductor," Proc. Cambr. Philos. Soc., vol. 34, p. 568, 1938.
- [17] W. Schottky, "Halbleitertheorie der Sperrschicht," Naturwissenschaften, vol. 26, p. 843, 1938.
- [18] H. A. Bethe, "Theory of the boundary layer of crystal rectifiers," MIT Radiat. Lab. Rep. 43-12, 1942.
- [19] C. R. Conwell and S. M. Sze, "Current transport in metal-semiconductor barriers," Solid State Electron., vol. 9, p. 1035, 1966.
- [20] J. W. Conley, C. B. Duke, G. D. Mahan, and J. J. Tiemann, "Electron tunneling in metal-semiconductor barriers," Phys. Rev., vol. 150, p. 466, 1966.
- [21] R. Stratton, J. Phys. Chem. Solids, vol. 23, p. 1177, 1962.
- [22] F. A. Padovani and R. Stratton, "Field and thermionic-field emission in Schottky barriers," Solid State Electron., vol. 9, p. 695, 1966.
- [23] P. R. Gray and R. G. Meyer, "Analog integrated circuits," Wiley, New York, 1977.
- [24] H. E. Elder and V. J. Glinski, "Detector and mixer diodes and circuits," in Microwave Semiconductor Devices and their circuit applications, H. A. Watson ed., McGraw-Hill, New York, 1968.
- [25] H. C. Torrey and C. A. Whitmer, Crystal Rectifiers, McGraw-Hill, New York, 1948.
- [26] W. M. Sharpless, "Wafer-type millimeter-wave rectifiers," Bell Syst. Tech. J., vol. 35, pp. 1385-1402, 1956.

- [27] G. C. Messenger and C. T. McCoy, "Theory and operation of crystal diodes as mixers," *Proc. IRE*, vol. 45, pp. 1269-1283, 1957.
- [28] L. Mania and G. B. Stracca, "Effects of the diode junction capacitance on the conversion loss of microwave mixers," *IEEE Trans. Communications*, vol. COM-22, pp. 1428-1435, 1974.
- [29] A. R. Kerr, "Low-noise room-temperature and cryogenic mixers for 80-120 GHz," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 781-787, 1975.
- [30] M. R. Barber, "Noise figure and conversion loss of the Schottky barrier mixer diode," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-15, pp. 629-635, 1967.
- [31] L. E. Dickens and D. W. Maki, "An integrated-circuit balanced mixer, image and sum enhanced," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-23, pp. 276-281, 1975.
- [32] A. A. M. Saleh, *Theory of Resistive Mixers*, Cambridge, MA, MIT Press, 1971.
- [33] D. N. Held and A. R. Kerr, "Conversion loss and noise of microwave and millimeter-wave mixers: part I -- theory," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26, pp. 49-55, 1978.
- [34] M. E. Heine, "Inherent signal losses in resistive-diode mixers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-29, pp. 281-292, 1981.
- [35] M. J. O. Strutt, "Noise figure reduction in mixer stages," *Proc. IRE*, vol. 34, pp. 942-950, 1946.
- [36] A. van der Ziel and R. L. Watters, "Noise in mixer tubes," *Proc. IRE*, vol. 46, pp. 1426-1427, 1958.
- [37] A. van der Ziel, *Noise: Sources, Characterization, and Measurement*, Prentice-Hall, Englewood Cliffs, NJ, 1970.

- [38] C. S. Kim, "Tunnel diode converter analysis," IRE Trans. Electron Devices, vol. ED-8, pp. 394-405, 1961.
- [39] C. Dragone, "Analysis of thermal shot noise in pumped resistive diodes," Bell Syst. Tech. J., vol. 47, pp. 1883-1902, 1968.
- [40] A. R. Kerr, "Shot noise in resistive-diode mixers and the attenuator noise model," IEEE Trans. Microwave Theory Tech., vol. MTT-27, pp. 135-140, 1979.
- [41] A. Jelenski, M. V. Schneider, A. Y. Cho, E. R. Kolberg, and H. Zirath, "Noise measurements and noise mechanisms in microwave mixer diode," in 1984 IEEE MTT-S Int. Microwave Symp. Digest, pp. 552-554, 1984.

Figure Captions

- 4.1. "Reverse microscope" optical system for an imaging array (after Rutledge et al. [1]).
- 4.2. A planatic arrangement for the reverse microscope optical system, where R is the radius of the substrate lens and n_1 is the index of refraction of the lens material (after Rutledge et al. [7]).
- 4.3. Bow-tie antenna array designed for use on $\epsilon_r = 12$ substrate (after Neikirk et al [2]).
- 4.4. Feed patterns of a single element in the bow-tie array (a) for $\epsilon_r = 12$ and (b) for $\epsilon_r = 4$. The pattern were measured for a substrate lens satisfying the aplanatic condition (after Neikirk et al. [2], [3]).
- 4.5. Frequency-independent resistance of long bow ties. The lines show the calculated results from (4.1). The circles and crosses are measured value (measured at 2-4 GHz) (after D. B. Rutledge et al. [7]).
- 4.6. Energy-band diagram of a metal in contact with an n-type semiconductor.
- 4.7. Avalanche breakdown voltage versus impurity concentration for one-sided abrupt junctions in Ge, Si, <100>-oriented and GaP (after Sze, [13]).
- 4.8. Primary conduction mechanisms of majority carrier transport at a Schottky contact.
- 4.9. Equivalent circuits for video detection; (a) the radio frequency equivalent circuit, (b) the low frequency equivalent circuit.
- 4.10. Computed results for double side-band mixer; (a) intrinsic conversion loss, (b) RF impedance, and (c) IF impedance (after Dickens et al., [30]).

CHAPTER 5

EXPERIMENTS WITH SCHOTTKY DIODE IMAGING ARRAYS

Experimental methods and results are presented in this chapter. First, the device fabrication is discussed. It is not my intention to give the detail sequence of device fabrication. Rather, I would like to present the reasoning and understanding behind the particular fabrication technology we developed or adopted. Section 2 describes the video detection experiments. It is used not only as a detector, but also as a diagnostic tool for the diode and the imaging optics. Heterodyne detection experiments are discussed in section 3. The final section is devoted to performance of the optics of the imaging array.

5.1. Device Fabrication Technologies

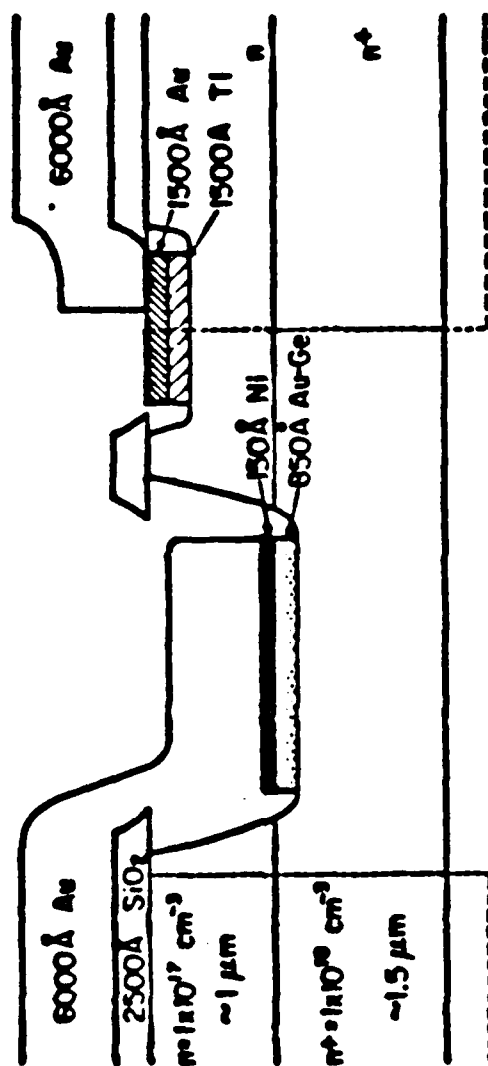
There are four major parts in the fabrication process and the sequence of fabrication is: (1) ohmic contact; (2) multiple energy proton implantation; (3) Schottky contact; and (4) antenna metalization.

The importance of ohmic contact was made obvious in the discussion of video and heterodyne detection. There are two reasons for device isolation. DC isolation of the individual devices is certainly necessary. Antennae on lossy dielectric, such as the epi-layer in this case, behave very differently from free space antennae. In order to avoid unnecessary free carrier absorption and interference with the antenna pattern, all the epi-layers other than the device area have to be rendered nonconductive. This can be achieved either by mesa etch or by proton implantation [1], [2]. Proton implantation is favored because it preserves the planar surface of the wafer. Antenna metalization is straightforward. I shall describe the pertinent facts about ohmic contact, proton implantation and Schottky contacts in the following.

The device dimensions and the cross section of the finished device are shown in Figs. 5.1 and 5.2, respectively. The starting wafer is a Chromium doped semi-insulating GaAs wafer with two layers of LPE grown epitaxial layers, $\sim 1 \mu\text{m}$ of n-type active layer with doping concentration $\sim 1 \times 10^{17} \text{ cm}^{-3}$ on top of a $\sim 1.5 \mu\text{m}$ heavily doped buried layer with doping concentration $\sim 1 \times 10^{18} \text{ cm}^{-3}$. The doping concentration of the n-type active layer was chosen as high as possible without sacrificing the diode ideality factor (section 4.2.3). The doping concentration of the buried layer is the upper limit practically achievable by LPE in order to minimize the spreading resistance. Because of the small dimensions involved, the best way for patterning is by positive photoresist liftoff. All the patternings are done by the Canon 4 times reduction projection aligner.

The etchant used throughout the fabrication process is the Ammonium Hydroxide (NH_4OH) based etchant. The NH_4OH based etchant is known to give better surface Stoichiometry than others [3]. This is an important consideration especially for Schottky contacts. Another advantage is that the concentration can be varied to give a wide variety of etching rates.

The finished chip is a 2×10 array. Pictures of the device and chip are shown in Figs. 5.3 and 5.4, respectively. The DC characteristics of the typical diode are as follows: series resistance $R_s \sim 20 \Omega$, ideality factor $n \leq 1.06$, reverse breakdown voltage $\geq 10 \text{ v}$. The junction capacitance is estimated to be $\sim 12 \text{ fF}$. This gives the cutoff frequency of the diode $f_0 = 1/2 \pi R_s C_0 = 663 \text{ GHz}$. The DC I-V characteristics of three adjacent devices on a typical chip are shown in Fig. 5.5. A typical forward and reverse I-V characteristic is shown in Fig. 5.6.



S. I. GaAs

Fig. 5.2

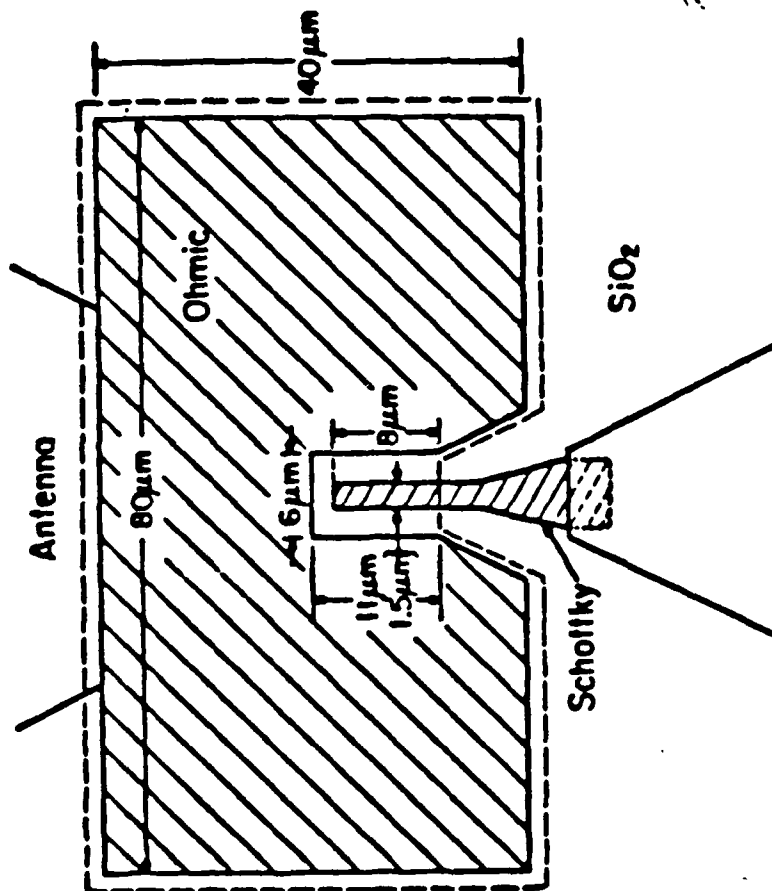


Fig. 5.1

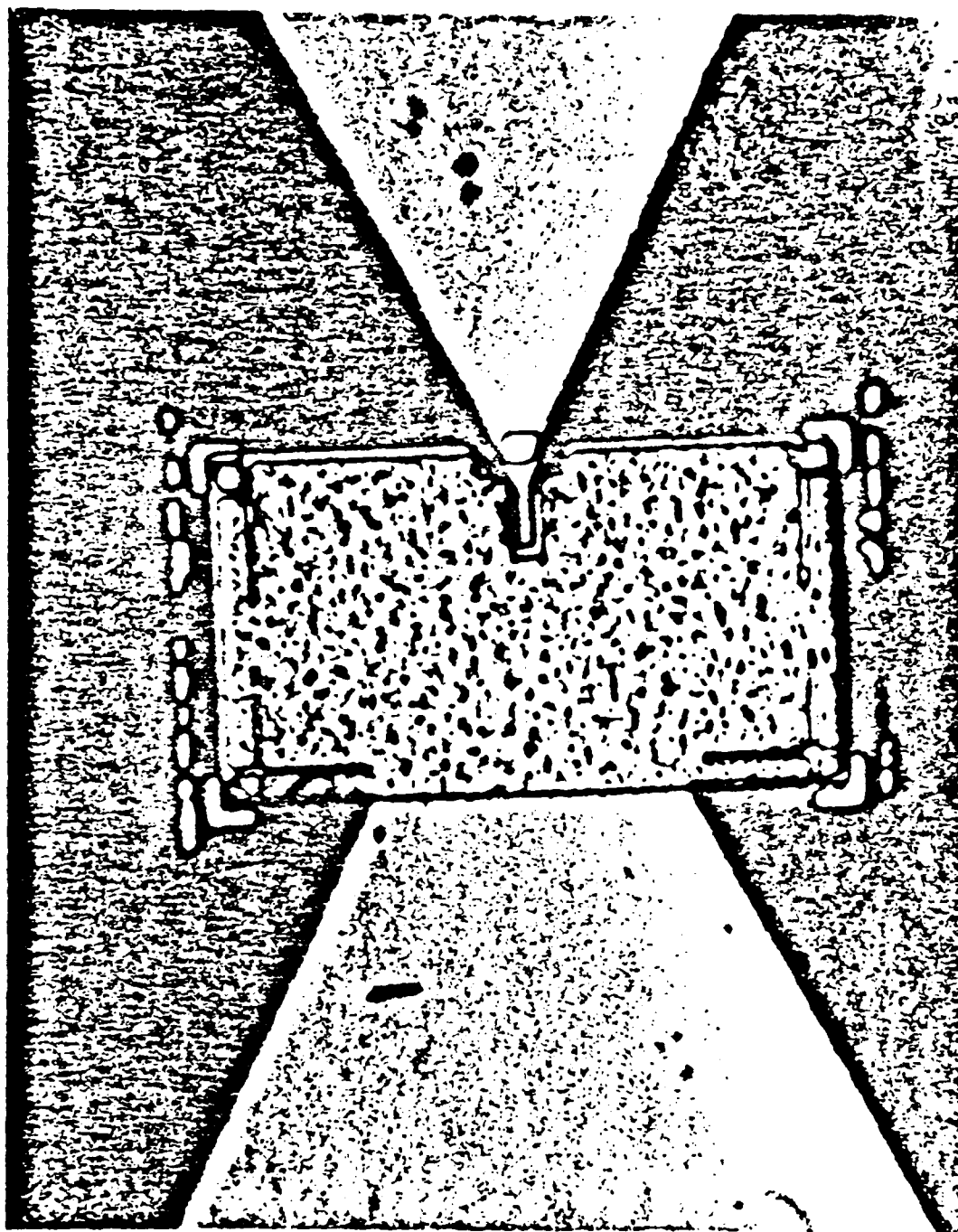


Fig. 5.3

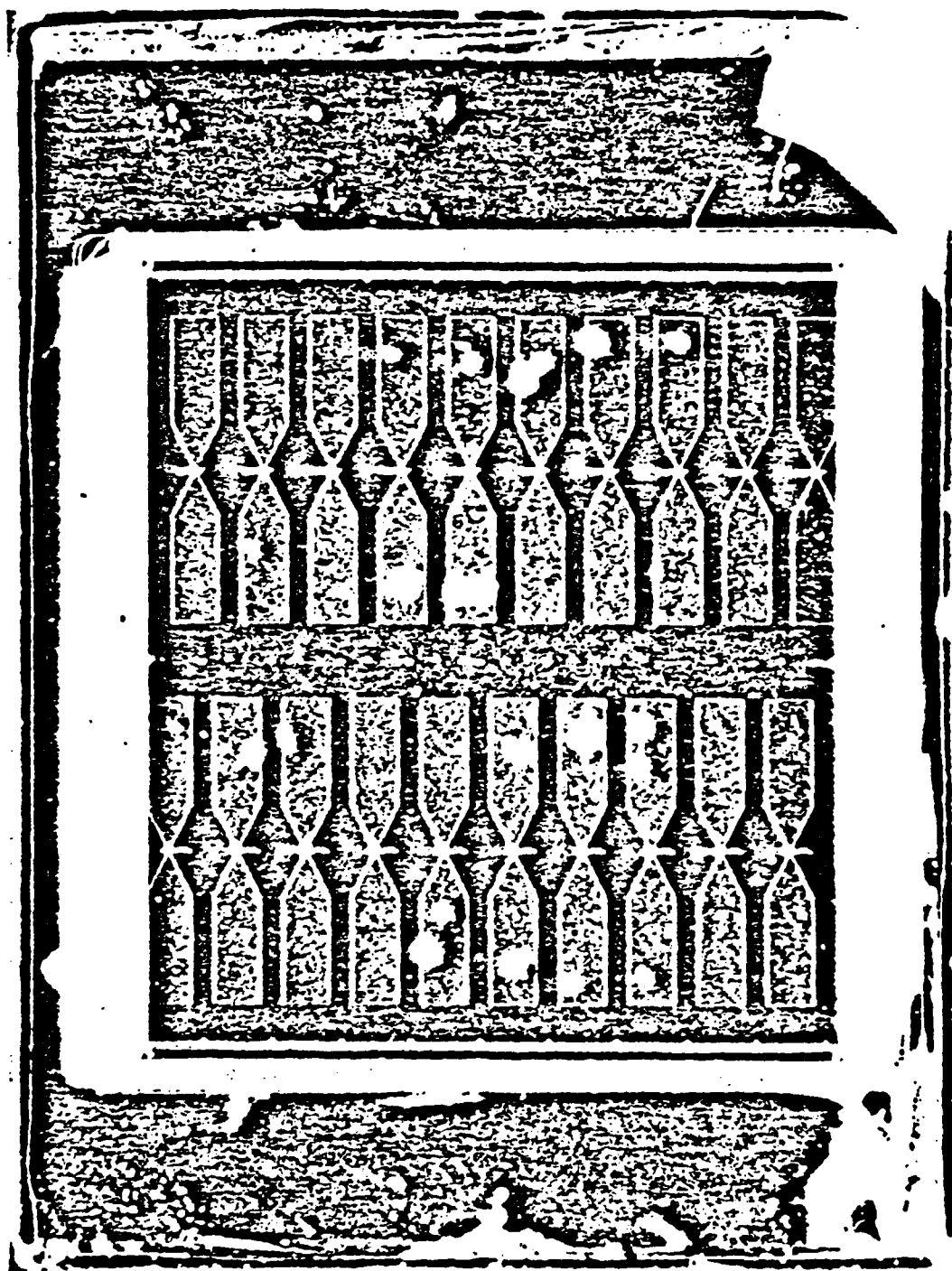


Fig. 5.4

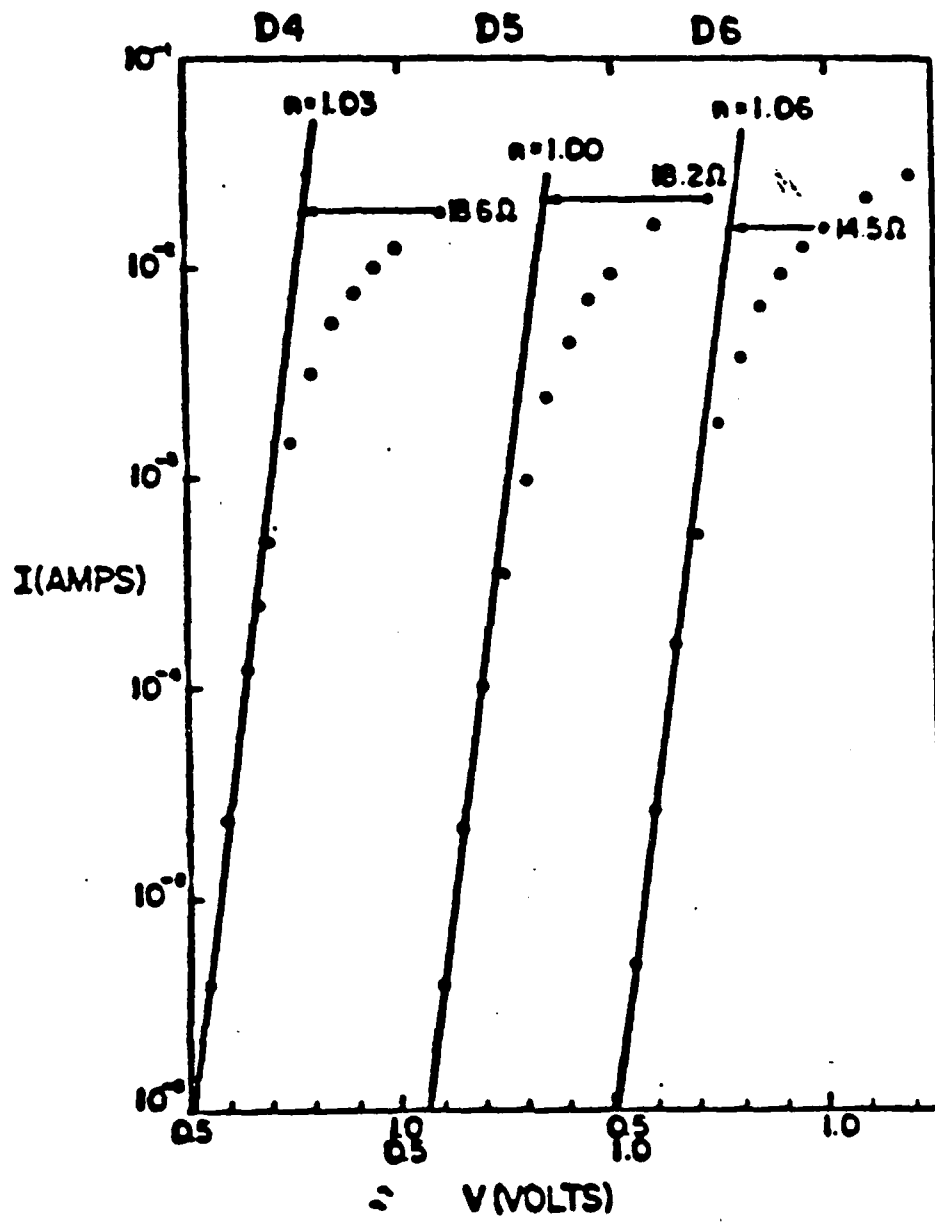


Fig. 5.5

5 V/div, .02 mA/div | 0.5 V/div, 0.2 mA/div

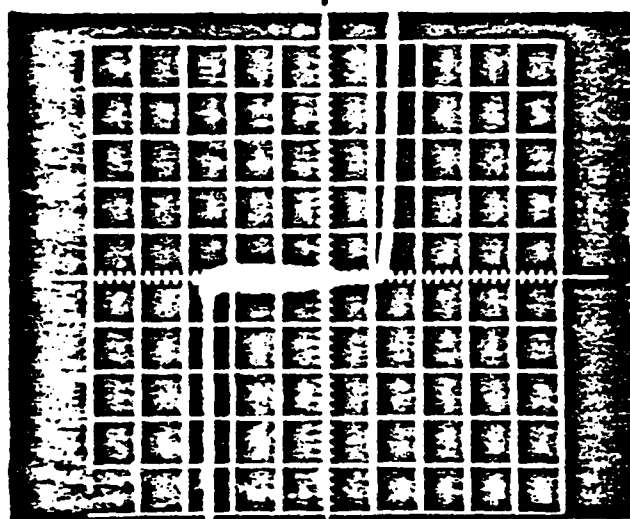


Fig. 5.6

5.1.1. Schottky Contacts

There are several criteria that the metal has to meet for use as Schottky contact metal. Long-term stability of the junction is certainly an important consideration. The metal should be easy to be patterned to the small dimensions necessary for millimeter-wave frequency applications. Some metal contacts are more accessible to surface contamination. In order not to unduly complicate the fabrication process, these metals are to be avoided.

Tungsten was considered first, for its excellent long term stability even at temperature as high as 500 C [4], [5]. It was ruled out, however, since it is extremely difficult to pattern, especially for small junctions. The ideality factor of the diode after patterning is poor, and has high leakage current and low breakdown voltage.

Another drawback of Tungsten is that it has to be deposited by an r.f. sputtering process. Although, at lower r.f. power the first order effect does not affect diode DC parameters, it does result in degradation of diode noise properties [6]. At higher r.f. power, both noise properties and the diode DC characteristics are degraded. This resulted from the energetic ion bombardment during the sputtering process. Such damage has been shown to be in the form of dislocation loops in the epitaxial layer or vacancies near the metal-semiconductor interface.

Chromium [7] was chosen next, for it can be deposited and patterned easily by thermal evaporation and positive photoresist liftoff. [Positive photoresist liftoff is particularly suitable for defining small dimensions]. Excellent diodes have been made in this way. Unfortunately, it is very accessible to surface contamination resulting in diodes of unpredictable properties. The long term stability of the chromium Schottky contact is also doubtful.

Titanium [8] was finally tried. It can be easily deposited and patterned by evaporation and positive photoresist liftoff. It is believed that, because of its highly reactive nature, it readily reacts with the surface contaminant and hence forms a nice clean interface with the GaAs substrate. The resulting diode has nearly ideal DC characteristics; the leakage current is low, the ideality factor is close to 1.0 and the junction breakdown voltage is about the theoretically predicted value (Fig. 4.7). Long term stability of the junction is reasonably good also; no interdiffusion was observed at temperatures as high as 350 K [8].

5.1.2. Ohmic Contacts

The quality of an ohmic contact to GaAs devices is one of the most significant factors affecting the performance of such a device. Rideout [9] has provided an excellent theoretical treatment of ohmic contact to GaAs but admits that "ohmic contact technology has developed thus far more as a technical art than as a science." The comment is still true almost 10 years since this statement was made. Research in this field is still active.

The most commonly used material system for ohmic contact to GaAs is the Au-Ge/Ni contact. The contact is usually formed by evaporating a eutectic mixture of Au-Ge (88-12% by wt., respectively) followed by a thin layer of Ni. It is then alloyed for a minute or so at about 450 C. The conventional wisdom about the formation of A-Ge/Ni ohmic contact on GaAs is summarized as follows.

It is believed that Ge dopes the underlying GaAs so heavily that tunneling may occur through the formed Au-GaAs Schottky barrier. The sequence of evaporation does not seem to change the final arrangement of the constituents, [10]. Immediately under the contact a highly disordered

region occurs. Gold precipitates, and hexagonal AuGa rich regions with high density of dislocation lines extending through more than 300 nm deep appear [11]. The role of Ni [10] is to enhance the indiffusion of Ge (by increasing the solubility of GaAs in the AuGe Ni melt). It also improves the contact's adherence to the surface. Au enhances the outdiffusion of Ga so that most of the Ge sits on substitutional Ga sites creating the desired n^+ -layer. It was also noted that an excess of Au [12] or Ni [13] degrades the contact resistivity. Excess Au, most probably, leads to excess Ga outdiffusion which leave excess As and not enough Ge to replace the Ga. Also, Au and Ni act as acceptors and may compensate the underlying n^+ -layer. It is advisable to use thin contact layers to form ohmic contacts and to enhance the conductivity of the contact by post-alloy deposition, or a pre-alloy deposition with some barrier layer. The resistivity of the GaAs under the contact was found to be 5-10 times higher than the original uncontacted material and it varies very slightly down to a depth of more than 400 nm [14]. (It is even hypothesized that the different conductivity layer under the contact extend throughout the total depth of the epi -layer!) For the ohmic contact used in the imaging arrays, we try to keep the total Au-Ge/Ni thickness below 100 nm and a Ni to Au-Ge thickness ratio of 1 to 6.

Usually the contact is alloyed, without a cap, in H_2 or forming gas environment. The surface of the contact made in this way is very rough; it is rough on the scale of a micron (considering that the total Au-Ge/Ni thickness is less than 100 nm!). The contact also tends to become open at relatively low current. The uniformity of the contact is poor also. Experimentally, we found that all these can be remedied by alloying with a CaF_2 cap of 500 nm thick. Best results are obtained for contacts alloyed at 450 C for 1 min. As the alloy time increases,

the uniformity becomes poorer. The resulting contact is reproducible and has a specific contact resistance of $\sim 1 \times 10^{-6} \Omega\text{-cm}^2$.

5.1.3 Proton Implantation

The rule of thumb for the penetration depth of proton into GaAs is 1 μm per 100 KeV of energy. The optimum dosage for multiple energy proton implantation has been established by Donnelly et al. [2]. For a total epi-layer thickness of $\sim 2.5 \mu\text{m}$ for the wafer, the following energies and dosages are adopted: 350 KeV, 250 KeV, and 150 KeV with dosages of $5 \times 10^{-14} \text{cm}^{-2}$, $3 \times 10^{-14} \text{cm}^{-2}$ and $1.5 \times 10^{-14} \text{cm}^{-2}$, respectively.

As marked by the dash line in Fig. 5.1, the boundary of the proton implanted area should be as close to the device as possible so as to avoid extra capacitance introduced by any epi-layer underneath the metal antenna. The fine patterning required for the proton implantation mask rule out the possibility of using plating for depositing the proton implantation mask. I adopt the implantation mask developed by Van Tuyt et al. [15], a three layer mask composed of $\text{CaF}_2/\text{SiO}_2/\text{Au}$ with thicknesses of 150 nm/150 nm/2 μm , respectively. Au is used as the proton stopping material because of its high stopping power. CaF_2 is used as a liftoff agent, since it dissolves readily in dilute HCl ($\text{HCl}:\text{H}_2\text{O} \sim 1:10$). SiO_2 is used to ensure that the Au film remains a single piece so as to be removed completely. The patterning is also done by a positive photoresist liftoff, as usual, except that a special type of photoresist, AZ 1375, is used in order obtain photoresist thicker than 3 μm for proper liftoff of the implantation mask.

5.2. Video Detection Experiments

The designed center frequency of the device was 100 GHz. Testing was performed, however, at 69 GHz because of availability of Klystron and

Gunn diode sources at that frequency. Since the diode cutoff frequency is about 660 GHz, performance should be little affected. In the first part of the section, experiments to measure voltage responsivity are described and results presented. The second half of the section is concerned with the noise properties of the diode, namely the noise equivalent power.

5.2.1. Voltage Responsivity

A block diagram of the experimental setup used for the 69 GHz video detection is shown in Fig. 5.7. A Harris 819-A Universal Klystron power supply was used to square wave modulate the power of an OKI 70 V 10A Klystron. A TRG 112E-125/387 isolator was used to isolate the Klystron. A TRG 510E-387 variable attenuator was used to control the output power. A TRG 551E/387 frequency meter was used to monitor the Klystron oscillation frequency. The power at the coupled port of a Baytron 3E-30/20 20 dB directional coupler was monitored by means of a IN53 diode. Power at the through port of the coupler was fed to a TRG E681 standard gain horn and radiated out toward the imaging system. The power radiated was measured by replacing the standard gain horn with a TRG V981 Water Calorimeter.

The distance between the horn antenna and the TPX Objective lens is chosen to be greater than 100 cm so that it satisfies the far-field condition [16]. The power incident on the TPX objective lens was calculated by the antenna formula [16]. The power was coupled into the imaging elements (the antenna and the diode) through the silicon substrate lens and the GaAs substrate. The rectified voltage was measured by a PAR 124 lock-in amplifier with a PAR-118 preamplifier.

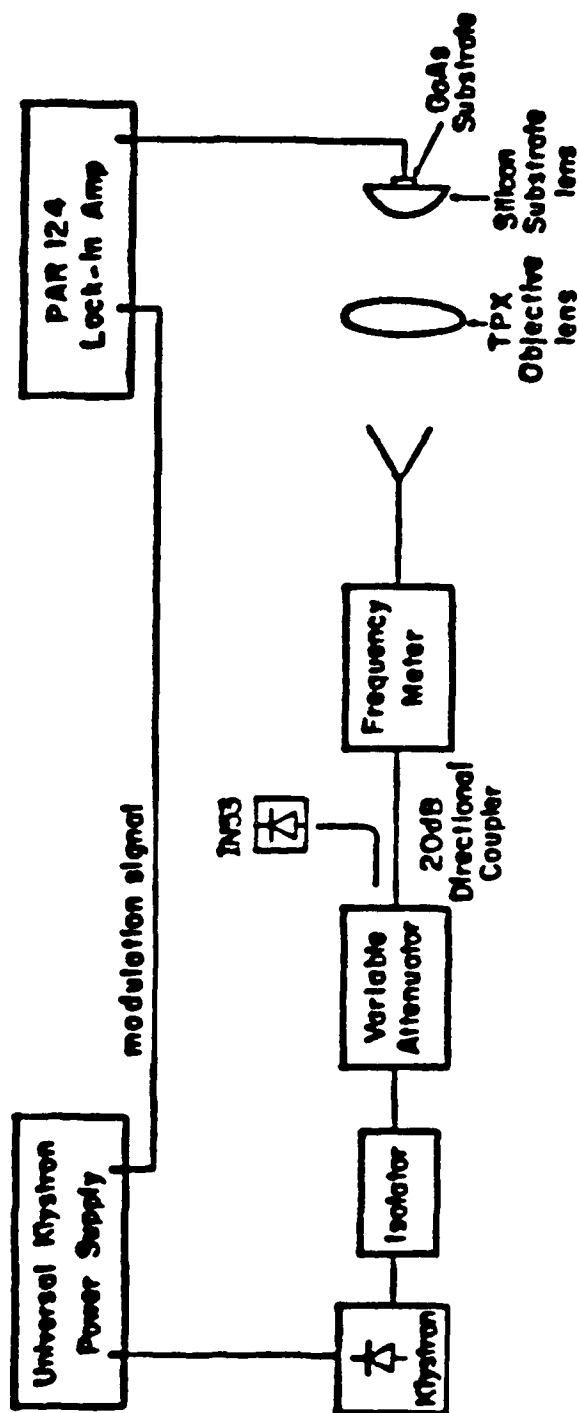
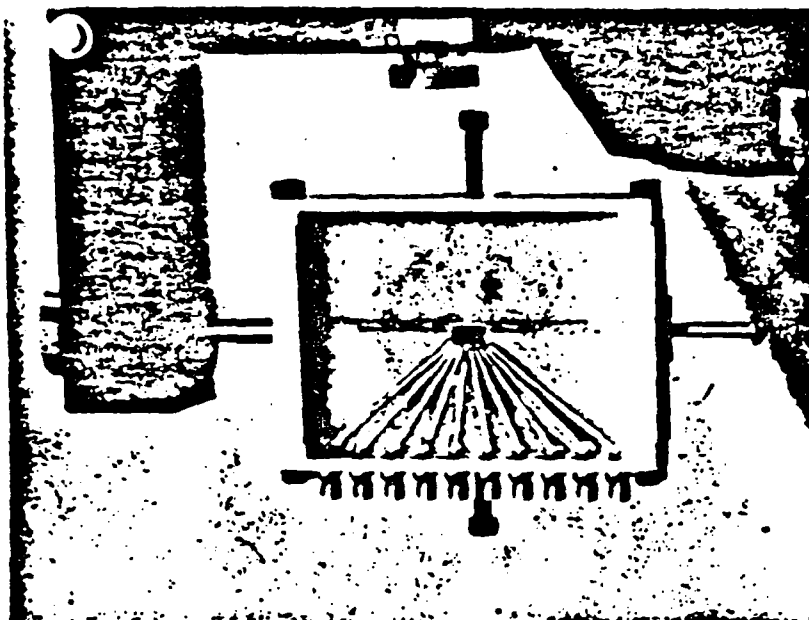


Fig. 5.7

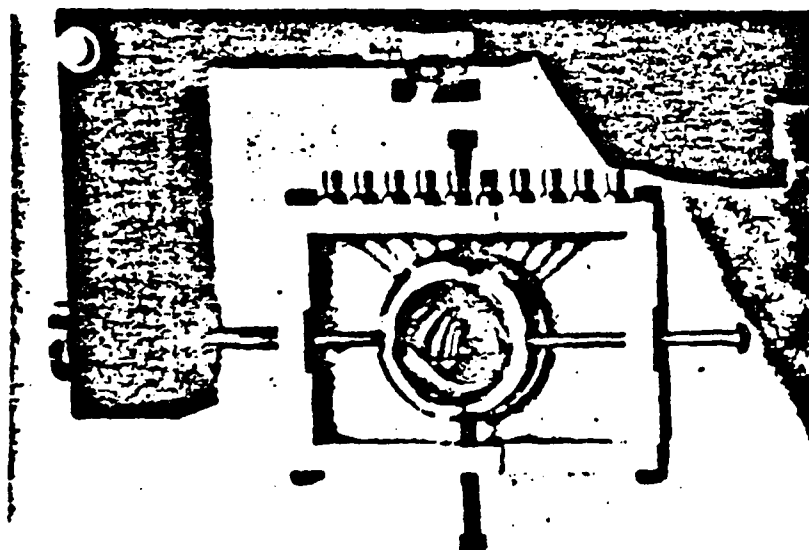
The Schottky diode was forward biased by a battery power supply and the bias voltage was monitored by a Data Precision 1450 digital multimeter. The reason for using a battery power supply was to avoid the ground loop problem and to have a precisely defined incremental junction resistance R_j ; for an error of R_j smaller than 5%, the bias voltage has to be accurate to within 1mV at room temperature. With the battery bias supply, it is easy to control the bias voltage down to 0.1 mV.

The imaging Schottky array measured is a 1×10 linear array. The finished chip of 2×10 array was scribed into 2 1×10 arrays. The linear array was glued to a pc board by silver epoxy and glue. The pc board was mounted into an aluminum chip and substrate lens holder. Connection of each individual device to a SMA connector was made. The substrate lens was mounted on the backside of the GaAs substrate and held in place by four screws. The details of the chip mount design are shown in Fig. 5.8. The aluminum holder was then mounted on a five degrees translation (three degrees of freedom) and rotation stage (two degree of freedom) and the TPX objective lens was mounted on a vertical translation state. The imaging optics is shown in Fig. 5.9.

The alignment and focusing procedures are as follows. The alignment is performed without the objective lens. When the axis of the silicon substrate lens is exactly aligned with the antenna/detector element, optimum power coupling occurs when the incident direction is parallel to the substrate lens axis. Because of the difficulty in exact alignment of the antenna/detector to the substrate lens, the two rotational axes can be adjusted for optimum power coupling. The objective lens is then put into place. The height of the objective lens mount and the three translational adjustments are then adjusted for optimum power coupling.



(b)



(a)

Fig. 5.8

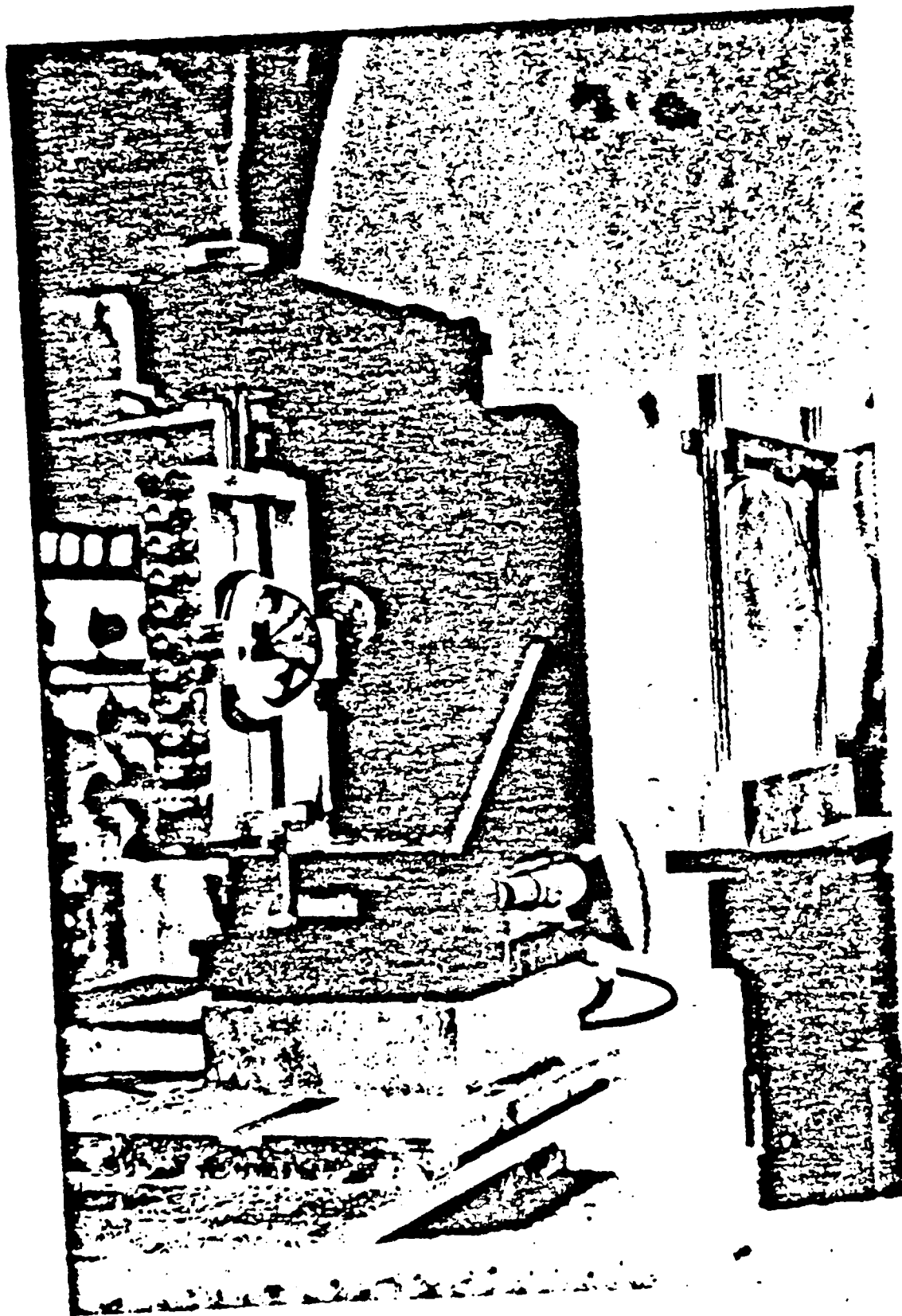


Fig. 5.9

Voltage responsivity R_V is measured as a function of DC bias voltage. The best voltage responsivity measured is 534 V/W at DC bias voltage of 0.56 V. The normalized voltage responsivities (with peak normalized to 1) as a function of DC bias voltage V_D , for the experimental results and the theoretical curves are plotted in Fig. 5.10. Notice the close fit of the shape of the experimental curve (and in particular the voltage at which the responsivity peaks) to the theoretical curve. Because the voltage at which the responsivity peaks depends very strongly on the junction capacitance, we conclude from the measurement that the junction capacitance is indeed very close to what we estimated from the known doping concentration of the epi-layer and the junction area. This also shows that the device can be modeled by the equivalent circuit model presented in section 4.3. By comparing the magnitude of the theoretical and experimental responsivity values, we can also draw the conclusion that the optical system loss is 7 dB. We shall come back to the details of this problem in section 5.4.

5.2.2. Noise-Equivalent Power

Diode noise at room temperature was measured by using the PAR 124 lock-in amplifier with a PAR 118 preamplifier operated in the RMS voltmeter mode with the signal channel quality factor Q set to 10% Equivalent Noise Bandwidth (10% ENBW). Since every amplifier inevitably contributes noise to the signal being amplified, in order to obtain the true measure of noise contributed by the diode, the noise power contributed by the lock-in amplifier has to be subtracted. The noise figure F of the lock-in amplifier as a function of source resistance and frequency was calibrated by using a metal-wire-wound precision potentiometer.

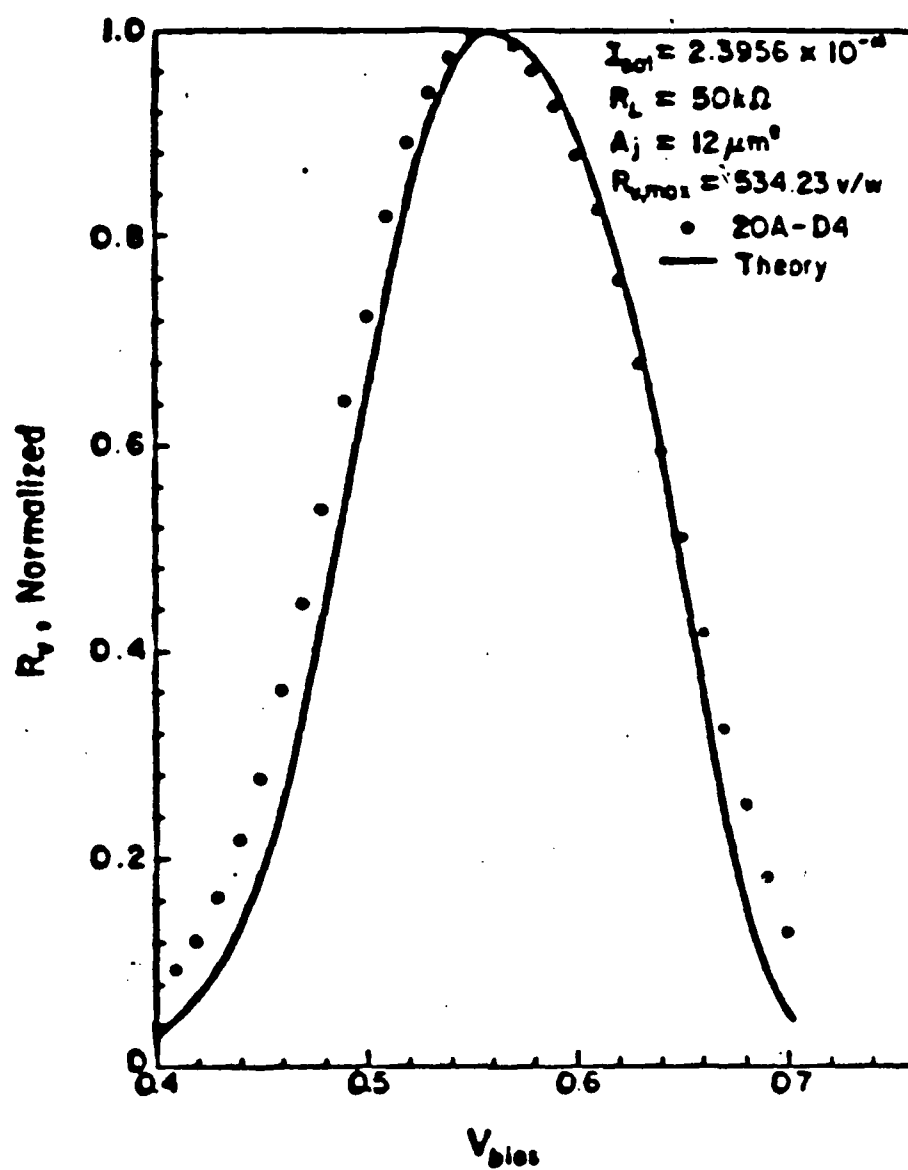


Fig. 5.10

The noise power contributed by the diode itself can then be determined by subtracting the noise power contributed by the amplifier. By the known equivalent circuit model (Fig. 4.9), we can find the equivalent noise voltage source by (4.35). Figure 5.11 shows the equivalent noise voltage source as a function of frequency for two bias voltages. It was found that mean-square noise voltage $\overline{V_N^2}$ can be fit by (4.35), with $\alpha = 2/3$,

$$\overline{V_N^2} = 2KTB R_j \left[1 + \left(\frac{f_0}{f} \right)^{2/3} \right] \quad (5.1)$$

R_j is the incremental junction resistance. The cross-over frequency f_0 as a function of bias voltage is shown in Fig. 5.12. The cross-over frequency depend on bias voltage V as

$$f_0 \sim \exp \left[\frac{1.2 qV}{KT} \right] \quad (5.2)$$

up to the limit of the lock-in amplifier ($f = 100$ KHz), the diode noise is still dominated by the flicker noise.

The noise equivalent power, NEP, can now be calculated by (4.35). The results are shown in Fig. 5.13. The best measured NEP is 1.46×10^{-11} W/ $\sqrt{Hz}$ at a bias voltage of 0.59 V for a frequency of 100 KHz. Since the diode noise is still limited by the flicker noise, it is possible to reduce the NEP by a factor of 2 simply by raising the frequency to higher than f_0 .

5.3. Heterodyne Detection Experiments

There are three methods of conversion loss measurement suitable for the quasi-optical mixer; they are the Y-factor method (Dick's Radiometry Method) [17] the DC incremental loss method [18] and the direct heterodyne method [19]. The direct heterodyne method was chosen for this work

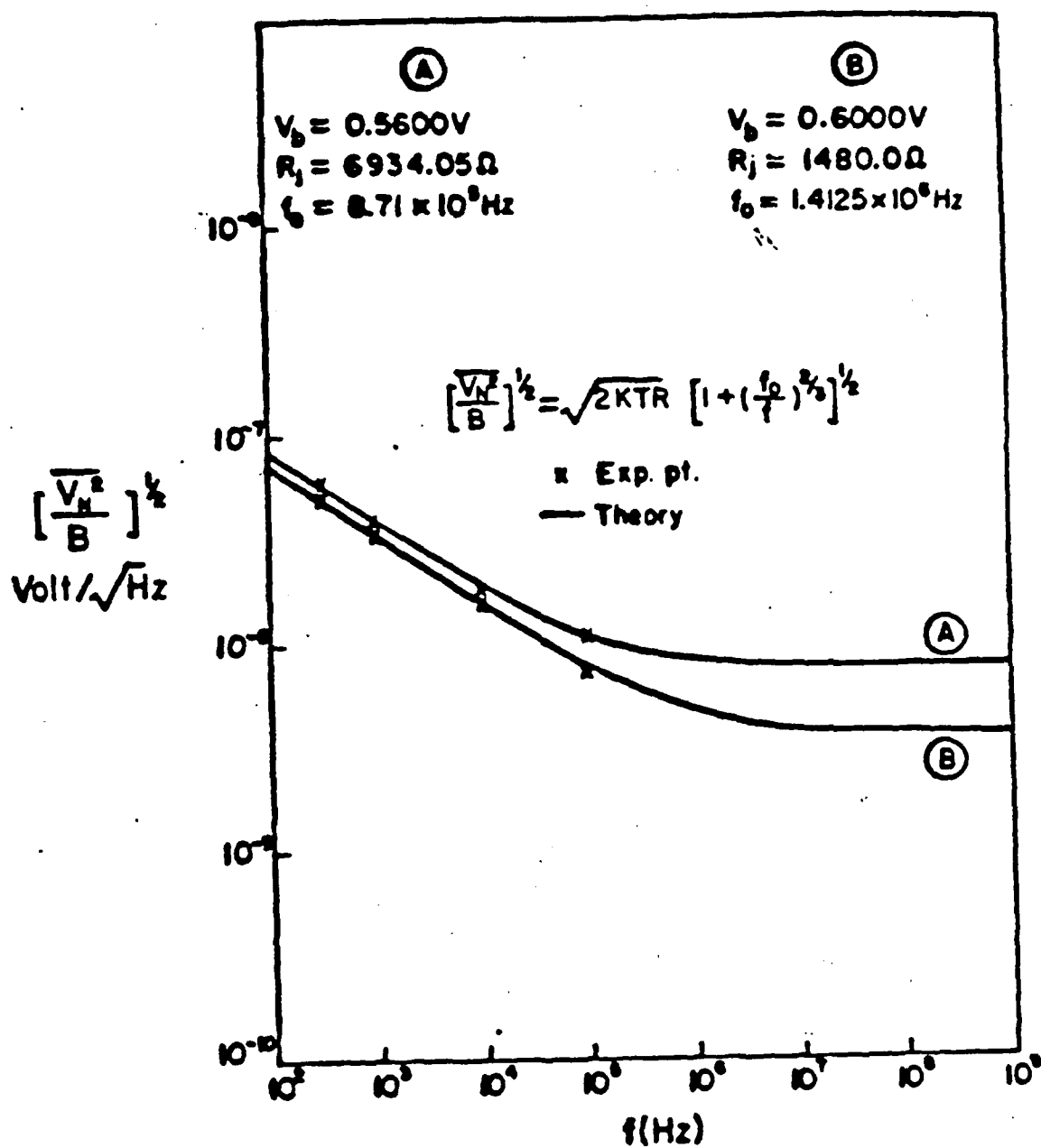


Fig. 5.11

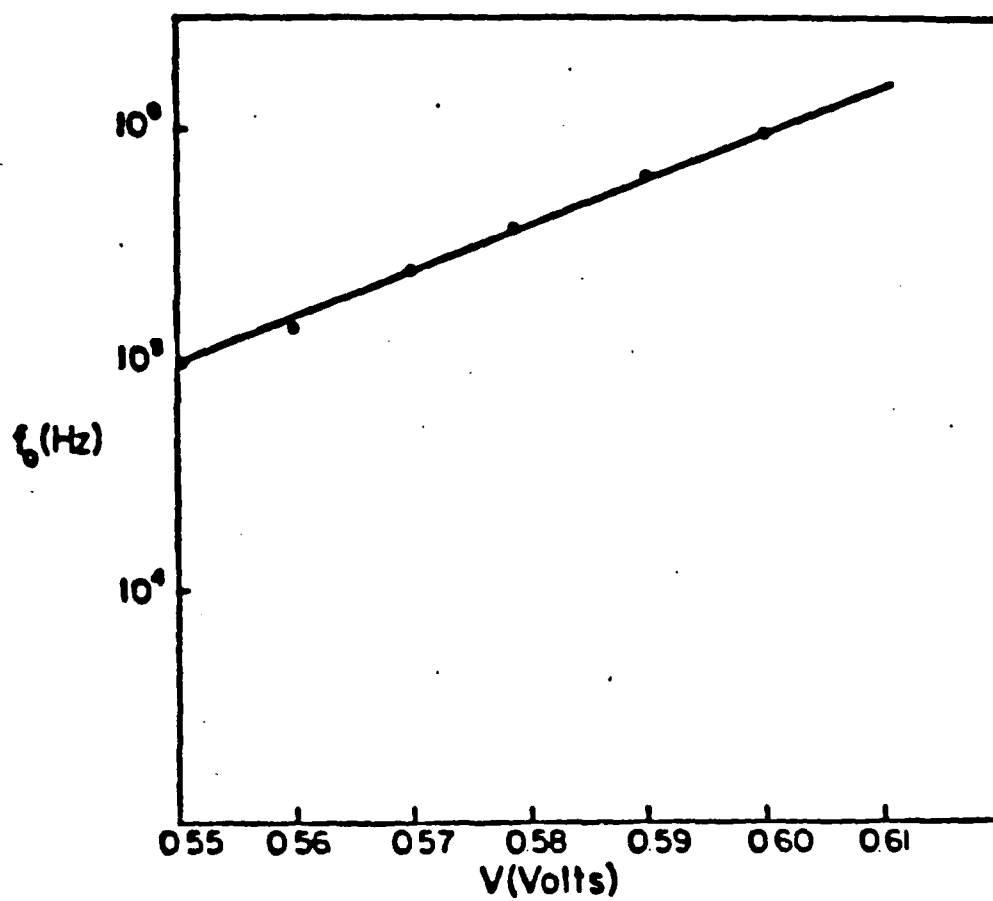


Fig. 12

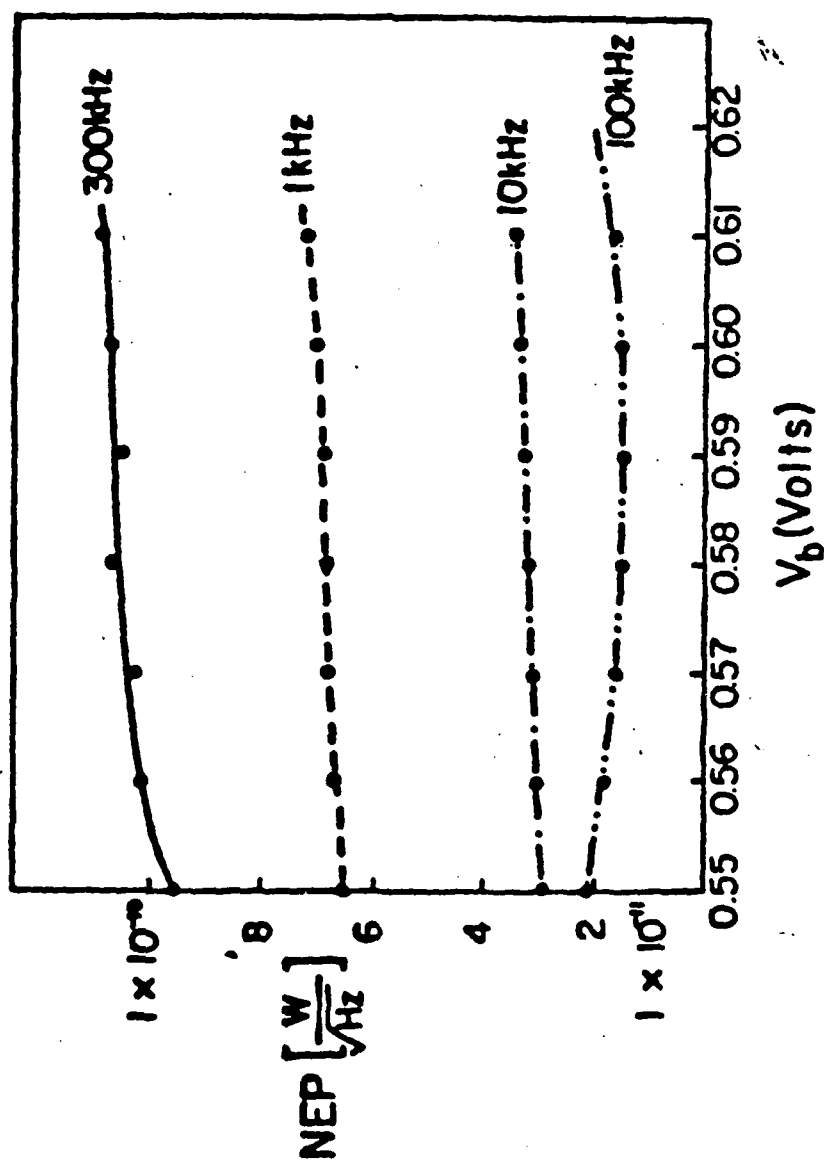


Fig. 5.13

because it is the most direct and straightforward way. The conversion loss is measured directly from the known RF signal power and the measured IF power.

5.3.1. Conversion Loss

The experimental setup for the mixing experiment is shown in Fig. 5.14. The Klystron section identical to the one described in the last section was used as the local oscillator. The signal was provided by a Hughes 47224H-1305 Gunn diode. A TRG V110-296 isolator and a TRG V510-211 variable attenuator were used to isolate the signal source and for power adjustment, respectively. The LO power was internally modulated at 200 Hz in order to maximize instantaneous power output. The 75 MHz IF signal from the mixer was fed into a MITEQ AU-2A-0110 amplifier. The amplified IF signal was either monitored by a Tektronix 7904 oscilloscope with 7A19 amplifier and 7B92 dual time base or by measuring power with an HP 435B power meter. The DC bias circuit was the same as for the video detection experiment.

The bias dependence of the conversion loss is shown in Fig. 5.15, for a LO power of 2.1 mW and an RF power of 1 μ W per diode. The optimum conversion loss of 19.4 dB is achieved at a forward DC bias of 0.73V. The pulse duty ratio, as measured by the Tektronix Oscilloscope is 40%. We can compare conversion with the estimated conversion loss given by the modified Barber theory.

Because of the 20 Ω series resistance of the diode, the actual bias voltage drop across the junction is 0.72 V when V_b is 0.73 V. The capacitance of the junction at this bias voltage is 32.81 fF. This gives a diode cutoff frequency of 242.5 GHz and hence $f/f_{c0} = 0.3$. Knowing that the pulse duty ratio $t = 0.4$, we find from Fig. 4.10(a) that the

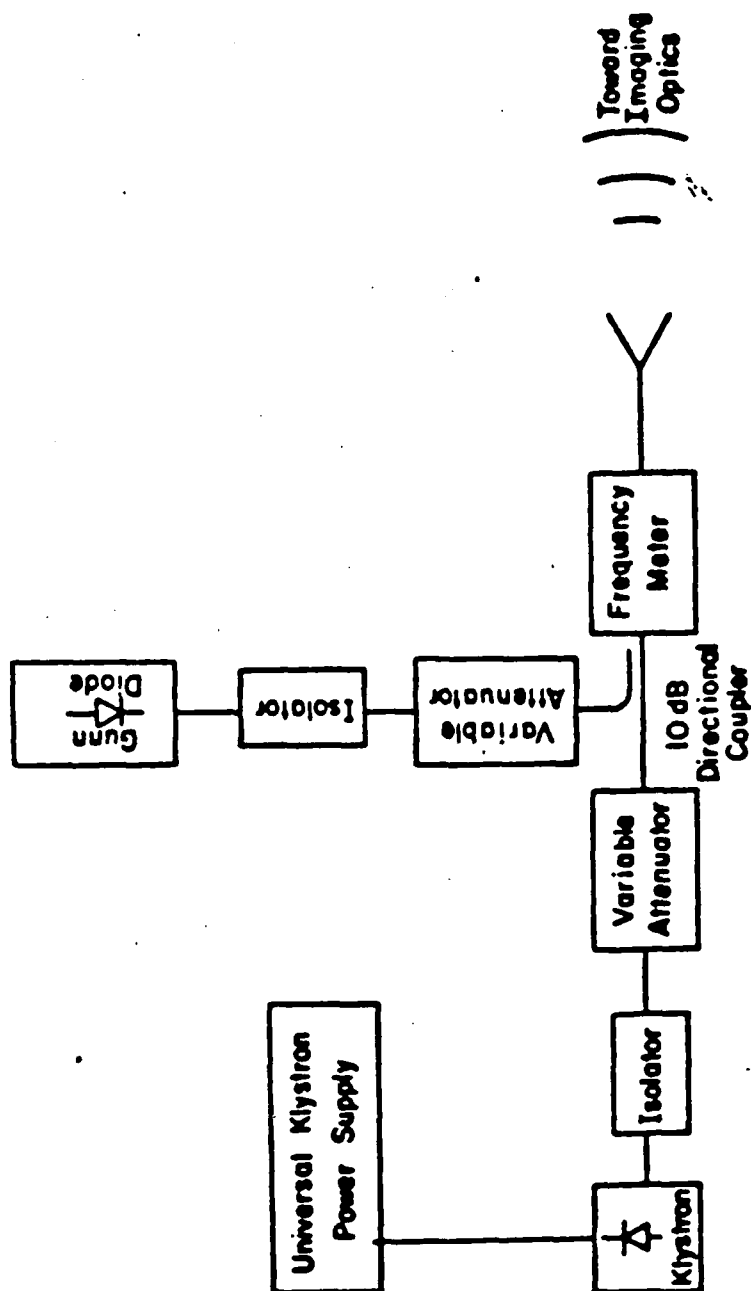


Fig. 5.14

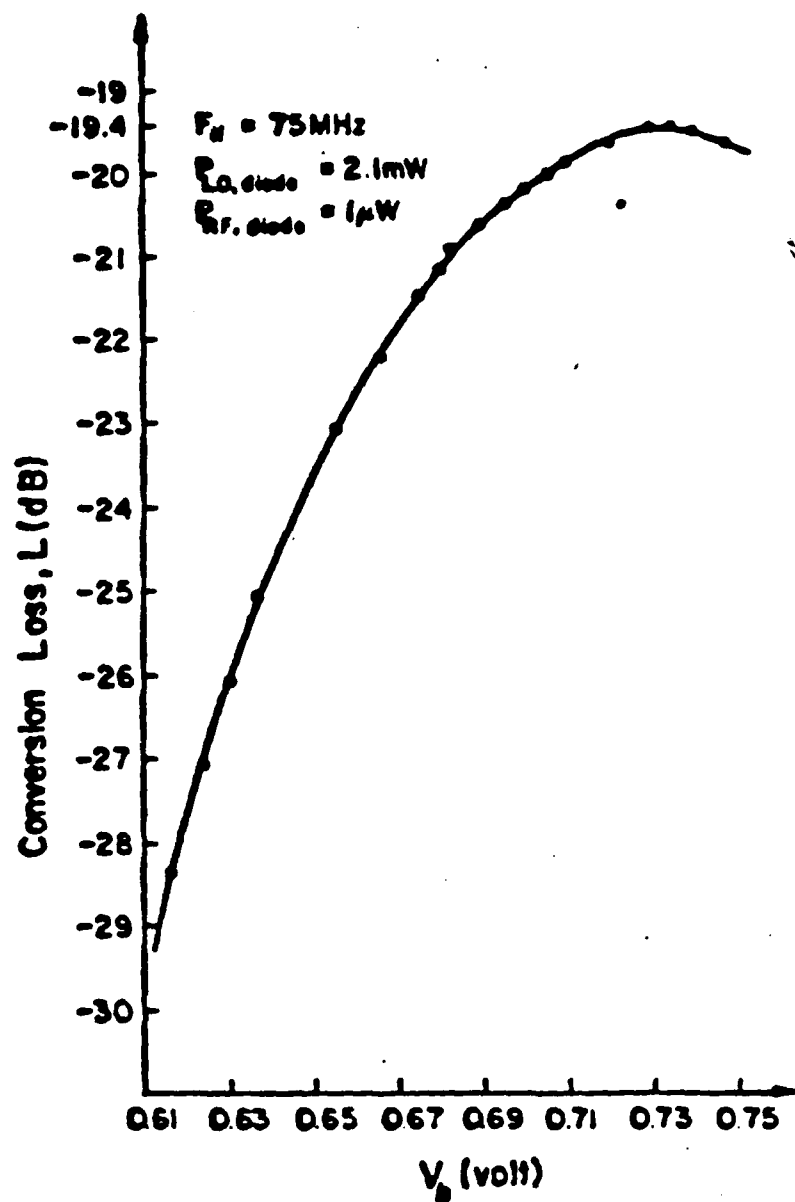


Fig. 5.15

intrinsic diode conversion loss is $L_i = 9$ dB. The average LO impedance is $R_0 = R_s/t = 50 \Omega$. By Fig. 4.10(b) and Fig. 4.10(c), we find that $R_{RF} = 40 \Omega$ and $R_{IF} = 70 \Omega$, respectively. The RF and IF mismatch loss are then calculated to be $L_{RF} = 0.78$ dB and $L_{IF} = 0.1$ dB. Together with the 7 dB optical conversion loss, the theoretical conversion loss L_{th} is

$$\begin{aligned} L_{th} &= L_{op} + L_i + L_{RF} + L_{IF} \\ &= 17 \text{ dB} \end{aligned}$$

There is still a conversion loss of 2.4 dB not accounted for. This is presumably due to the unknown terminations at the harmonic frequencies.

Figure 5.16 shows the change in conversion loss ΔL , as a function of attenuation of the variable attenuator (V/A) of the LO section. At the maximum LO power available to diode of 2.1 mW, the conversion loss is still decreasing at a rate of 0.1 dB change in ΔL per 1 dB increase of LO power at the optimum bias point. It is conceivable that as more LO power becomes available, the conversion loss will decrease and no DC bias will be needed.

5.3.2. NEP

The IF section used in noise measurement is shown in Fig. 5.17. Two MITEQ AU-2A-0110 low noise amplifiers are used as preamplifiers. A 75 MHz band pass filter is used to limit the noise bandwidth. A 16 dB fixed attenuator is used in order to avoid saturation of the final amplification stage (a Watkins-Johnson WJ-6201-344 amplifier). The amplified noise power is measured by a HP 435B power meter. The whole IF section has been calibrated to have a gain of $G = 89$ dB and bandwidth B of 9.65 MHz and a noise figure F of 1.55 dB at a source impedance of 50Ω .

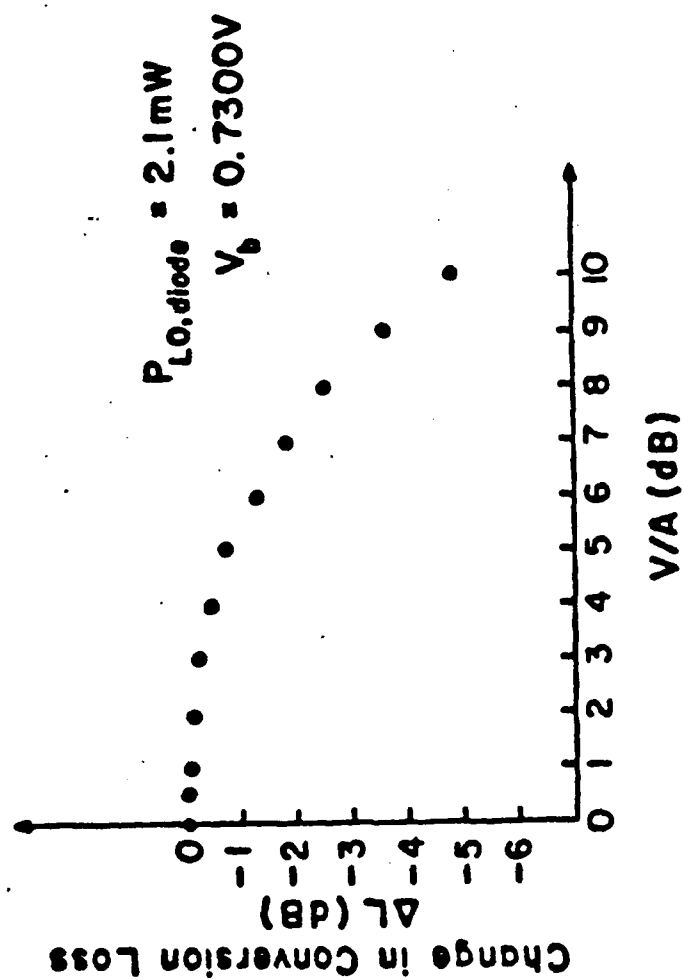


Fig. 5.16

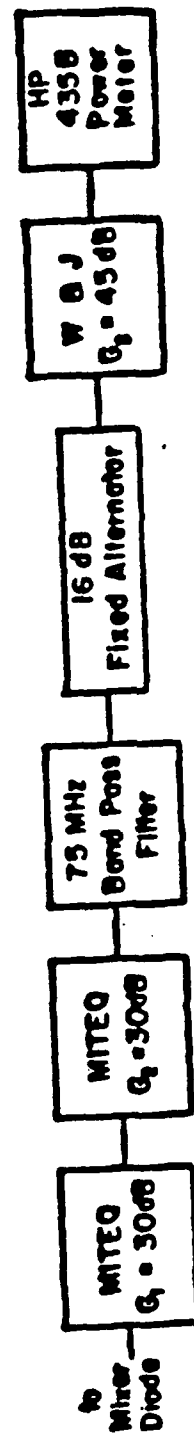


Fig. 5.17

Recall that the SSB and DSB noise equivalent
the noise power measured at the output of the IF section by

$$\text{Measured Noise Power} = \text{NEP}_{\text{SSB}} \text{ GB } [L_{\text{SSB}}]^{-1}$$

and

$$\text{NEP}_{\text{SSB}} = \text{KT}_{\text{SSB}}$$

Since the bow-tie antenna itself is fairly broadband, the mixer acts as a DSB receiver. The SSB and DSB cases, however, are simply related when $L_1 = L_{-1} \gg 1$, which is the case. We can, therefore, characterize the mixer either in terms of the SSB case or of the DSB case, and we have $\text{NEP}_{\text{SSB}} = \text{NEP}_{\text{DSB}}$.

The DSB NEP has been measured as functions of both DC bias and LO power level. The dependence of the DSB NEP on DC bias voltage is shown in Fig. 5.18. The minimum NEP achieved is $7 \times 10^{-19} \text{ W/Hz}$ at a LO power per diode of 2.1 mW and DC bias of 0.69 V. The corresponding DSB system noise temperature is $5.1 \times 10^4 \text{ K}$.

At the optimum NEP condition, the bias voltage is 0.69 V. The corresponding junction resistance R_j is 44.3 Ω . The equivalent diode noise temperature T_{eq} , according to (4.42), is $T_{\text{eq}} = 221.9 \text{ K}$. The mixer output noise temperature T'_M , by (4.43), is $T'_M + 216.8 \text{ K}$. The DSB mixer noise temperature $T_{M,\text{DSB}}$ is then 9440.6 K, according to (4.46). Knowing the noise figure ($F = 1.55 \text{ dB}$) of the IF amplifier, a system noise temperature of $1.44 \times 10^4 \text{ K}$ is obtained through (4.52).

The measured noise temperature is a factor of 3.5 higher than the predicted value from the noise model based on thermal and shot noise. Since the diode operated at a rather high current density, the excess noise mechanisms mentioned in section 4.2.4 are expected to contribute significantly. Unfortunately, these can still not be characterized

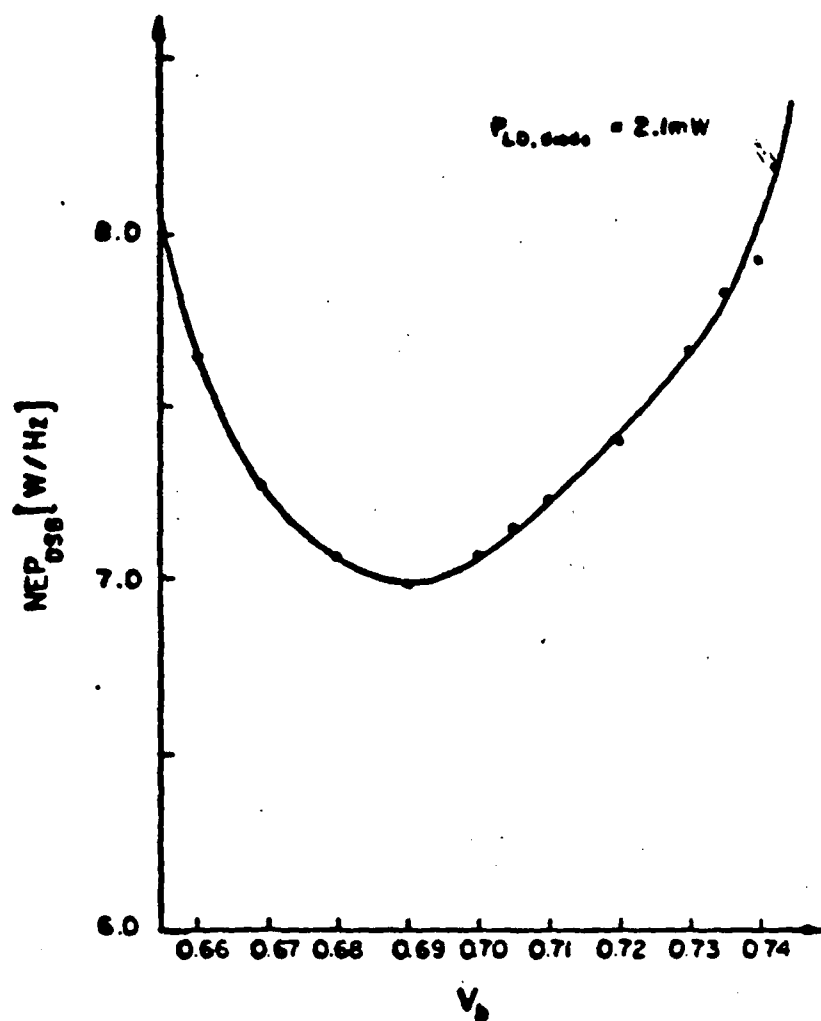


Fig. 5.18

quantitatively. There are two other possible sources for the excess noise measured. Firstly, the local oscillator noise beats with itself and can generate noise at the IF. The noise figure of the IF amplifier may be different from what is measured because the IF source impedance is not known. It is possible to include an impedance matching network between the mixer and the input to the IF section to minimize this portion of noise.

5.4. Optical System Performance

5.4.1. Optical Coupling Loss

From the video detection experiment, we concluded that the optical system coupling loss L_{op} is 7 dB. It is attributed to the reflection and dielectric loss from the TPX objective lens and the substrate lens, and the electromagnetic coupling loss of the antenna itself. Because of its low loss and small index of refraction ($n \approx 1.45$), the loss from the TPX lens is negligible.

We can estimate the reflection loss L_r from the silicon substrate lens by the reflection coefficient of a plane wave incident normally on a semi-infinite silicon. This is a good approximation because the wave impinges on silicon lens at a small angle. The reflection loss is

$$L_r = 1 - |R|^2 = 1 - \left| \frac{\sqrt{\epsilon_r} - 1}{\sqrt{\epsilon_r} + 1} \right|^2 \quad (5.3)$$

$$= -1.6 \text{ dB}$$

where R is the plane wave reflection coefficient, and ϵ_r is 12 for silicon.

Dielectric loss L_d in the silicon substrate lens can be estimated by the known resistivity. Inside the substrate lens, power P depends upon its radial traveling distance z by

$$P = P_0 e^{-\alpha l} \quad (5.4)$$

where

$$\alpha = \sigma \sqrt{\frac{\mu}{\epsilon}} \quad (5.5)$$

and σ is the conductivity of silicon, μ is the free space magnetic permeability and ϵ is the dielectric permittivity of silicon. For the resistivity of silicon equal to 1700 Ω -cm, and a thickness of the substrate lens of 25.85 mm, we have

$$L_d = e^{-\alpha l} = 0.85 \quad (5.6)$$

$$= -0.7 \text{ dB}$$

The dielectric loss and reflection loss can, in principle, be corrected by using lens material of higher resistivity and an anti-reflection coating, respectively.

5.4.2. Electromagnetic Coupling Loss

The remaining 4.7 dB is due to the electromagnetic coupling loss. It includes the mismatch of the antenna feed pattern to the incoming radiation and radiation falling outside of the antenna capture cross section. As is evident from (4.7)-(4.9), as the objective lens $f^\#$ increases, the incidence angle of the radiation θ_1 decreases (Fig. 4.2) and hence the direction of the incident wave is inside the antenna feed pattern. On the other hand, by (4.11), the diffraction-limited spot size Δr_s increases. Since the antenna capture cross section is fixed, less energy is captured. If the $f^\#$ decreases, the diffraction-limited spot size decreases and hence the energy is more concentrated. The incidence angle, however, become so large that the antenna becomes insensitive to it. Therefore, there exists an

optimum $f^\#$ for the objective lens.

To determine the optimum $f^\#$, we measured the relative coupling efficiency η^* by stopping down the diameter of the objective lens. The relative coupling efficiency η^* is the ratio of the voltage responsivity at a fixed bias, to the responsivity when $f^\# = 2$. Since the voltage responsivity of a diode at a given bias voltage is constant, η^* measures the efficiency of the coupling of power available from the objective lens into the antenna. The experimental result is shown in Fig. 5.19.

Let us consider the match of the antenna feed pattern first. The antenna has a main radiation lobe for $\theta_1 \leq 40^\circ$ and the gain is negligible for $\theta_1 \geq 45^\circ$ (Fig. 4.3). We can estimate the $f^\#$ to match the antenna by (4.4). Because of $n_0 \sin \theta_1 = n_1 \sin \theta_0$, for $\theta_1 = 45^\circ$, $n_0 = 1$, and $n_1 = \sqrt{2}$, we have $\theta_0 = 11.8^\circ$. The corresponding f-number is

$$f^\# = \frac{1}{2 \sin \theta_0} = 2.45$$

Indeed, the relative coupling efficiency peaks around $f^\# = 2.5$.

The corresponding diffraction-limited spot size is, by (4.11)

$$\begin{aligned} \Delta r_s &= 1.22 \frac{\lambda f^\#}{n_1} \\ &= 1.08 \text{ mm} \end{aligned}$$

As the f-number decreases, although the antenna pattern becomes mismatched, the energy becomes more concentrated because of the smaller Δr_s . Thus the relative coupling efficiency η^* stays constant with decreasing η^* before it rolls off for $f^\#$ smaller than 2. The fact that the relative coupling efficiency stays fairly constant for $f^\#$ between 2.5 to 4 is remarkable. Especially since the energy density decreases by a factor of $(4/2.5)^2 \sim 2.6$. It finally rolls off for $f^\#$ greater than 4 because of the mismatch between

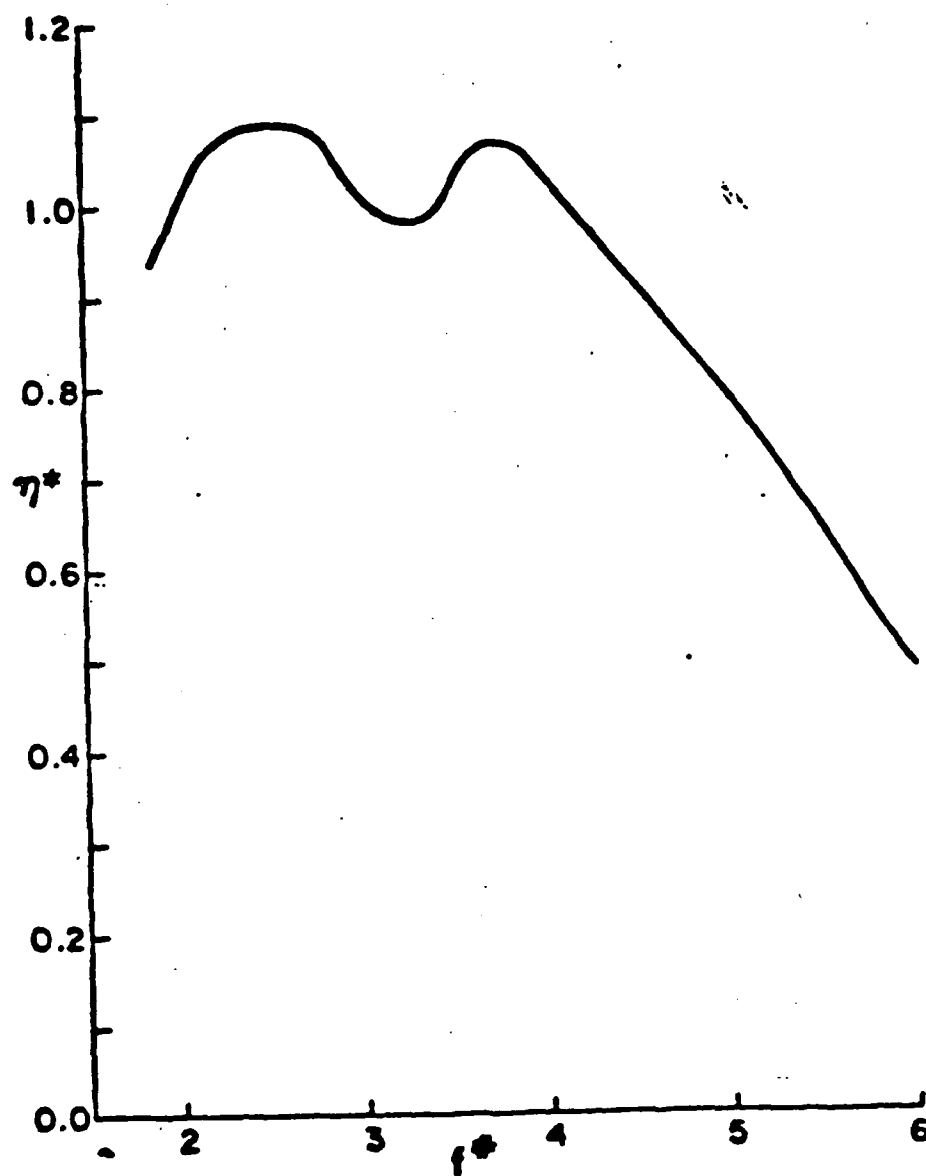


Fig. 5.19

the spot size and antenna capture cross section.

5.4.3. Crosstalk

Crosstalk between antenna elements is another important system characteristic. It is defined as the ratio of the response of the nearest neighbor antenna to the response of the antenna at the center of the image spot for a incident plane wave. It is obviously a function of the objective lens $f^\#$ through its effect on the diffraction-limited spot size. The measurement was again done by stopping down the aperture of the objective lens while measuring the video responsivity of the center antenna and its nearest neighbor simultaneously. The result is shown in Fig. 5.20. The lowest measured crosstalk was found to be 11.6 dB at $f^\#$ of 2.8.

Judging from Figs. 5.19 and 5.20, it appears that the optimum $f^\#$ for the objective lens is 2.8.

5.5. Conclusion

Clearly each individual mixer in these arrays is inferior in performance to a state-of-the-art receiver. However, in imaging applications what matters is the rate of information acquisition. For example, if each element of the array has a heterodyne NEP ten times as large as that of a discrete receiver, an exposure time 100 times as long is needed [see (5.7) below] for each data point. On the other hand, if we are dealing with a 32x32 imaging array, 1024 data points can be collected simultaneously. Thus the time required for scanning an image is ten times less than that which would be required for scanning the discrete receiver through 1024 different positions.

As an example, let us consider the imaging application in radiometry. The minimum detectable temperature change, $NE\Delta T$, is related to the system noise temperature $T_{S,DSB}$ by [20]

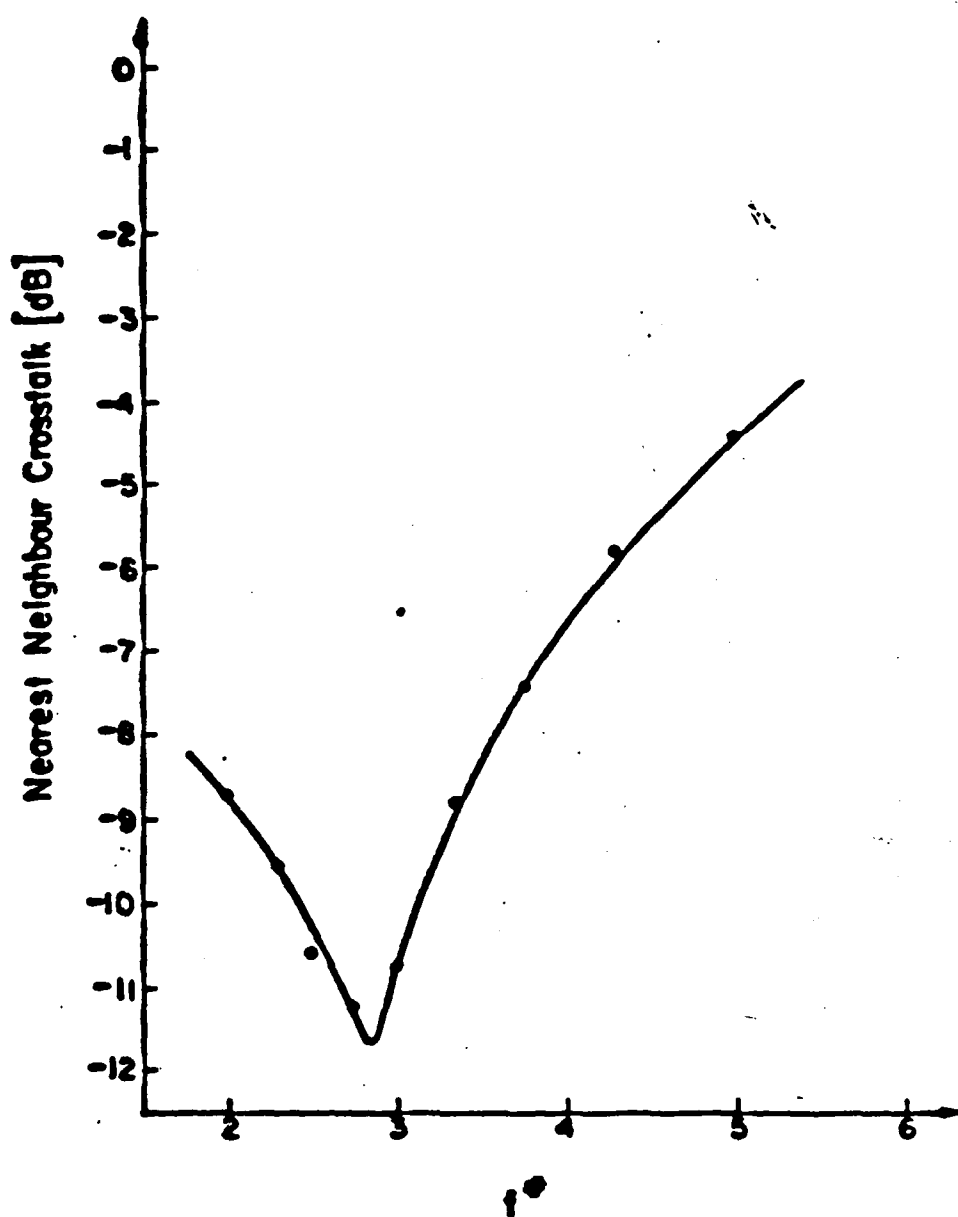


Fig. 5.20

$$NE\Delta T = \frac{\sqrt{\pi} T_{S,DSB}}{2\sqrt{B\tau}} \quad [^{\circ}K] \quad (5.7)$$

where B is the IF bandwidth and τ is the integration time. Suppose the bandwidth is 1 GHz and the integration time is 1 second, the NE Δ T is

$$\begin{aligned} NE\Delta T &= \frac{\sqrt{\pi} 5.1 \times 10^4}{2\sqrt{10^9 \times 1}} \\ &= 1.4^{\circ}K \end{aligned}$$

The measured system noise temperature is used in the above calculation. As noted before, if the excess noise temperature can be eliminated through the use of epitaxial wafers of better quality, etc. the NE Δ T can be reduced to 0.4 $^{\circ}$ K with the present mixer performance. It should also be noted that improvements in performance of several dB should be obtainable through such straightforward means as anti-reflection coating the lenses, optimizing the doping profile of the substrate to reduce the diode parasitics, and correcting the IF mismatch. It is estimated that a DSB conversion loss of 8 dB should be achievable. The NE Δ T in this case can be estimated as follows. Assuming the same IF amplifier is used, the receiver system noise temperature can be calculated by (4.42), (4.43) to be 2044 $^{\circ}$ K. With the same IF bandwidth and integration time, the NE Δ T is 0.06 $^{\circ}$ K. This makes the Schottky diode imaging array promising for imaging applications.

References

- [1] A. G. Foyt, W. T. Lindley, C. M. Wolfe, and J. P. Donnelly, "Isolation of junction devices in GaAs using proton bombardment," *Solid State Electronics*, vol. 12, pp. 209-214, 1969.
- [2] J. P. Donnelly and F. J. Leonberger, "Multiple-energy proton bombardment in n^+ -GaAs," *Solid State Electronics*, vol. 20, pp. 183-189, 1977.
- [3] C. C. Chang, P. H. Citrin, and B. Schwartz, "Chemical preparation of GaAs surfaces and their characterization by Auger electron and x-ray photoemission spectroscopies," *J. Vac. Sci. Tech.*, vol. 14, pp. 943-952, 1977.
- [4] A. K. Sinha and J. M. Poate, "Effect of alloying behavior on the electrical characteristics of n-GaAs Schottky diodes metalized with W, Au, and Pt," *Appl. Phys. Lett.*, vol. 23, pp. 666-668, 1973.
- [5] A. K. Sinha, "Metalization scheme for n-GaAs Schottky diodes incorporating sintered contacts and a W diffusion barrier," *Appl. Phys. Lett.*, vol. 26, pp. 171-173, 1975.
- [6] Y. Armand, A. Christon, and H. Day, "Ideality and noise figure characteristics of r.f. sputtered millimeter GaAs diodes," *Electronics Lett.*, vol. 16, pp. 581-583, 1980.
- [7] C. C. Chang, private communication.
- [8] A. K. Sinha, T. E. Smith, M. H. Reas, and J. M. Poate, "n-GaAs Schottky diodes metalized with Ti and Ti/Pt," *Solid State Electronics*, vol. 19, pp. 489-492, 1976.
- [9] V. L. Rideout, "A review of the theory and technology for ohmic contacts to group III-V compound semiconductors," *Solid State Electronics*, vol. 18, pp. 541-550, 1975.

- [10] M. Wittmer, R. Pretorius, J. M. Mayer, and M-A. Nicolet, "Investigation of the Au-Ge-Ni system used for alloyed contacts to GaAs," *Solid State Electronics*, vol. 20, pp. 433-439, 1977.
- [11] T. J. Nagee and J. Peng, *Phys. Stat. Sol. (a)*, vol. 32, p. 695, 1975.
- [12] M. N. Yoder, "Ohmic contacts in GaAs," *Solid State Electronics*, vol. 23, pp. 117-119, 1980.
- [13] M. Otsubo, H. Kumabe, and H. Mike, "Liquid phase epitaxy growth of GaAs from Au-Ge-Ni melts," *Solid State Electronics*, vol. 20, pp. 617-621, 1977.
- [14] M. Heiblum, M. I. Nathan, and C. A. Chang, "Characteristics of AuGeNi ohmic contacts to GaAs," *Solid State Electronics*, vol. 25, pp. 185-195, 1982.
- [15] R. L. Van Tuyl, V. Kumar, D. C. D'Avanzo, T. W. Taylor, V. E. Peterson, D. P. Hornbuckle, R. A. Fisher, and D. B. Estreich, "A manufacturing process for analog and digital Gallium Arsenide integrated circuits," *IEEE Microwave Theory Tech.*, vol. MTT-30, pp. 935-942, 1982.
- [16] H. Jasik, *Antenna Engineering Handbook*, McGraw-Hill, New York, 1972.
- [17] E. R. Carlson, M. V. Schneider, and T. F. McMaster, "Subharmonically pumped millimeter-wave mixers," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-26, pp. 706-715, 1978.
- [18] S. Deutsch, "Receiver characteristics," in *Handbook of Microwave Measurements*, Third Ed., M. Sucher and J. Fox Ed., Polytechnic Press of the Polytechnic Institute of Brooklyn, New York, 1970.
- [19] E. Santer and G. V. Schultz, "Comparison of methods of sensitivity determination of point-contact diodes at submillimeter wavelength," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-25, pp. 468-470, 1977.
- [20] N. B. Davenport, Jr. and W. L. Root, *Random Signals and Noise*, McGraw-Hill, New York, 1958.

Figure Captions

- 5.1. Geometry and dimensions of the surface-oriented Schottky diode.
- 5.2. Cross-section of the Schottky diode.
- 5.3. A finished Titanium Schottky diode.
- 5.4. A 2x20 monolithic Schottky diode array.
- 5.5. DC I-V characteristics of three adjacent devices on a typical finished chip.
- 5.6. The forward and reverse I-V characteristics of a typical diode.
The reverse breakdown voltage is about 10 V.
- 5.7. Block diagram of the experimental setup for 69 GHz video detection.
- 5.8. The chip mount design. The chip was glued to the pc board from the back side. The ohmic contacts of the diodes were connected together to the ground on the backside of the pc board and then to the aluminum holder. The Schottky contacts were individually bounded to the copper line connected to the inner conductor of the SMA connectors.
- 5.9. The setup of the imaging optics.
- 5.10. Normalized experimental and theoretical voltage responsivity as a function of DC bias voltage for device 20A-D4.
- 5.11. Frequency dependence of the equivalent noise voltage source for two values of DC bias voltage.
- 5.12. DC bias dependence of the cross-over frequency f_0 of (5.1).
- 5.13. Noise equivalent power as a function of DC bias voltage for four different frequencies.
- 5.14. Experimental setup for the 69 GHz heterodyne detection experiment.
- 5.15. DC bias dependence of the mixer conversion loss.

- 5.16. The mixer conversion loss as a function of LO power level at an optimum bias voltage of 0.73 V. The LO power per diode is 2.1 mW when the variable attenuator (V/A) is at 0 dB.
- 5.17. IF section for noise measurement.
- 5.18. DC bias dependence of the DSB NEP. The LO power per diode is 2.1 mW.
- 5.19. The relative optical coupling efficiency as a function of the objective lens f-number $f^\#$.
- 5.20. Dependence of the objective lens f-number $f^\#$ of crosstalk between the center detector and its nearest neighbor.

CHAPTER 6

FUTURE DIRECTIONS

In this thesis, I have discussed the dielectric ridge waveguide approach to the M³IC and a monolithic Schottky diode imaging array. I believe there are good opportunities for further work along these lines.

For the dielectric ridge waveguide-based M³IC, there are three areas worth working on; alternative coupler, further integration, and DRG with substrate. Although the V-coupler works reasonably well, it is desirable to reduce coupling loss further. It is also advantageous to reduce the length of the coupler so that more devices can be built on the same chip. To achieve these objectives, the shape of the coupler has to be optimized. This can probably be done only through carrying out the full-wave analysis of the coupler. Another interesting area is to use DRG to design M³IC. Waveguide devices such as directional couplers, ring resonators, channel-dropping filters can be designed in DRG, and integration of a V-coupler with Schottky diodes have been demonstrated. It is possible to further integrate the waveguide devices, passive semiconductor devices and active semiconductor devices (e.g., Gunn diode) to form useful circuits, such as a multi-channel receiver front end. As was mentioned before, the DRG is most useful at higher end of the millimeter-wave spectrum. As frequency becomes higher, the web will become thinner. At higher frequencies, therefore, it may become necessary to coat the DRG with a supporting material to increase the mechanical strength of the circuit. Since the modal properties of an asymmetric slab guide are different from those of the symmetric slab guide. It is necessary to redesign the DRG.

With regard to Schottky diode imaging arrays, aside from the straightforward improvement of the present imaging array mentioned in section 5.5, three areas come to mind where opportunities for further work exist. First of all, a way of coupling LO power uniformly to each mixer in the array has to be found. In the present arrangement, the LO power is coupled into the mixer through the image-forming lens, and hence the LO power is focused to a single spot only; the array is not uniformly pumped. Direct coupling of LO power from the front side of the wafer is not efficient, because the antenna pattern of the bow-tie antenna on the substrate is highly asymmetric. Until this problem is solved, the imaging array will be useful only for video detection. The second area is to build a subharmonically pumped imaging array. This is especially important at higher frequencies where LO power is difficult to come by. The third area is to raise the level of integration. It is interesting to integrate either the LO, the IF amplifiers or both of them into the same chip. To achieve this objective, ways of coupling the LO in and the IF out of the diode without disturbing the antenna pattern and introducing signal loss have to be worked out.

In any event, the field of M³IC is still young. There are good opportunities for further work in the near future.

END
FILMED

5-86

DTIC