.

PICOSECOND ELECTRICAL WAVEFRONT GENERATION AND PICOSECOND OPTOELECTRONIC INSTRUMENTATION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY.

by
Mark J.W. Rodwell
December 1987

.

.

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

(Principal Advisor)

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Approved for the University Committee on Graduate Studies:

Dean of Graduate Studies

Abstract

While electronic components have been demonstrated over narrow bandwidths at millimeter- wave frequencies, electronic components for broadband and pulse applications have shown much poorer performance. This discrepancy in performance between pulse and narrowband circuits also applies to the related instrumentation. This thesis describes the development of picosecond electronic devices and the related instrumentation.

Picosecond electrical wavefronts are generated by propagation on a GaAs monolithic nonlinear transmission line consisting of a high-impedance coplanar-waveguide transmission line periodically loaded by a series of Schottky contacts. The variation of wave velocity with voltage permits compression of the falltimes of step-functions propagating on these lines, with a minimum compressed falltime set by the diode and periodic-line cutoff frequencies. With appropriate design, the characteristic impedance can be set at 50 ohms, permitting low-reflection interfaces. With first-generation devices, compression of 20 picosecond edges to 7.8 ps has been attained, and 38 ps wavefronts have been compressed to 10 ps on a cascade of two devices. Recently, with second-generation devices, compression from 20 ps to 5 ps and from 30 ps to 7 ps has been achieved. Theory, fabrication, and evaluation of these devices is discussed.

Direct electrooptic sampling is a noncontact test method for measuring the voltage waveforms at the internal nodes of GaAs integrated circuits. The technique exploits the electrooptic effect of the GaAs substrate to obtain a voltage-dependent polarization-modulation of a probe beam passing through the circuit substrate. Sampling techniques result in $\sim$ 100 GHz bandwidth. The factors determining system bandwidth and sensitivity are discussed, and measurement results are reviewed.

To attain picosecond time resolution in electrooptic sampling, the timing fluctuations of the pulsed laser system must be reduced to significantly below 1 ps. Timing fluctuations in the 50 Hz–25 kHz frequency range are reduced from 1.2 ps to 0.24 ps by phase-locking the laser to a precision radio-frequency oscillator. Stabilizer design considerations are discussed.

Acknowledgements

The thesis describes three research topics, all of which depended to a large extent on collaborations with other students and with researchers from outside Stanford. In some cases the projects were extensions of another's work, in other cases the projects drew heavily on the experimental infrastructure developed by other students for other projects.

In the development of the nonlinear transmission line, the biggest bow must go to Chris Madden and the nine months he spent in the microfabrication laboratory. Without him, this project would have failed. Thanks also to Lance Goddard and Tom Carver for understanding our time pressures and handling jobs for us at such short notices. Steve Swierkowski, Kay Mayeda, and Lori Erps at Lawrence Livermore developed workable masks from our rough drawings, guided us extensively through processing, and invested substantial time in a parallel fabrication effort. Yi-Ching Pao, Nancy Gabriel (now at HP), and George Bechtel at Varian III-V device center provided us with the MBE material and gave freely of their considerable expertise in the design and fabrication of microwave diodes. Roe Hemenway and Scott Diamond invested many months in the selection and installation of GaAs processing equipment for the Ginzton microstructures facility, and spent half their vacation in a premature fabrication effort last Christmas.

Kurt Weingarten and Brian Kolner were the leaders on the electrooptic sampling work, and I thank them both. Kurt tolerated a new user on the system who was always most interested in using it for his own measurement needs, and for two years collaborated on microwave amplifier testing and laser timing stabilization. Working with him was tremendous. Steve Swierkowski collaborated on testing some digital circuits and Majid Riaziat and George Zdasuik at Varian Research Center collaborated on testing microwave distributed amplifiers. Majid's coplanar-waveguide distributed amplifiers, combined with some old papers on ferroelectric nonlinear transmission lines found by Bert Auld and Brian Kolner, together seeded the Schottky- diode nonlinear transmission line. Brian suggested stabilizing the laser timing fluctuations by means of a phase-lock-loop and Tom Baer (Spectra-Physics) guided me in interfacing to the laser.

Three members of the faculty guided these projects: Dave Bloom, Bert Auld, and Pierre Khuri-Yakub. Their contributions include conception, guidance with theory and implementation, seaching for literature, help with experiments, and management and financial support of the programs. IBM provided a two-year

fellowship.

Finishing a thesis seemed to demand a singlemindedness bordering on obsession; thanks to my family and friends both for their support and for their patience. Thanks especially to Ginny, who professed to not mind the hours worked, and even provided midnight taxi service from Ginzton, knowing otherwise I would most likely fall off my bike on the way home.

Table of Contents

Part 2: Picosecond Optoelectronic Instrumentation.

Introduction

The development of high-speed digital devices and pulse-waveform analog circuits has recently been spurred by a number of applications, including fiber-optic data transmission at gigahertz rates, high-throughput computing, and wideband signal processing. State of the art digital logic gates exhibit propagation delays and risetimes on the order of 20-100 ps, resulting in maximum clock rates of 4-20 GHz [1], while (with the exception of superconducting devices) analog circuits and devices for impulse and step generation, pulse amplification, and gating, have shown pulse widths, risetimes, and aperture times no shorter than 20 ps, equivalent to 18 GHz bandwidth. In marked contrast to these results, narrowband millimeter- wave circuits performing simple electronic functions (modulation, difference-frequency generation, harmonic generation, and envelope detection) have been demonstrated at frequencies as high as 450 GHz [2]. This discrepancy in performance between pulse and narrowband circuits also applies to the related instrumentation. For example, while sampling oscilloscopes used for waveform and timing measurements in analog and digital systems have risetimes of 18-25 ps, equivalent to bandwidths of only 14-20 GHz, the vector network analyzers used for microwave 2-port parameter analysis are available for waveguide bands as high in frequency as 60-90 GHz.

This thesis describes the development of picosecond electronic devices and the instrumentation used to characterize them. Part 1 of the thesis (chapters 1–4), concerns the design, fabrication and evaluation of Schottky-diode nonlinear transmission lines, monolithic GaAs devices for the generation of step-functions or wavefronts having picosecond transition times. The short, large-amplitude wavefronts generated by these monolithic structures have potential application in picosecond-resolution time-domain electronic instrumentation. Currently, the fastest electronic pulse generators used in sub-nanosecond instrumentation are tunnel diodes and step-recovery diodes. Step-recovery diodes generate $\sim$10 V transitions of 35 ps risetime, while tunnel diodes generate $\sim$ 20 ps edges, but with amplitude ($\sim$0.2 V) insufficient for many applications. The performance of these devices has not improved significantly in the last 20 years. In contrast, generation of $\sim$ 1 ps wavefronts appears to be feasible with nonlinear transmission lines.

A nonlinear transmission line is a waveguide incorporating nonlinear reactive elements which introduce a variation in the propagation velocity with either voltage or current. This variation in velocity results in steepening of either the positive-going or negative-going wavefronts of signals propagating on the line. As the transition

time decreases, wavefront dispersion arising from a variety of effects competes with the wavefront compression arising from the nonlinearity. A final, limited transition time is reached at which these two processes are balanced. While nonlinear transmission lines can be constructed using nonlinear inductance arising from ferromagnetic [3] or superconducting [4] materials or nonlinear capacitance arising from ferroelectric [5] materials, nonlinear transmission lines employing the voltage-variable capacitance of a reverse-biased semiconductor junction (variable-capacitance diode or varactor) are most often discussed in the literature [6]. Despite extensive published mathematical studies of wavefront compression on nonlinear transmission lines, experimental results have been confined to generation of $\sim$ 1ns wavefronts on large-scale experimental devices using discrete diodes [7].

Schottky diodes having terahertz cutoff frequencies are readily integrated on GaAs substrates with transmission lines having low dispersion and low loss at millimeter-wave frequencies; with these parameters, analysis predicts (Chapters 1 and 2) picosecond compressed wavefronts. Initial monolithic GaAs devices built at Ginzton Lab were designed with very low resolution (10μm) design rules; these devices have generated $\sim$ 4 volt transitions with 7.8 picosecond risetimes. Recently, improved monolithic devices fabricated with (3μm) design rules have generated 5 ps wavefronts. Integrated with Schottky-diode sampling bridges, $\sim$100 GHz electrical waveform sampling should be possible with a monolithic room-temperature device.

The second part of the thesis reviews optoelectronic instrumentation; i.e. optoelectronic techniques for measurement of very-high-speed electronics and optics. The field is now in active development, with many competing techniques and many participating research groups. The thesis discusses one technique developed at Stanford: direct electrooptic sampling (chapter 5), a picosecond-resolution internal-node probing technique for GaAs integrated circuits. After $3\frac{1}{2}$ years of development by several graduate students, an instrument has been developed which can *routinely* provide $\sim$ 100 GHz-bandwidth waveform or transfer function measurements, probing small conductors within high-speed gallium arsenide (GaAs) integrated circuits. A research project itself, the electrooptic sampling system was also a critical tool in the development of the nonlinear transmission line.

Among several factors critical to the success of electrooptic sampling was suppression of the timing fluctuations inherent in the mode-locked laser used for time sampling. The very short optical pulses from mode-locked laser systems are widely used in picosecond time-resolved experiments. Most such experiments are performed using a pump-probe technique, whereby the timing fluctuations of the laser system do not influence the measurement. In electrooptic sampling, pump-probe techniques are inappropriate, and laser timing fluctuations directly degrade the time

resolution of the measurement. When these timing variations were first investigated, fluctuations of approximately 10 ps were found in a laser system producing 1.5 ps pulses. With feedback stabilization of the laser and with improvements to the laser system itself, 0.24 ps timing fluctuations are now achieved, and stabilization of the jitter to below 0.1 ps seems feasible. Apart from the data of this thesis on Nd:YAG mode-locked lasers, and data from AT&T Bell Laboratories on semiconductor diode lasers, there is no published data on other mode-locked laser systems. Basic noise theory of free-running and injection-locked oscillators suggests that most mode-locked laser systems will show significant timing jitter; and a timing stabilization system (Chapter 6) will very likely be a necessary component in any picosecond optical probing technique competing with electrooptic sampling.

References

[1] J.F. Jensen, L.G. Salmon, D.S. Deakin, and M.J. Delaney: "Ultra-high speed GaAs static frequency dividers," Technical Digest of the 1986 International Electron Device Meeting, p. 476.

[2] T. Takada and M. Hirayama: "Hybrid Integrated Frequency Multipliers at 300 and 450 GHz", IEEE Trams. on MTT, vol. MTT-26, no. 10, October 1978, p. 733.

[3] C.S. Tsai and B.A. Auld, J. Appl. Phys. **38**,2106 (1967)

[4] A. Scott, *Active and Nonlinear Wave Propagation in Electronics*, Wiley-Interscience, 1970.

[5] Landauer, R. :"Parametric Amplification along Nonlinear Transmission Lines", J. Appl. Phys., 1960, Vol. 31, No. 3, pp. 479-484.

[6] D. Jäger: "Characteristics of travelling waves along the nonlinear transmission lines for monolithic integrated circuits: a review", Int. J. Electronics, 1985, vol. 58, no. 4, pp. 649-669.

[7] M. Birk and Q.A. Kerns: "Varactor Transmission Lines", Engineering Note EE-922, Lawrence Radiation Laboratory, University of California, May 22, 1963.

.

Part 1:
Picosecond Electrical Wavefront Generation.

Chapter 1: Shock Wave Generation on a Distributed Schottky Diode

Nonlinear wave propagation, shock-wave formation, and soliton propagation are observed in a variety of physical systems, including supersonic fluid flow [1], Josephson junction transmission lines [2], and electromagnetic wave propagation in ferromagnetics [3] and ferroelectrics [4]. Scott [2] discusses a variety of electrical systems for nonlinear wave propagation, while Courant [5] gives general methods of solution of the hyperbolic nonlinear partial differential equations describing nonlinear wave propagation.

Transmission lines having a nonlinear shunt capacitance arising from either ferroelectric dielectrics or from shunt loading by voltage-variable capacitors have been proposed by a number of authors [4,6,7,8,10], in papers dating back to the early 1960's. Parametric amplification [9] on nonlinear lines, suggested by in 1960 by Bell and Wade [10], was shown by Landauer in 1960 [4] to be unattainable in nondispersive lines, due to the formation of shock wavefronts. Khokhlov [6] analyzed the propagation of sinusoidal signals along lines with nonlinear shunt capacitance, calculating the shock wavefront transition time for the case where capacitor's series resistance is the dominant limitation. Similar calculations for shock wavefront formation given step-function input signals were reported by Peng and Landauer, 1973 [11]; the transition times were found to be the same as calculated by Khokhlov for sinusoidal excitation.

The potential application of electromagnetic shock-wave formation to electronic switching and signal generation was recognized by Birk and Kerns in 1963 [7], and by Khokhlov, 1960 [6]. Birk and Kerns generated wavefronts of ~ 300 ps risetime using a discrete inductor-varactor diode ladder network, the risetimes being limited by the minimum capacitance feasible in a discrete packaged diode. These transition times were uncompetitive with those generated by compact step-recovery diodes under development in the same period [12].

The recent development of gallium arsenide integrated circuit technology should permit the fabrication of inductor-varactor diode monolithic nonlinear transmission lines and the generation of picosecond wavefronts. A recent body of literature, reviewed by Jäger [13] has concentrated on monolithic varactor transmission lines formed by microstriplines or coplanar waveguides loaded with a continuous or interrupted Schottky contact. Despite the potential of picosecond step-function formation, the papers reviewed all concentrate on parametric amplification, harmonic generation, or slow-wave propagation, although Jäger [14] discusses shock-wave for-

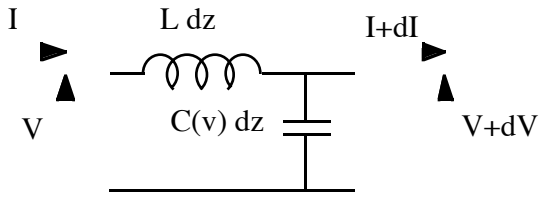


Figure 1.1: Equivalent circuit of Idealized Nonlinear Transmission Line.

mation without reference to the wavefront transition time.

The analysis begins by studying wavefront compression on an idealized nonlinear transmission line whose circuit model consists of a uniformly distributed (i.e. continuous) nonlinear shunt capacitance and a uniformly distributed linear series inductance. The resulting wavefront transition times will then be calculated for the case where the series resistance of the nonlinear capacitance is the dominant limitation; the analysis projects feasible transition times below 1 ps. The correspondence between a transmission line loaded with a continuous Schottky contact and the idealized circuit model will then be examined critically, and two difficulties with fully-distributed Schottky contact transmission lines will be discovered.

1.1 Nonlinear Propagation and Wavefront Compression

The equivalent circuit of a differential element dz of the transmission line is shown in Fig 1.1. Assuming that $V(z,t)$ and $I(z,t)$ are differentiable *single-valued* functions of (z,t), and writing nodal equations,

$$\frac{\partial I}{\partial z} + \frac{dQ}{dV}\frac{\partial V}{\partial t} = 0 \ , \tag{1.1}$$

and

$$\frac{\partial V}{\partial z} + L\frac{\partial I}{\partial t} = 0 \ , \tag{1.2}$$

where L is the inductance, $Q(V)$ is the charge per unit length of transmission line, and the shunt capacitance is defined by $C(V) \equiv dQ/dV = Q_V(V) = Q_V$. For a Schottky contact to a uniformly-doped semiconductor, $C(V) = C_{j0}/\sqrt{1 - V/\phi}$, where ϕ is the junction potential. Equations (1.1) and (1.2) are written more compactly as $I_z + Q_V V_t = 0, V_z + LI_t = 0$. This notation will be adopted whenever partial and total derivatives can be distinguished by context.

As written, the relations above are nonlinear differential equations giving $(V, I) = \bar{f}(z,t)$. Following a similar derivation by Courant [5], the equations can be solved

by the method of characteristics, as illustrated below. Form 2 linear combinations of Eqs. (1.1) and (1.2) :

$$I_z + \lambda_1 L I_t + \lambda_1 V_z + Q_V V_t = 0 \ , \tag{1.3}$$

$$I_z + \lambda_2 L I_t + \lambda_2 V_z + Q_V V_t = 0 \ . \tag{1.4}$$

These equations relate the derivatives of I and V with respect to two independent parameters z and t. By introducing an auxiliary pair of variables, (α, β), each of the two equations is reduced to a form having derivatives with respect to only one independent parameter. The two independent variables (z, t) are varied as nonlinear functions of the parameters $(z, t) = \bar{g}(\alpha, \beta)$. Both the nonlinear mapping and the weighting parameters $\lambda_{1,2}$ are chosen so that in each equation all derivatives of I and V are with respect to a single parameter, α for Eq. (1.3), β for Eq. (1.4). This requires

$$1 : \lambda_1 L = \lambda_1 : Q_V = z_\alpha : t_\alpha \ ,$$

$$1 : \lambda_2 L = \lambda_2 : Q_V = z_\beta : t_\beta \ ,$$

where the subscripts α and β denote derivatives. Hence the nonlinear mapping is

$$\begin{aligned} \delta t =& \sqrt{LQ_V}\delta\alpha - \sqrt{LQ_V}\delta\beta \\ \delta z =& \delta\alpha + \delta\beta \ , \end{aligned}$$

or

$$\begin{aligned} 2\delta\alpha =& \delta z + \delta t/\sqrt{LQ_V} \\ 2\delta\beta =& \delta z - \delta t/\sqrt{LQ_V} \ , \end{aligned}$$

and the parameters are

$$\lambda_1 = \sqrt{Q_V/L} \qquad \lambda_2 = -\sqrt{Q_V/L} \ .$$

When these substitutions are made, two wave equations are obtained,

$$I_\alpha = -\sqrt{Q_V/L}\, V_\alpha \quad \text{and} \quad I_\beta = \sqrt{Q_V/L}\, V_\beta \ . \tag{1.5}$$

That is, along lines of constant α or constant β, the total derivative of I is a simple function of the total derivative of V; note that the mapping is a function of the

solution $V(\alpha, \beta)$. No additional constraints are placed on $V(\alpha, \beta)$, and one solution to Eq. (1.5) is thus

$$V = V(\beta) \quad \text{and} \quad 2\delta\beta = \delta z - \delta t/\sqrt{LQ_V} \ .$$

Given this solution, examine a point of constant β; because $V = V(\beta)$, this is also a point of constant voltage V. Hence, for each β, $2\beta = z - t/\sqrt{LQ_V}$. Points of constant amplitude of the waveform are thus points for which $z - t/\sqrt{LQ_V}$ is constant, and the solution represents a wave in which each individual point on the waveform has a voltage-dependent propagation velocity $u(V) = 1/\sqrt{LC(V)}$:

$$V(z,t) = V_0(t - z\sqrt{LQ_V}) = V_0(t - z\sqrt{LC(V)}) \ . \tag{1.6}$$

Integrating Eq. (1.5) with respect to β, shows that the wave current and wave voltage are related by a voltage-dependent characteristic impedance $Z_0(V) = \sqrt{L/C(V)}$: that is,

$$I(\beta) = I_0 + \sqrt{Q_V/L}\, V(\beta) \ . \tag{1.7}$$

If $C(V)$ increases with V, the variation in propagation velocity with voltage will result in compression of negative-going wavefronts propagating on the line. As an input signal $V_0(t)$, a falling step-function with initial voltage v_h, final voltage v_l, and falltime $T_{f,in}$, propagates along the line, the falltime $T_f(z)$ will at first decrease linearly with distance:

$$T_f(z) = T_{f,in} - z\Big(\sqrt{LC(V_h)} - \sqrt{LC(V_l)}\Big) \ . \tag{1.8}$$

For sufficiently large z, Eqs. (1.6) and (1.8) predict that $V(z,t)$ will become multi-valued and the falltime will become negative (Fig. 1.2). At the point where $T_f(z) = 0$, $V(t)$ becomes discontinuous and the change in variables used in the method of characteristics becomes degenerate. Equations (1.6) and (1.8) then apply only on the continuous portions of $V(z,t)$. As shown by a simple integration of Eqs. (1.1) and (1.2) across any discontinuities in $V(z,t)$ [11] , the discontinuities, called *shock fronts*, propagate at a velocity $u = 1/\sqrt{LC_{ls}}$, where $C_{ls} = (Q(V_l) - Q(V_h))/(V_l - V_h)$ is the large-signal capacitance. Over any discontinuity, $\Delta V/\Delta I = Z_{ls}$, where $Z_{ls} \equiv \sqrt{L/C_{ls}}$ is a large-signal wave impedance. A corrected form of Eq. (1.8), giving the evolution of step-function falltimes on an idealized nonlinear capacitance transmission line is then

$$T_f(z) = \max \begin{cases} T_{f,in} - z\Big(\sqrt{LC(V_h)} - \sqrt{LC(V_l)}\Big) \\ 0 \end{cases} \ , \tag{1.9}$$

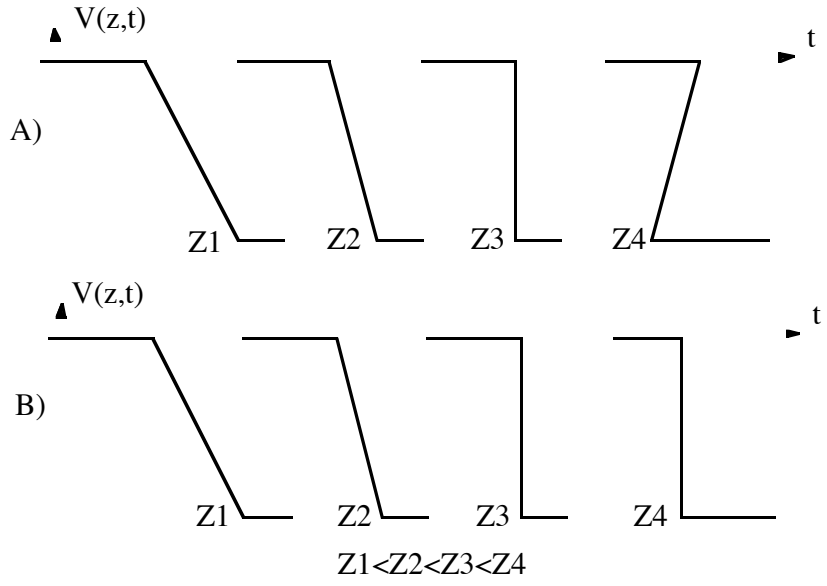


Figure 1.2: Wavefront compression on nonlinear transmission line. A) Propagation and wavefront compression as per Eq. 6, resulting in multi-valued waveforms. B) After formation of a shock wavefront the wavefront propagates unchanged, as per Eq. 9.

where $C(V) = C_{j0}/\sqrt{1 - V/\phi}$ is the capacitance of a step-junction diode.

1.2 Diode Cutoff Frequency as a Limitation to Transition Time

Various parasitic elements result in a nonzero minimum compressed falltime. Significant among these are inductance and resistance in series with the nonlinear shunt capacitance, and skin impedance and dispersion of the transmission structure. For ferroelectric nonlinear transmission lines , waveguide dispersion can be eliminated by use of a line with uniform dielectric, there is no parasitic inductance, and the dielectric relaxation time appears in the equivalent circuit as a resistance in series with the nonlinear capacitance. The skin effect, while possibly significant, is difficult to treat analytically in these nonlinear propagation problems, thus theoretical investigations [4,6] of transmission lines with nonlinear capacitance have concentrated on the influence of the parasitic series resistance.

For the varactor-diode nonlinear transmission lines described in this thesis, several effects can be equally significant in setting the minimum compressed falltime. In the early calculations on the feasibility of generation of picosecond transients, skin effect, waveguide dispersion, and (for periodic structures, Ch. 2) periodicity

effects were regarded as parameters to be minimized by appropriate design, while the varactor-diode $1/r_{diode}C(V)$ cutoff frequency, set by minimum dimensions and maximum doping, was viewed to be a fundamental limitation. Under these highly simplified conditions, the minimum compressed falltime can be calculated as a function of diode cutoff frequency. The analysis below follows the method of Peng and Landauer [11].

The equivalent circuit (Fig. 1.3a) of a differential element of the line now includes a resistance in series with the diode, and the nodal equations are

$$V_z + LI_t = 0 \quad I_z + Q_t = 0 \quad V = V^c(Q) + RQ_t \ . \tag{1.10}$$

For $V^c(Q)$ the Q-V characteristics are assumed as in a reverse- biased Schottky diode whose anode is connected to the transmission line signal conductor and whose cathode is connected to ground (Fig. 1.3b): $\phi - V^c(Q) = kQ^2$ where Q is negative. Both V and Q are assumed to be always negative. The diode zero-bias capacitance is $C_{jo} = 1/2\sqrt{\phi k}$. To calculate the final profile of the shock front, it is assumed that the waves are of constant profile and hence $Q = Q(\zeta)$, $V = V(\zeta)$, and $I = I(\zeta)$, where $\zeta = z - ut$, and u is the shock front propagation velocity. Therefore

$$V_\zeta = Lu^2 Q_\zeta \ . \tag{1.11}$$

Assuming that the propagating signal (Fig. 1.4) is a step-function with zero initial voltage, a final (low and negative) voltage V_l, initial charge $-\sqrt{\phi/k}$, and final (low) charge Q_l, Eq. (1.11) is integrated from ζ to $+\infty$:

$$V = Lu^2(Q + \sqrt{\phi/k}) \tag{1.12}$$

Integration across the shock front from $\zeta = -\infty$ to $\zeta = +\infty$ gives (after some manipulation) the shock-front velocity $u = \sqrt{1/LC_{ls}}$, where the large-signal capacitance

$$C_{ls} \equiv \frac{\Delta Q}{\Delta V} = 2C_{j0}\frac{\phi}{-V_l}\left(\sqrt{1 - V_l/\phi} - 1\right) \tag{1.13}$$

is defined as in the previous section. Substitution of Eq. 1.12 into the third part of Eq. 1.10 then yields

$$\frac{\phi}{k} - \frac{Lu^2}{k}\sqrt{\frac{\phi}{k}} - Q\frac{Lu^2}{k} - Q^2 = \frac{Ru}{k}Q_\zeta \ ,$$

which can be rearranged as a tabulated integral:

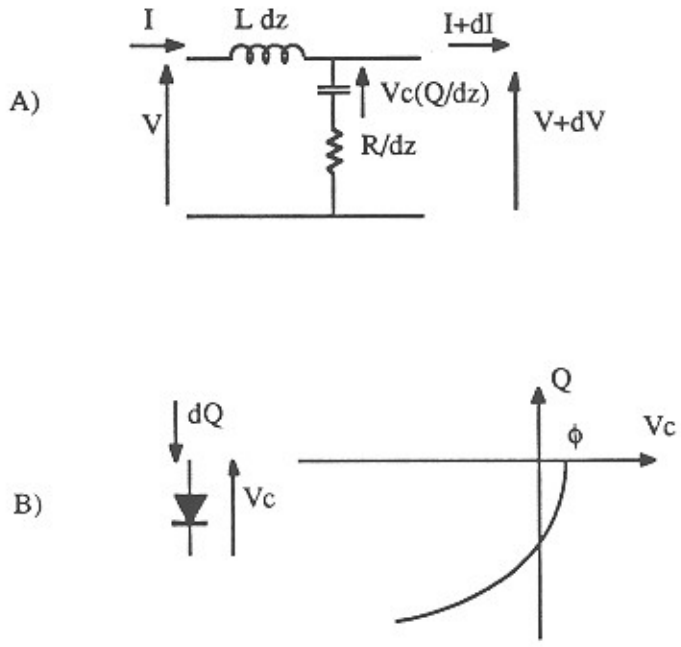


Figure 1.3: Equivalent circuit including diode series resistance, A) and Q-V characteristics of diode, B).

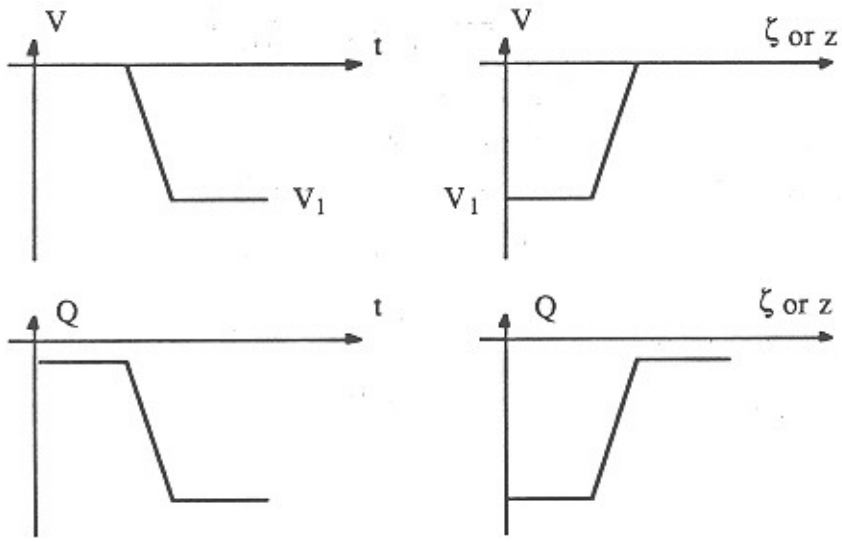


Figure 1.4: Shock wavefront charge and voltage profiles.

$$\zeta = \int_{Q_l}^{Q(\zeta)} \frac{\frac{Ru}{k}\, dq}{\frac{\phi}{k} - \frac{Lu^2}{k}\sqrt{\frac{\phi}{k}} - q\frac{Lu^2}{k} - q^2}$$

After much manipulation,

$$\begin{aligned} V(z,t) =& \frac{V_l}{2} + \frac{V_l}{2}\tanh\left(\frac{t - z/u}{\tau}\right) \\ I(z,t) =& I_0 + \frac{V(z,t)}{Z_{ls}} \end{aligned} \quad , \tag{1.14}$$

where

$$u = \frac{1}{\sqrt{LC_{ls}}} \qquad Z_{ls} = \sqrt{\frac{L}{C_{ls}}} \ ,$$

$$\tau = 4RC_{j0}\frac{1}{\sqrt{1 - V_l/\phi} - 1} = \frac{4}{\omega_{diode}}\frac{1}{\sqrt{1 - V_l/\phi} - 1} \ ,$$

and ω_{diode} is the zero-bias RC cutoff frequency of the diode. The 10%-90% risetime of $\tanh(t/\tau)$ is 2.2τ, hence

$$T_{f,90\%-10\%} = \frac{8.8}{\omega_{diode}}\frac{1}{\sqrt{1 - V_l/\phi} - 1} = \frac{1.4}{f_{diode}}\frac{1}{\sqrt{1 - V_l/\phi} - 1} \ . \tag{1.15}$$

Under the assumption that series resistance is the dominant limitation, Eq. (1.15) permits calculation of the minimum compressed falltime as a function of diode cutoff frequency and breakdown voltage. Discrete millimeter-wave step-junction varactor diodes have been fabricated with 15 V reverse breakdown voltage and 13 THz cutoff frequencies [15], while $\sim$ 10V diodes having 1-2 THz cutoff frequencies are attainable in monolithic form. For most GaAs Schottky diodes, $\phi \cong 0.8$V. Feasible compressed falltimes are then

$$T_{f,90\%-10\%} \cong \begin{cases} 0.25 \text{ to } 0.50\text{ps}, & \text{monolithic diodes;} \\ 32\text{fs}, & \text{best reported diodes.} \end{cases}$$

These transition times are two to three orders of magnitude smaller than the transition times (risetimes) generated by tunnel diodes and step-recovery diodes. Wavefront compression on varactor nonlinear transmission lines thus justifies a more careful investigation. In particular, the details of the device structure have been neglected and it was assumed that the varactor's nonlinear shunt capacitance could be continuously incorporated into a transmission line. Consideration will be given to a fully-distributed varactor-diode nonlinear transmission line, i.e., a microstripline

or coplanar waveguide on a semiconductor substrate in which the signal conductor makes a continuous Schottky contact to the substrate. This approach will be rejected.

1.3 On the Undesirability of a Fully Distributed Varactor Diode

How is the nonlinear diode capacitance incorporated into a transmission- line structure? In the case of ferroelectric nonlinear transmission lines, the dielectric of a linear transmission line is simply replaced with the ferroelectric material. For a varactor-diode nonlinear transmission line it is natural to consider wave propagation on an extended Schottky contact . These structures are microstrip or coplanar waveguide transmission lines on GaAs where the signal line and ground plane form (respectively) Schottky and ohmic contacts to the lightly-doped substrate. Slow-wave propagation on extended Schottky contacts has been considered by a number of authors [13,16,17] for microwave applications. Extended Schottky contacts may be suitable for shock-wavefront generation by nonlinear propagation but they have several disadvantages: the wave impedance is very low, and at high frequencies skin effect in the semiconductor layers introduces loss and dispersion.

1.3.1 Skin Effect in the Semiconductor Layers.

Fig. 1.5 shows the cross-sections of two proposed fully distributed nonlinear transmission lines. The first of these, in Fig. 1.5a , is a coplanar waveguide transmission line in which the center conductor makes a Schottky contact to a N- layer, with the two surrounding ground planes making ohmic contacts to a highly conductive buried N+ layer which serves as the diode cathode connection. With typical GaAs fabrication procedures, the minimum lateral dimensions of this structure would be 1 to 2 μm. Figure 1.5b shows a cross-section of a varactor-diode nonlinear transmission lines in which the metallic interconnections approximate a microstrip transmission line. Since GaAs wafers cannot be reliably thinned to less than $\sim 100\mu$m because of breakage, the substrate is heavily doped (N+ layer) to reduce the resistance in series with the semiconductor depletion-layer capacitance.

Upon first inspection of the nonlinear transmission lines of Fig. 1.5, it might be assumed that the incremental series inductance is that of the uncontacted transmission line (i.e., a similar metallic line on an insulating substrate), and that the incremental shunt capacitance is the Schottky diode depletion-layer capacitance. Wave propagation would then be described by Eq. 1.10, and sub-picosecond wave-

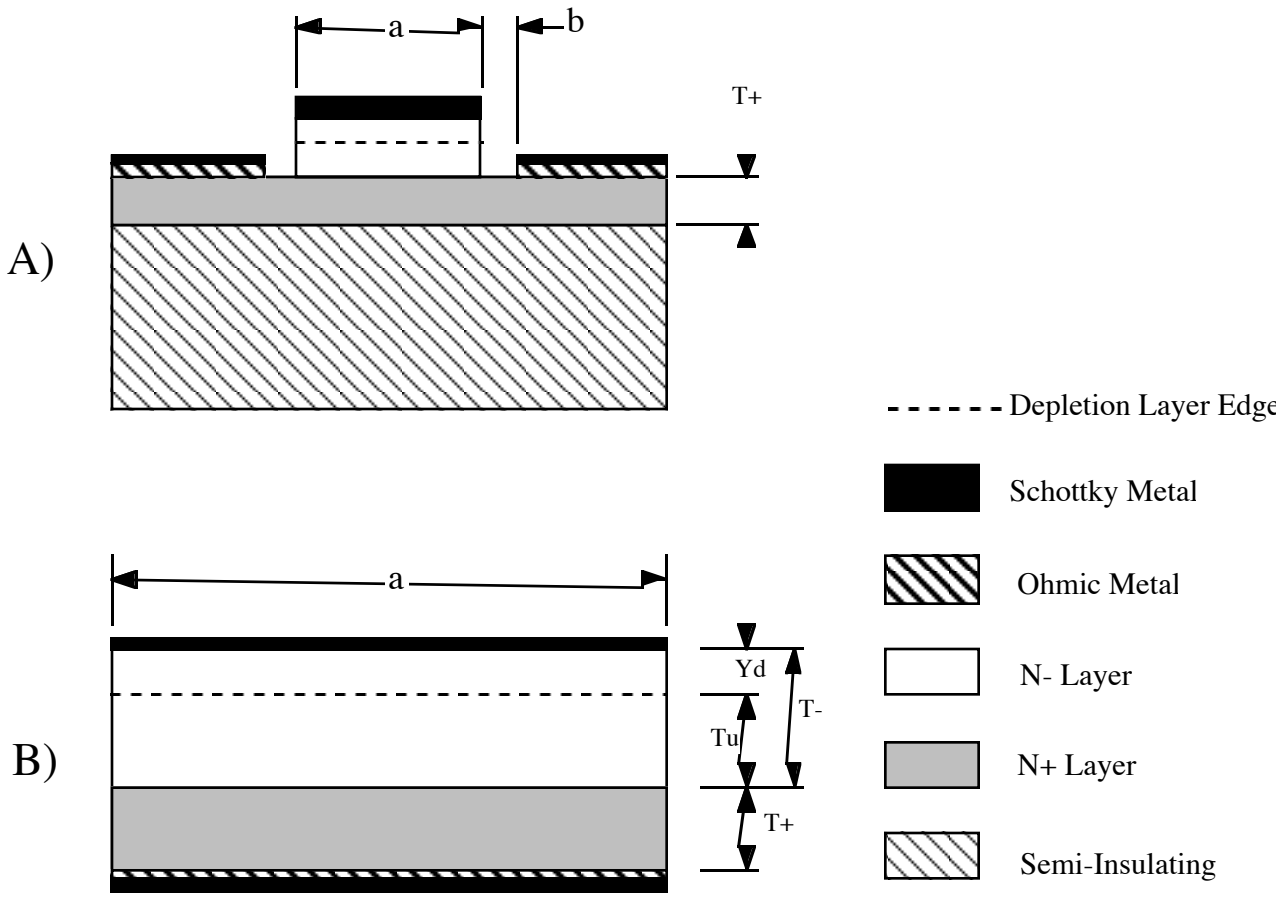


Figure 1.5. A): Cross-section of distributed varactor diode incorporated into a coplanar waveguide transmission line. B): Cross-section of a microstrip transmission line loaded with a distributed varactor diode. T- and T+ denote the thicknesses of the N-and N+ layers, Tu denotes the thickness of the undepleted portion of the N- layer, and Yd denotes the depletion layer thickness.

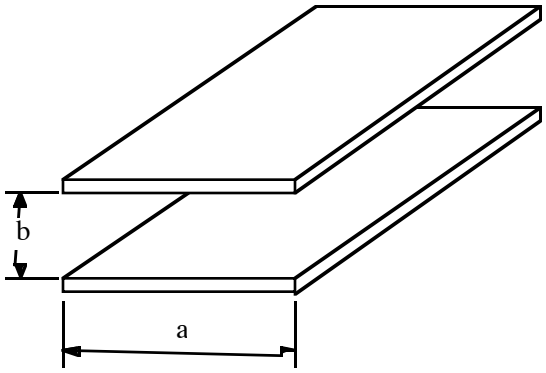


Figure 1.6: Simple parallel-plate transmission line.

front generation would be feasible. These assumptions concerning the equivalent circuit would be correct if the currents flowing through the semiconductor layers were strictly traverse to the direction of propagation, and if the longitudinal currents were confined to the metal interconnections. However, because of its nonzero conductivity, some fraction of the total longitudinal current flows in the N+ layer. At higher frequencies the N+ layer skin depth decreases, resulting in decreased current in the metallic ground conductor, and increased current in the N+ layer, resulting in increased ohmic loss and reduced line impedance. Distributed Schottky diodes investigated for slow-wave propagation and microwave phase modulation have exhibited small-signal attenuation in excess of 5dB/mm at 20GHz [16,17].

While analysis of the skin effect in the coplanar structure is difficult, exact analysis is possible for a parallel-plate geometry, and the equivalent circuit and the physical interpretation are simple if the N+ layer is thin in comparison with the skin depth. Consider first the parallel-plate transmission line of Fig. 1.6. If the plate width is much greater than the plate separation (i.e.: $b \ll a$) then the series inductance L and the shunt capacitance C per unit length are $L = b\mu/a$ and $C = a\epsilon/b$. The equivalent circuit is as in Fig. 1.1, with $C(v) = a\epsilon/b$ being voltage-independent. The permittivity of the surrounding dielectric is denoted by ϵ.

Consider now the nonlinear transmission line structure of Fig. 1.5b. Neglecting longitudinal currents in the undelpeted portion of the N- layer, the equivalent circuit of the transmission line consists of the series inductance $L_{N-} = T_-\mu/a$ of a parallel-plate transmission line formed between the Schottky contact and the N-/N+ interface, the shunt capacitance $C_{depl} = a\epsilon/Y_d$ of a parallel-plate transmission line formed between the Schottky contact and the depletion edge, and the series surface impedance of the N+ layer arising from the skin effect. ϵ now denotes the permittivity of the (GaAs) substrate. With skin effect neglected in the metal con-

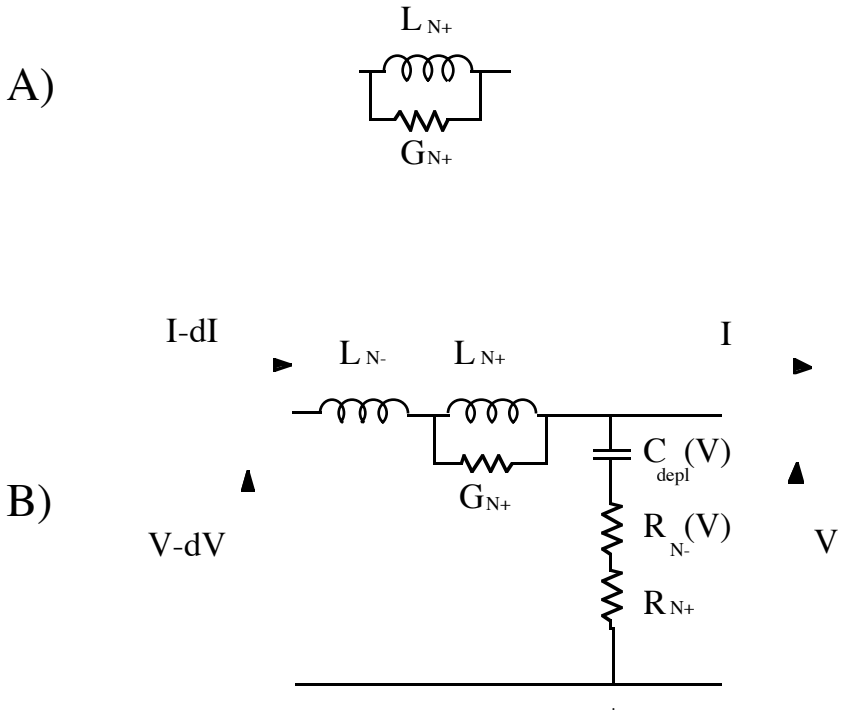


Figure 1.7: Equivalent circuits representing the skin impedances of the N+ layer in a fully distributed varactor diode transmission line. A) Surface impedance of the N+ layer. C) Equivalent circuit of distributed varactor line.

ductors, the surface impedance seen by a TEM wave propagating longitudinally between the Schottky contact and the N-/N+ interface is

$$Z_{skin} = \frac{\sinh(\kappa T_+)/a}{(\kappa/j\omega\mu) + (\sigma_-/\kappa)[\cosh(\kappa T_+) - 1]} \ , \tag{1.16}$$

where T_+ is the thickness of the N+ layer and σ_+ its conductivity, and $\kappa = (1 + j)/\delta_+$, where $\delta_+ = \sqrt{2/\omega\mu\sigma_+}$ is the skin depth. Typical N+ layer dopings for millimeter-wave diodes are 10^{18}–$10^{19}/cm^3$, resulting in skin depths of $\sim$ 5–10μm at 100GHz. For the microstrip case of Fig. 1.5b, the N+ layer thickness will be $\sim$ 25– 100μm. Although only at low frequencies will the N+ skin depth be small in comparison with the N+ layer thickness, the approximation is helpful in understanding the wave propagation. Equation 1.16 can then be expanded to second order in κT_+:

$$Z_{skin} \cong \frac{1/a}{1/j\omega\mu T_+ + \sigma_+ T_+} = j\omega\mu T_+/a \parallel 1/a\sigma_+ T_+ = j\omega L_{N+} \parallel 1/G_{N+} \ . \tag{1.17}$$

The series surface impedance of the N+ layer is thus a parallel combination (Fig. 1.7a) of an inductance $L_{N+} = \mu T_+/a$ and a conductance $G_{N+} = a\sigma_+ T_+$. From Fig. 1.6, L_{N+} is recognized as the series inductance of a parallel-plate transmission line between the N-/N+ interface and the ohmic contact, while G_{N+} is the conductance of an N+ doped layer of dimensions $T_+ \times a$.

A full equivalent circuit (Fig. 1.7b) also includes the traverse resistance of the N- and N+ layers in series with the depletion capacitance: $R_{N-} = T -_u /a\sigma_-$ and $R_{N+} = T_+/a\sigma_+$. Note that because the depletion-layer width $Y_d(V_c) = \sqrt{2\epsilon(\phi - V_c)/qN_{d-}}$ varies with the voltage V_c across the depletion layer, both C_{depl}, and R_{N-} vary with V_c:

$$C_{depl}(V) = a\sqrt{\frac{q\epsilon N_{d-}}{2(\phi - V_c)}} \qquad R_{N-}(V) = \frac{1}{a\sigma_-}\left(T_- - \sqrt{\frac{2\epsilon(\phi - V_c)}{qN_{d-}}}\right) \tag{1.18}$$

The equivalent circuit permits calculation and interpretation of the line's small-signal propagation characteristics. Both the N+ layer parasitic conductance G_{N+} and the diode series resistance introduce loss: for $\omega \ll \omega_{diode},\ \omega_+$, the N+ layer introduces a small- signal attenuation $\alpha_{semiconductor}$ increasing as the square of frequency:

$$\alpha_{semiconductor} \simeq \omega^2\sqrt{(L_{N-} + L_{N+})C_{depl}}\left(\frac{L_{N+}G_{N+}}{2} + \frac{(R_{N-} + R_{N+})C_{depl}}{2}\right) \tag{1.19}$$

At a critical frequency $\omega_+ = 1/L_{N+}G_{N+}$ the longitudinal current redistributes itself into the N+ layer. For an N+ layer having $10^{19}/cm^3$ doping, $\omega_+ \cong 3$THz for $T_+ = 1\mu$m and $\omega_+ \cong 130$GHz for $T_+ = 5\mu$m. At frequencies well below ω_+, the equivalent circuit is as in Fig. 1.3, where the series inductance $L = L_{N-} + L_{N+}$ is that of a parallel-plate transmission line between the Schottky and Ohmic metallizations. The structure then behaves as analyzed in the previous sections. At frequencies ω approaching ω_+, the skin effect in the N+ layer introduces increasing dispersion and loss, while for frequencies $\omega \gg \omega_+$, the ground current becomes confined to a layer of thickness $\delta_+ = \sqrt{2/\omega\mu\sigma_+}$ at the N+/N- interface. Substantial loss results from the series resistance of the N+ layer , and the total transmission-line series inductance $L = L_{N-}$ is that of a parallel-plate transmission line between the Schottky metallization and the N-/N+ interface. The line impedance has dropped, the phase velocity has increased, and loss has been introduced.

With less accuracy, the undepleted portion of the N- layer can be similarly treated. A redistribution of the longitudinal current from the N-/N+ interface to the

depletion edge will occur at a frequency $\omega_- = 1/L_{N-}G_{N+}$, where $L_{N-} = \mu T_u/a$ and $G_{N-} = a\sigma_- T_u$. For an N- layer of 0.5μm thickness and $10^{17}/cm^3$ doping, typical of a millimeter-wave diode, $\omega_- \simeq 5 \times 10^{14}$ Hz; unless the N- layer is extraordinarily thick, its skin effect can be neglected.

While calculation of the exact shock-front profile for the fully distributed structure is intractable, it is probable that the strong loss and dispersion occurring at frequencies approaching ω_+ will result in shock wavefront transition times $T_{f,min} \gg 1/\omega_{N+}$: shock-front formation and profile for a similar but simpler case is discussed briefly in Peng and Landauer [11]. For microstrip nonlinear transmission lines of practical wafer thickness, $\omega_{N+} \sim$ 5–50 GHz. Picosecond wavefront generation, requiring ω_+ in excess of $\sim 2\pi \times$ 1THz, will not be possible on this structure.

The N+ layer current distribution is more complex in coplanar lines (Fig. 1.5a) than in wide microstrip lines (Fig. 1.5b) having uniform field distributions, and exact calculation of the N+ layer impedance does not appear tractable. By analogy with the microstrip case, where to second order in ω the impedance is the series inductance of a TEM wave propagating in the N+ layer shunted by the N+ layer series conductance, the N+ layer impedance of a coplanar Schottky line is approximated as the series conductance of the N+ layer underlying the line shunted by series inductance of the coplanar-waveguide TEM wave:

$$\begin{aligned} G_+ &\simeq \sigma_+ T_+ (a + 2b) \\ L_+ &\simeq \text{inductance of a coplanar waveguide of width } a \text{ and gap } b \end{aligned} \tag{1.20}$$

For coplanar Schottky contact lines, the transverse dimension of the N+ layer must be constrained. For either microstrip or coplanar waveguide Schottky lines, prohibitive losses arise from currents in the semiconductor layers unless the lateral dimensions are on the scale of a few μm. In addition, the wave impedance will be undesirably low unless the Schottky contact widths are very small.

1.3.2 Wave Impedance of the Fully Distributed Structure

The nonlinear line wave impedance $Z_{ls} \equiv \sqrt{L/C_{ls}}$ is determined by the ratio of incremental series inductance to incremental shunt capacitance. As shown above, with fixed diode cutoff frequency and fixed wavefront voltage, wavefront compression and final shock-front profile are independent of the line impedance. Input and output interfaces and power dissipation limits constrain the line impedance required of a useful device.

In standard microwave systems, the interface impedances of instruments, modules transmission lines, wafer probes, and connectors are standardized at 50Ω. Except in restricted cases where the nonlinear transmission line can be integrated with both the driving pulse generator and the output (load) device, the nonlinear line must interface to a 50Ω system. Larger or smaller wave impedance Z_{ls} will result in source and load reflections. The load reflection will interact with the forward wave, varying both its wavefront profile and its propagation delay. If Z_{ls} is substantially below 50Ω, the voltage launched onto the nonlinear transmission line will be much smaller than the open-circuit source voltage, and the wavefront compression will be reduced (Eq. 1.9). Source and load reflections with a low-impedance line can be eliminated by integration of the line with a driving generator and driven load device having matched impedances, but the power $P = V_{rms}^2/Z_{ls}$ provided by the generator and dissipated in the load varies inversely with the line impedance, while the rms wavefront voltage V_{rms} is set by compression requirements. For 50% duty-cycle, square-wave input voltage varying between zero and -5 volts, the average load power dissipation, 250 mW for a 50Ω system, increases to a substantial 2.5 W for a 5Ω system. A nonlinear transmission line having a large-signal wave impedance Z_{ls} ∼10 to 20Ω will interface poorly to other high-frequency devices and will demand substantial power from the driving generator. $Z_{ls} = 50\Omega$ is preferable, but is difficult to attain in a fully distributed nonlinear transmission line.

To illustrate the difficulties in attaining a wave impedance approaching 50Ω, consider the design of coplanar-waveguide extended Schottky contact having a geometry as in Fig. 1.5a. The center conductor width $a = 5\mu$m is chosen for ease of lithography (required line lengths for useful compression ratios are 1-10 mm), and is typical of slow-wave Schottky transmission lines reported in the literature [13,16,17]. The N+ layer doping $N_{d+} = 10^{19}/cm^3$ and thickness $T+ = 1\mu$m are chosen to minimize the N+ layer resistance R_{N+}, while the small gap size $b = 3\mu$m minimizes both R_{N+} and G_{N+}. The N- layer doping $N_{d-} = 3 \times 10^{16}/cm^3$ is typical of a microwave diode; its thickness $T+ = 0.5\mu$m is selected so that the layer is fully depleted at -5 Volts, the minimum anticipated signal voltage. The barrier potential $\phi = 0.8$V for a Ti-GaAs junction. The wave impedance is determined by the Schottky contact capacitance and the transmission line inductance. For $a/b = 5\mu m/3\mu m$ the line series inductance L is $4.4(10^{-7})$H/m. The zero-bias capacitance C_{j0} of the 5μm contact is $2.9(10^{-9})$F/m; for a 0 to -5 V step-function, the large-signal capacitance $C_{ls} = 0.54C_{j0}$ (Eq. 1.13) is $1.57(10^{-9})$F/m. The resulting large- signal wave impedance is $Z_{ls} = 16.8\Omega$, lower than is desirable.

The wave impedance can be increased by increasing the series inductance or decreasing the shunt capacitance. To increase the inductance, the transverse dimen-

sions of the interconnecting lines must be substantially increased, introducting large semiconductor-layer losses. The capacitance can be decreased by either decreased N- layer doping or decreased Schottky contact width, but decreased doping rapidly degrades the diode cutoff frequency, while decreased Schottky contact width results in lithographic difficulties and high metallic losses.

Consider first increasing the wave impedance by increasing the line series inductance. The inductance can be increased only slightly by increasing the ratio of dimensions b/a. The inductance L is given by $L = Z_1/v_1$, where Z_1 and v_1 are the characteristic impedance and phase velocity of transmission line on an insulating substrate. For a coplanar line on GaAs, the phase velocity $v_1 = 0.38c$ is independent of geometry, while [18] line impedances greater than $\sim 100\Omega$ are difficult to attain: at $b/a = 100$, $Z_1 = 152\Omega$, and series inductances greater than $1.5(10^{-6})$H/m are not feasible. Microstrip transmission line is similarly constrained [18,19]. If the conductor spacing is increased to $b = 100\mu$m, the series inductance increases to only 10^{-6}H/m and the wave impedance has increased to only $Z_{ls} = 25\Omega$. The large spacing b results in substantial N+ layer losses: from Eqs. (1.17,1.18), $\omega_+ \simeq 32$GHz, while $\omega_{diode} \simeq 275$ GHz, and increases in ω_+ attained by decreases in N+ layer doping or thickness will rapidly degrade ω_{diode}. With strong propagation loss occurring at frequencies approaching 50 GHz, the structure will not support the generation of picosecond wavefronts.

The capacitance can be decreased by decreasing either the Schottky contact width or the N- layer doping. Decreasing the doping increases the depletion layer thickness, and hence decreases the capacitance per unit area (Eq. 1.18). The decreased capacitance per unit area comes at the expense of degraded diode r_sC cutoff frequency. With decreased doping, the variation in delpetion depth with voltage is greatly increased, and the conductivity of the layer is decreased. Thus, when the N- layer is not fully depleted, the undepleted portion of the layer has a greater thickness and a greater resistivity. The diode series resistance and hence the diode cutoff frequency and the compressed falltime are degraded, as we now show in detail:

The depletion-layer capacitance will only decrease with voltage until the depletion edge reaches the N+/N- interface; thereafter, it remains constant. The N- layer thickness T_- is thus selected so that the layer is fully depleted at the peak negative voltage V_l:

$$T_- = \sqrt{2\epsilon(\phi - V_l)/qN_{d-}}\ , \tag{1.21}$$

where $\epsilon \simeq 13.1\epsilon_0$ is permittivity of GaAs, and ϵ_0 is the permittivity of vacuum. From section 1.3.1 (eq. 1.17), the undepleted fraction of the N- layer contributes

to the varactor series resistance. Combining Eqs. 1.17 and 1.21, and neglecting the N+ layer resistance, we find that the diode zero-bias rC cutoff frequency is:

$$R_{N-}C_{j0} = \frac{\epsilon}{\sigma_-}\left(\sqrt{1 - V_l/\phi} - 1\right) = \frac{\epsilon}{q\mu_n N_{d-}}\left(\sqrt{1 - V_l/\phi} - 1\right) \tag{1.22}$$

If we then neglect the variation of R_{N-} with voltage (and thus overestimate T_f), Eq. (1.15) permits calculation of the minimum compressed falltime as limited by R_{N-}:

$$T_{f,90\%-10\%} \sim 8.8\frac{\epsilon}{\sigma_-} = 8.8\frac{\epsilon}{q\mu_n N_{d-}} \quad \text{if } \omega_{diode} \ll \omega_{N+} \tag{1.23}$$

The line impedance varies as the inverse of the square root of capacitance, the capacitance varies as the square root of the N- layer doping N_{d-}, and the diode $R_{N-}C_{j0}$ time constant varies as the inverse of N_{d-}; if ω_{diode} dominates, adjusting N- layer doping will result in T_f varying as the fourth power of the line impedance. Returning to the design case above, if N_{d-} is decreased from $3(10^{16})/cm^3$ to $3(10^{15})/cm^3$, Z_{ls} will increase to 30Ω, but ω_{diode} will decrease to 200 GHz and T_f will increase to 4.2 ps. Diode active (N-) layer doping cannot be decreased below $\sim 10^{16}/cm^3$ in a structure intended for generation of 1-5 ps wavefronts.

Decreasing the Schottky contact width is a more promising method of decreasing the shunt capacitance and hence increasing the line impedance. If we return to our initial $3(10^{16})/cm^3$ doping but decrease the contact width a to 1μm, C_{ls} increases to $5.7(10^{-10})$F/m, L increases slightly to $6.8(10^{-6})$H/m, and the large-signal wave impedance becomes $Z_{ls} = 47\Omega$. The diode cutoff frequency and the N+ layer critical frequency (ω_+) remain high.

The narrow linewidth a introduces both lithographic difficulties and high metallic losses. For compression of $\sim$ 25ps wavefronts, lines of $\sim$ 5mm $\times$ 1μm will be required; definition of such structures is feasible but difficult and low-yield with optical lithography and metal liftoff techniques. The small conductor cross-section and periphery results in significant small-signal attenuation from the conductor's resistivity ρ_{metal}:

$$\alpha_{metal} \simeq \frac{\rho_{metal}/aT_{metal}}{2Z_0(V)} + \frac{\rho_{metal}/a\delta_{metal}}{2Z_0(V)} \tag{1.24}$$

where T_{metal} is the metallization thickness, ρ_{metal} is its conductivity ($\rho = 2.4\times 10^{-8}\Omega$-m for gold), and $\delta_{metal} = \sqrt{2\rho/\omega\mu}$ is the skin depth. For a 50Ω distributed Schottky contact of 1μm width and 1.5μm thickness, α_{metal} is approximately 1.5

dB/mm at 0 Hz and increases to 4.4 dB/mm at 25 GHz. In contrast, the initial design ($a = 5\mu$m) the loss increases from 0.8 dB/mm at DC to 2.4 dB/mm at 25 GHz. Both the line impedance and the conductor series resistance vary with linewidth; to attain metallic losses less than c.a. 1 dB/mm will require linewidth well in excess of 5μm, resulting in line impedances well below 15Ω. While we have not addressed the influence of these skin losses on the shock wavefront transition time, it is likely that the high loss at relatively low frequencies will prevent the formation of wavefronts with transition times below 10 ps.

Extended Schottky contact nonlinear transmission lines have low wave impedance and high losses arising from both longitudinal currents in the N+ layer and ohmic losses in the metallic transmission lines. Variations of the line's geometry or material characteristics intended to increase the wave impedance also substantially increase the line losses. While the literature has not studied their application to shock wavefront generation, extended Schottky contacts have been considered extensively for microwave phase- shifting and slow-wave (delay) application; no such device yet reported has shown either usefully high characteristic impedance or acceptably low attenuation.

These intrinsic difficulties are eliminated by adandoning the fully distributed structure in favor of a periodic structure. The continuous Schottky contact covering the full area of the transmission line center conductor is replaced by a series of small-area Schottky contacts at regular spacings along the line, reducing the average capacitance per unit line length. Further, the transmission lines between the Schottky contact can be placed on a semi-insulating substrate, eliminating the losses from longitudinal currents in N+ layers beneath the line. Lines having low loss and 50Ω large-signal impedance can be readily designed, and picosecond shock wavefronts can be generated. We will consider such periodic structures in Chapter 2.

References.

[1] R. Courant & K.O. Friedrichs, *Supersonic Flow and Shock Waves*, Wiley-Interscience, New York 1967.

[2] A. Scott, *Active and Nonlinear Wave Propagation in Electronics*, Wiley-Interscience, 1970.

[3] C.S. Tsai and B.A. Auld, J. Appl. Phys. **38**,2106 (1967)

[4] Landauer, R. :"Parametric Amplification along Nonlinear Transmission Lines",

J. Appl. Phys., 1960, Vol. 31, No. 3, pp. 479-484.

[5] R. Courant, *Methods of Mathematical Physics, Volume II: Partial Differential Equations*, Wiley-Interscience, 1962.

[6] Khokhlov, R.V. :"On the Theory of Shock Radio Waves in Non-Linear Lines", *Radiotekhnika i elektronica*, 1961, 6, No.6, pp. 917-925.

[7] M. Birk and Q.A. Kerns: "Varactor Transmission Lines", Engineering Note EE-922, Lawrence Radiation Laboratory, University of California, May 22, 1963.

[8] R.H. Freeman and A. E. Karbowiak: "An investigation of nonlinear transmission lines and shock waves", J. Phys. D: Appl. Phys. **10** 633- 643, 1977.

[9] J.M. Manley and H.E. Rowe: "Some General Properties of Nonlinear Elements-Part I. General Energy Relations," Proc. IRE, vol. 44, pp. 904-913; July 1956.

[10] C.V. Bell and G. Wade: "Iterative Traveling-Wave Parametric Amplifiers", IRE trans. Circuit Theory, vol. 7, no. 1, pp. 4-11, March 1960.

[11] S.T. Peng and R. Landauer: "Effects of Dispersion on Steady State Electromagnetic Shock Profiles", IBM Journal of Research and Development, vol. 17, no. 4, July 1973.

[12] J.L. Moll and S.A. Hamilton: "Physical Modeling of the Step Recovery Diode for Pulse and Harmonic Generation Circuits", Proc. IEEE, vol. 57, no. 7, pp. 1250-1259, July 1969.

[13] D. Jäger: "Characteristics of travelling waves along the nonlinear transmission lines for monolithic integrated circuits: a review", Int. J. Electronics, 1985, vol. 58, no. 4, pp. 649-669.

[14] D. Jäger and F.-J. Tegude :"Nonlinear Wave Propagation along Periodic-Loaded Transmission Line", Appl. Phys., 1978, 15, pp. 393-397.

[15] Lundien, K., Mattauch, R.J., Archer, J., and Malik, R. : "Hyperabrupt Junction Varactor Diodes for Millimeter-Wavelength Harmonic Generators", IEEE Trans. MTT-31, 1983, pp. 235-238.

[16] G.W. Hughes and R.M. White: "Microwave properties of nonlinear MIS and Schottky-barrier microstrip", IEEE Trans. on Electron Devices, **ED-22**, pp. 945-956, 1975.

[17] Y.C. Hietala, Y.R. Kwon, and K.S. Champlin, "Broadband Continuously Variable Microwave Phase Shifter Employing a Distributed Schottky Contact on Silicon", Elect. Lett., vol. 23, no. 13, pp. 675-676, June 1987.

[18] D.K. Ferry, Editor: *Gallium Arsenide Technology*, Howard Sams and co., 1985. See especially ch. 6.

[19] T.C. Edwards: *Foundations for Microstrip Circuit Design*, Wiley-Interscience, 1981.

Chapter 2: The Periodic Nonlinear Transmission Line

In the periodic nonlinear transmission line (Fig. 2.1a), a relatively high-impedance transmission line is loaded at regular spacings by a series of Schottky diodes serving as voltage-dependent shunt capacitances. The high- impedance interconnecting transmission lines can be approximated by L- C π-sections, and the equivalent circuit (Fig. 2.1b), an L-C ladder network in which C is voltage-dependent, resembles the incremental equivalent circuit of a fully distributed nonlinear line. For frequencies small in comparison with the ladder-network cutoff frequency $\omega_{per} = 2/\sqrt{LC}$, the small-signal propagation characteristics of the two circuits are identical. The periodic-line wave impedance, set by the ratio of L to C, can be increased by reducing the Schottky contact area (increasing C) or by increasing the diode spacing (increasing L). In contrast to fully distributed lines, a 50Ω wave impedance is feasible for periodic lines, independent of the line and diode minimum dimensions. In the periodic nonlinear transmission line, the interconnecting transmission lines are placed on a semi-insulating substrate, eliminating the losses in fully-distributed lines arising from longitudinal currents in semiconducting layers (Section 1.3.1).

Periodicity introduces an additional limit to the compressed falltime. At frequencies approaching ω_{per} the periodic line exhibits strong group velocity dispersion, while propagation is evanescent at frequencies above ω_{per}. With large signals propagating on the line, the minimum shock wavefront transition time can be limited both by line periodicity and by diode series resistance. Shock wavefront transition time, i.e. the minimum compressed falltime, in principle can be found by direct solution of the nonlinear wave equations, by approximate methods, by simulation, or by experiments with scale models. At this writing the wavefront transition time has not been found by direct solution of the nonlinear wave equations, although the analysis does provide simple and exact relationships for the large-signal wave impedance and the shock- wavefront propagation velocity.

If each section of the periodic nonlinear transmission line is approximated as a linear dispersive filter section cascaded with a nonlinear but nondispersive compression section, wave impedance, wavefront compression and minimum compressed falltime can be calculated by simple and approximate methods. Currently, the effects of skin impedance and interconnection dispersion can be calculated only by this method. Because it relates large-signal transition times to small-signal propagation characteristics, the separated disperse-compress model is appealing and

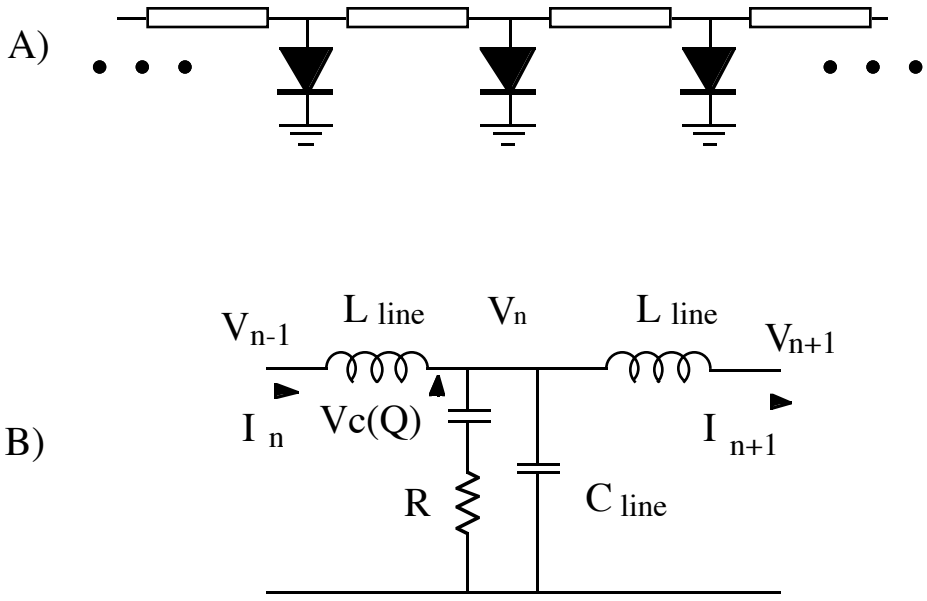


Figure 2.1: Periodic nonlinear transmission line a), and equivalent circuit of one line section b).

intuitive; however, consistency with experiment and computer simulations requires a one parameter fit.

A common circuit simulation program (SPICE) provides models for both diodes and for dispersionless transmission lines having frequency- independent loss. SPICE provides accurate waveform simulations provided that diode series resistance and line periodicity dominate skin effect and interconnect dispersion as limitations to the transition time. For the devices studied to date, skin effect and interconnect dispersion have been negligible, and SPICE has been the primary tool for circuit modeling.

Finally, because of some mistrust of the numerical methods used in the computer simulations (numerical convergence and memory overflow problems were encountered for one version of the simulation package), analysis was verified by measurement of wavefront compression on scale models of the designed monolithic device.

2.1 Partial Solution of the Nonlinear Equations

The periodic nonlinear transmission line (Fig. 2.1a) consists of a transmission line of characteristic impedance Z_1 which is loaded at electrical spacings of τ (in units of time) by a series of Schottky diodes in reverse bias. If the diode capacitance

is large in comparison with the shunt capacitance of the line $C_{line} = \tau/Z_1$, then (as will be shown later) the minimum compressed falltimes will be large in comparison with τ; to sufficient accuracy for hand calculations the transmission line segments can be replaced by π-section L-C equivalent circuits, where $L = L_{line} = \tau Z_1$ and $C_{line} = \tau/Z_1$. Modeling the diode in reverse bias as a nonlinear capacitor with a series resistance, we obtain the equivalent circuit of Fig. 1.2b. To simplify the nonlinear analysis, we will further assume that C_{line} is small in comparison with $C(V)$ and can be neglected. Writing nodal equations from the equivalent circuit:

$$\begin{aligned} V_n(t) - V_{n-1}(t) = L\frac{dI_n(t)}{dt} \ , \quad I_n(t) - I_{n+1}(t) = \frac{dQ_n(t)}{dt} \ , \\ V_n(t) = V_c(Q_n(t)) + R\frac{dQ_n(t)}{dt} \ . \end{aligned} \tag{2.1}$$

As before, the diode Q-V relationship is given by $\phi - V_c(Q) = kQ^2$, where Q is negative. Solution of the nonlinear wave equations for the periodic line can be pursued by the same method as for the continuous structure (Section 1.2); the models differ in that the derivatives with respect to distance of the continuous line are replaced by finite differences in the voltages and currents of adjacent line sections. To calculate the properties of the final shock wavefront, a wave of constant profile: $V_{n+1}(t) = V_n(t - T)$ is assumed. The wave equations become

$$\begin{aligned} V_n(t) - V_n(t + T) = L\frac{dI_n(t)}{dt} \ , \quad I_n(t) - I_n(t - T) = \frac{dQ_n(t)}{dt} \ , \\ V_n(t) = V_c(Q_n(t)) + R\frac{dQ_n(t)}{dt} \ . \end{aligned} \tag{2.2}$$

Combining the first two equations to eliminate I,

$$V_n(t + T) + V_n(t - T) - 2V_n(t) = -L\frac{d^2Q_n(t)}{dt^2} \tag{2.3}$$

After integrating Eq. (2.3) twice with respect to t,

$$V(t) * T \ \mathrm{tri}(t/T) = \int_{-\infty}^{t}\int_{-\infty}^{\beta} -L\frac{d^2Q_n(\alpha)}{d\alpha^2}d\alpha d\beta = L\left(Q_n(t) + \sqrt{\phi/k}\right) \ , \tag{2.4}$$

where $*$ denotes the convolution operation and

$$\mathrm{tri}(t) = \begin{cases} 1 - |t|, & |t| < 1 \\ 0, & |t| \geq 1. \end{cases}$$

Evaluation of Eq. (2.4) at $t = +\infty$, gives $T^2 \ \Delta V = L \ \Delta Q$. Hence the per-section delay T for the shock wavefront is

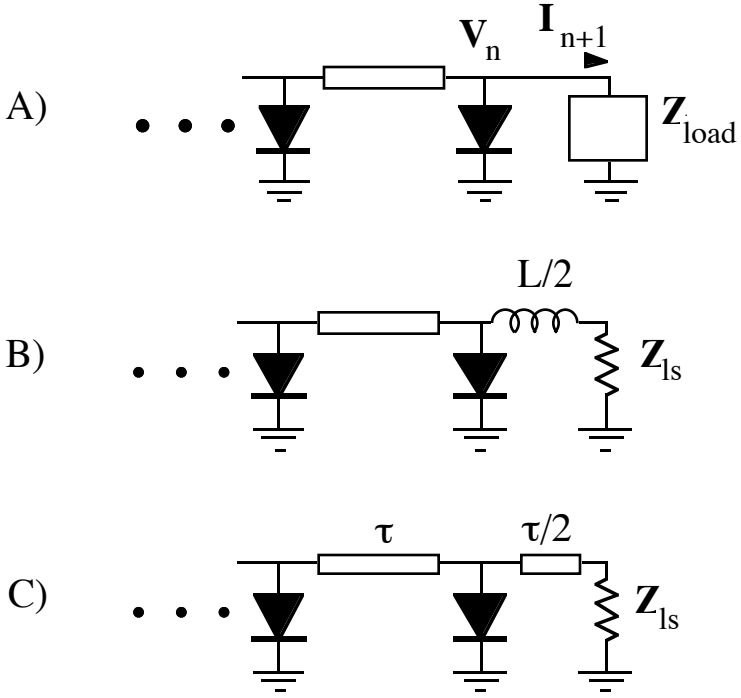


Figure 2.2: Termination of the periodic line with a load impedance conforming to Eq. (2.6), a). Synthesis of this load impedance using an inductance, b), or a half-section of line, c).

$$T = \sqrt{L\Big(\frac{\Delta Q}{\Delta V}\Big)} = \sqrt{LC_{ls}}, \tag{2.5}$$

where C_{ls} is given by Eq. (1.13). As with the fully distributed line, the shock wavefront propagation velocity is set by the large-signal capacitance. With some manipulation, Eq. (2.4) permits calculation of the current-voltage relationship of the shock wavefront. If the unit-amplitude rectangular pulse is defined by

$$\text{rect}(t) = \begin{cases} 1, & |t| < 1/2 \\ 0, & \text{otherwise,} \end{cases}$$

and if the first two parts of Eq. (2.2) are substituted into Eq. (2.4), the current of the shock wavefront is

$$I_{n+1}(t) = I_0 + \left(\frac{V_n(t)}{Z_{ls}} \; * \; \frac{1}{T} \; \text{rect}(t/T + 1/2) \right). \tag{2.6}$$

As previously, $Z_{ls} \equiv \sqrt{L/C_{ls}}$. The nonlinear transmission line is a nonlinear system, hence the principle of superposition cannot be applied, and eigenfunction expansions such as Fourier integrals cannot be used to calculate system responses.

Despite this, Fourier transformation of Eq. (2.6) is helpful; if the line output is loaded by an impedance conforming to the above current-voltage relationship (Fig. 2.2a), the shock wavefront will be transmitted to the load without deformation. The load impedance is a linear system and Fourier methods can be applied. Taking the Fourier transform of $\mathrm{rect}(t/T + 1/2)$, the correct load impedance is:

$$Z_{load} = Z_{ls} e^{j\omega T/2} \frac{\omega T/2}{\sin(\omega T/2)} \simeq Z_{ls} + j\omega L/2, \tag{2.7}$$

where the approximation results from a first-order expansion in $\omega T/2$. From simulations, shock wavefront transition times are in the range of $2T$–$4T$. Termination of the line with Z_{ls} in series with a matching inductance $L/2$ (Fig. 2.2b) or a half-length ($\tau/2$) matching section of line (Fig. 2.2c) is thus sufficient for line-load coupling with negligible wavefront broadening.

Following the strategy of Section 1.2, an attempt to calculate the minimum compressed falltime might be made by combining Eq. (2.4) with the third part of Eq. (2.2), and eliminating either Q or V. In this case, unfortunately, the resulting equation is not separable:

$$(\phi - kQ_n^2 + R\frac{dQ_n}{dt}) \ * \ T \ \mathrm{tri}(t/T) = L(Q_n + \sqrt{\phi/k}) \ .$$

Although complete analysis of the shock wavefront profile by solution of the nonlinear wave equations has not been successful, this approach does provide two worthwhile results. As with the fully distributed line, the shock-wavefront velocity is determined by the large-signal capacitance. The conclusion regarding load impedance is more noteworthy: despite nonlinear shunt capacitance, diode series resistance, and line periodicity, the line output is correctly terminated by the large-signal wave impedance in series with a half-length section of transmission line. To procede further it is necessary to use simulations providing accurate modeling and approximate methods providing understanding.

2.2 Approximate Analysis

Compression of wavefronts on the idealized nonlinear transmission line of Section 1.1 arises from a variation of propagation velocity with voltage. Introduction of diode series resistance and line periodicity into the model results in a nonzero shock wavefront transition time. Calculation of this transition time by direct solution of the nonlinear wave equations has proved fruitless, but alternative methods exist.

2.2.1 Energy-Momentum or Energy-Charge Methods

In a variety of physical systems exhibiting shock-wave formation, the shock wavefront transition time can be calculated through the methods of conservation of energy and momentum (mechanical systems), or conservation of energy and charge (electrical systems). As an illustration, consider the idealized continuous nonlinear transmission line of Chapter 1. Energy is lost as a shock wavefront propagates along the line. The loss [1] is the difference between the final stored energy in the nonlinear capacitors and the energy necessary to charge them during passage of the shock wavefront, and is equal to the difference in the stored energies of a linear capacitance C_{ls} and a nonlinear capacitance $C(V)$ having the same ΔQ and ΔV. The loss is a consequence of conservation of charge and is independent of the diode series resistance; yet, the lost energy must be dissipation in this resistance. Power dissipation in the resistance is I^2R and is proportional to $1/T_{f,min}^2$. The dissipated energy then varies as $1/T_{f,min}$, and hence $T_{f,min}$ can be calculated. The approximation is useful when the shock wavefront is limited primarily by small losses, including the case of resistive losses of the N+ layer in the lines of Section 1.3.1. For systems with small loss and strong dispersion, the shock wavefront will exhibit overshoot or oscillation [2]. The wavefront is no longer a simple monotonic edge which, approximated by a linear ramp, has a single degree of freedom. Instead, the simplest approximation of the wavefront is a damped sinusoid whose damping and natural frequency must be calculated. Because of the increased degrees of freedom in the form of the shock wavefront, calculation of the transition time requires additional information.

2.2.2 The Expand-Compress Model.

The profile of the shock wavefront can be viewed as a consequence of competition between wavefront compression arising from the voltage- dependent delay and wavefront expansion arising from various sources of loss and group velocity dispersion in the circuit. For the periodic nonlinear transmission line, loss is introduced by diode series resistance while dispersion results from both the line periodicity and from the group velocity dispersion of the interconnecting transmission lines. The skin impedance of the transmission lines introduces both loss and dispersion.

As a wavefront propagates along the line, its falltime will at first decrease linearly with distance. As the transition time approaches the characteristic time constants governing the line loss and dispersion, the wavefront expansion arising from dispersion and loss will increase. A final limited transition time will be reached at which at each section of the periodic line, the wavefront expansion due to loss and dispersion is in equilibrium with the wavefront compression due to the voltage-

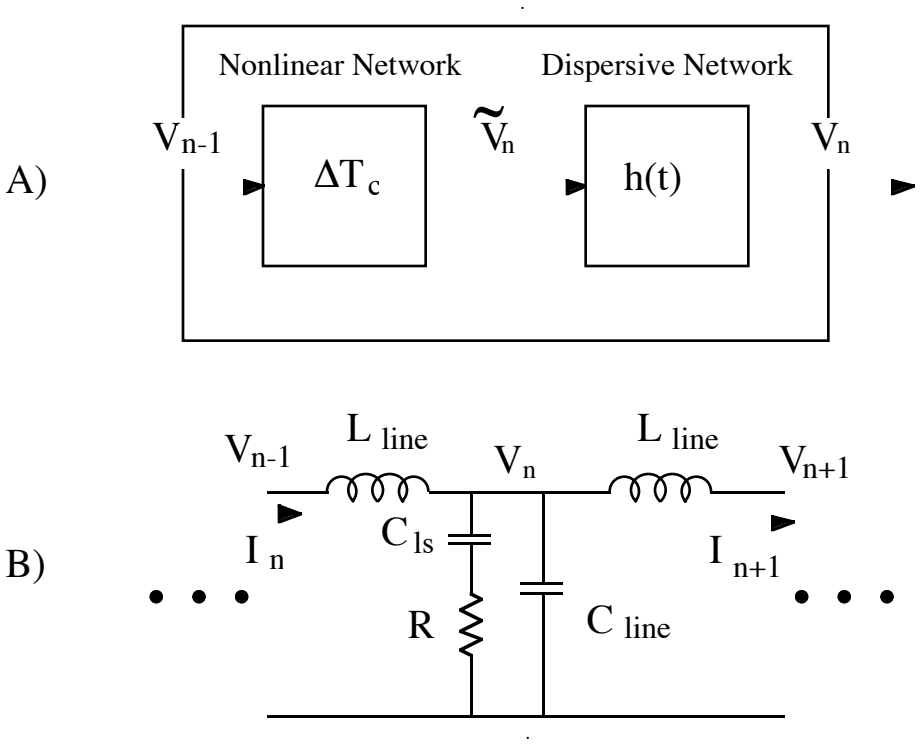


Figure 2.3: Section of the periodic line approximate by cascaded nonlinear nondispersive and linear dispersive subsections, a). Linearized equivalent circuit for calculation of the linear subsection transfer function, b).

dependent delay.

To model this competition between expansion and compression, we treat each section of the periodic line as a cascade of a nonlinear compression system with a linear expansive system (Fig. 2.3). The nonlinear compression section has the characteristics of an ideal nonlinear transmission line; the falltime $\tilde{T}_{f,n}$ of $\tilde{V}_n$ is given by

$$\begin{aligned} \tilde{T}_{f,n} &= T_{f,n-1} - \left(\sqrt{LC_T(V_h)} - \sqrt{LC_T(V_l)} \right) \\ \tilde{T}_{f,n} &= T_{f,n-1} - \Delta T_c \end{aligned} \tag{2.8}$$

The total line shunt capacitance $C_T(V)$ is the sum of the diode capacitance $C(V)$ and the interconnecting line capacitance $C_{line} = \tau/Z_1$.

The linear network has the impulse response of a section of the periodic line, where the network has been linearized by replacement of $C(V)$ by its average value C_{ls}. Its impulse response $h(t)$ is the Fourier transform of $H(\omega)$, the transfer function $V_n(\omega)/V_{n-1}(\omega)$ of the network in Fig. 2.3b. From $h(t)$ the step-response risetime T_{linear} of the linear network can be calculated. The falltime at the output of

the nth section of line can then be calculated using approximate sum-of-squares convolutions:

$$T_{f,n}^2 = \sqrt{\left(T_{f,n-1} - \Delta T_c\right)^2 + T_{linear}^2} \tag{2.9}$$

The evolution of falltime along the line can be calculated from this nonlinear difference equation. Near the input of the line, where the falltime T_f is much greater than the characteristic expansion time T_{linear}, the falltime decreases linearly with section number:

$$T_{f,n} \simeq T_{f,n-1} - \Delta T_c \quad \text{if} \quad T_{f,n} \gg T_{linear} \tag{2.10}$$

The minimum compressed falltime is calculated by setting $T_{f,n} = T_{f,n-1}$:

$$T_{f,min} = \frac{T_{linear}^2}{2\ \Delta T_c} \tag{2.11}$$

Calculation of the minimum compressed falltime now requires calculation of the risetime of the linear network. Through this method the influence of diode series resistance can be calculated with reasonable accuracy (eq. 1.15 provides the result without approximation), and the influences of line periodicity, skin impedance, and dispersion can be approximated.

2.2.3 Periodicity and Series Resistance.

From the equivalent circuit (Fig. 2.3b) of a single line section, the transfer function of the linear network can be calculated:

$$H(\omega) = e^{A+j\Theta} = e^{A_1(\omega)} e^{A_2(\omega)} e^{j\Theta_2(\omega)} = H_1(\omega)\ H_2(\omega)\ ,$$

where

$$\cosh(A + j\Theta) = 1 - \frac{\omega^2 L}{2}\left(C_{line} + \frac{C_{ls}}{1 + j\omega R C_{ls}}\right)\ . \tag{2.12}$$

Defining $Z_{ls} = \sqrt{L/(C_{line} + C_{ls})}$ and approximating to second order in ω:

$$\begin{aligned} H_1(\omega) &= \exp\left(-\frac{1}{2}\omega^2 C_{ls}^2 Z_{ls} R\right) \\ h_1(t) &= \frac{1}{\sqrt{2\pi}\ \tau_{RC}} e^{-t^2/2\tau_{RC}^2}\ , \end{aligned} \tag{2.13}$$

where $\tau_{RC} = C_{ls}\sqrt{Z_{ls}R}$. Within the accuracy of the approximations above, the impulse response $h_1(t)$ represents the pulse expansion due to diode series resistance. Because the impulse response is Gaussian, the step response is an integrated Gaussian, and the 10%–90% risetime T_{diode} of $h_1(t)$ is

$$T_{diode} \simeq 2.6C_{ls}\sqrt{Z_{ls}R} \; . \tag{2.14}$$

Substituting T_{diode} for T_{linear} in Eq. (2.11) yields an expression for $T_{f,min}$ proportional to RC_{j0} times a function of V/ϕ which differs in form from the exact solution of Eq. (1.15). The discrepancy in $T_{f,min}$ is generally less than 5%.

The treatment of line periodicity by the expand-compress model is less satisfying. Returning to eq. (2.12), and neglecting for this calculation the diode series resistance gives

$$\cos(\Theta_2 + jA_2) = 1 - \frac{\omega^2 L}{2}\left(C_{line} + C_{ls}\right) = 1 - 2\frac{\omega^2}{\omega_{per}^2} \; . \tag{2.15}$$

To include a nonzero resistance, omission of terms in R in Eq. (2.15) is correct only to third order in ω, and adequate analysis of Eq. (2.12) requires expansion to at least fourth order in Θ. Equation (2.15) is the dispersion relationship for a periodic L–C ladder network without resistance in series with the capacitors. As ω approaches the periodic line cutoff frequency ω_{per}, the phase velocity decreases. For $\omega > \omega_{per}$ propagation is evanescent with the loss $|H_2(\omega)|$ increasing exponentially with frequency. If the transfer function $H_2(\omega)$ is approximated as an ideal low pass filter,

$$H_2(\omega) \sim \begin{cases} 1, & \text{if } |\omega| < \omega_{per}; \\ 0, & \text{otherwise,} \end{cases} \tag{2.16}$$

then the 10%–90% risetime of the filter section is

$$T_{per} = 1.40\sqrt{L(C_{line} + C_{ls})} = 1.40Z_{ls}(C_{line} + C_{ls}) \; . \tag{2.17}$$

Unfortunately, when substituted into the relationship for minimum compressed falltime Eq. (2.11), the linear-filter risetime of Eq. (2.17) does not give results consistent with experiment or with computer simulation; omission of the dispersion of $H_2(\omega)$ apparently *overestimates* T_{per}.

Equation (2.16) approximates $H_2(\omega)$ as an ideal low-pass filter for all frequencies below the cutoff frequency; if the transfer function is instead approximated as an ideal low-pass filter having linear phase shift $\Theta = -\omega\sqrt{L(C_{line} + C_{ls})}$ for all Θ less than the cutoff phase ($\Theta = \pi$), then the 10%-90% risetime of the filter section becomes:

$$T_{per} = 0.7\sqrt{L(C_{line} + C_{ls})} = 0.7 Z_{ls}(C_{line} + C_{ls}) \tag{2.18}$$

The choice between approximations (2.18) and (2.17) is arbitrary, yet, using approximation (2.18) and setting

$$T_{linear} = \sqrt{T_{diode}^2 + T_{per}^2}\ , \tag{2.19}$$

gives expand-compress model results which conform well to the results of computer simulations. Figure 2.4 compares the results of falltime calculations using SPICE and the expand-compress model incorporating approximation (2.19). In the figure, the interconnecting line impedance is 80Ω, $V_l = -2$ volts, and $\phi = 0.8V$. For the upper graph, $C_{j0} = 50$fF, $\tau = 1.6$ps, and $R = 3.8\Omega$. In the lower graph the line cutoff frequency has been scaled 2:1 with constant ω_{diode} and constant Z_{ls}: $C_{j0} = 25$fF, $\tau = 0.8$ps, and $R = 7.6\Omega$. These results are due to Rob Marsland. The falltime initially decreases linearly as it propagates on the line, conforming to Eqs. (1.8) and (2.10), and asymptotically approaches a final, limited falltime.

The equations of this section are readily manipulated, and it is tempting to rearrange them to obtain an complex expression for $T_{f,min}$ in terms of R, L, C_{j0}, V_l, and ϕ. In doing so, we leave the scope of the approximate model. Weak approximations (Eq. 2.17 *vs.* Eq. 2.18) limit the accuracy, and the correlation of the data in Fig. 2.4 is best viewed as the result of a one-parameter fit. The expand-compress model provides simple and easily interpreted first-order estimates of the variation of $T_{f,min}$ with the circuit parameters, variations which can be hard to extrapolate from the results of a series of computer simulations. One further result is worth quoting:

$$\begin{aligned} T_{f,min} &\simeq f(V_l/\phi)\left(\frac{1}{\omega_{diode,ls}} + \left(0.14\ \frac{C_{line} + C_{ls}}{C_{ls}}\right)\frac{1}{\omega_{per}}\right) \\ f(V_l/\phi) &\simeq \begin{cases} 15, & \text{if } V_l = -2\text{V and } \phi = 0.8\text{V}; \\ 10, & \text{if } V_l = -5\text{V and } \phi = 0.8\text{V}; \end{cases} \end{aligned} \tag{2.20}$$

where an average diode cutoff frequency $\omega_{diode,ls} = 1/RC_{ls}$ has been defined, and where $f(V_l/\phi)$ decreases for more negative V_l . The diode cutoff frequency must be approximately 5 times the periodic-line cutoff frequency if the two effects are to have equal influence on the minimum compressed falltime.

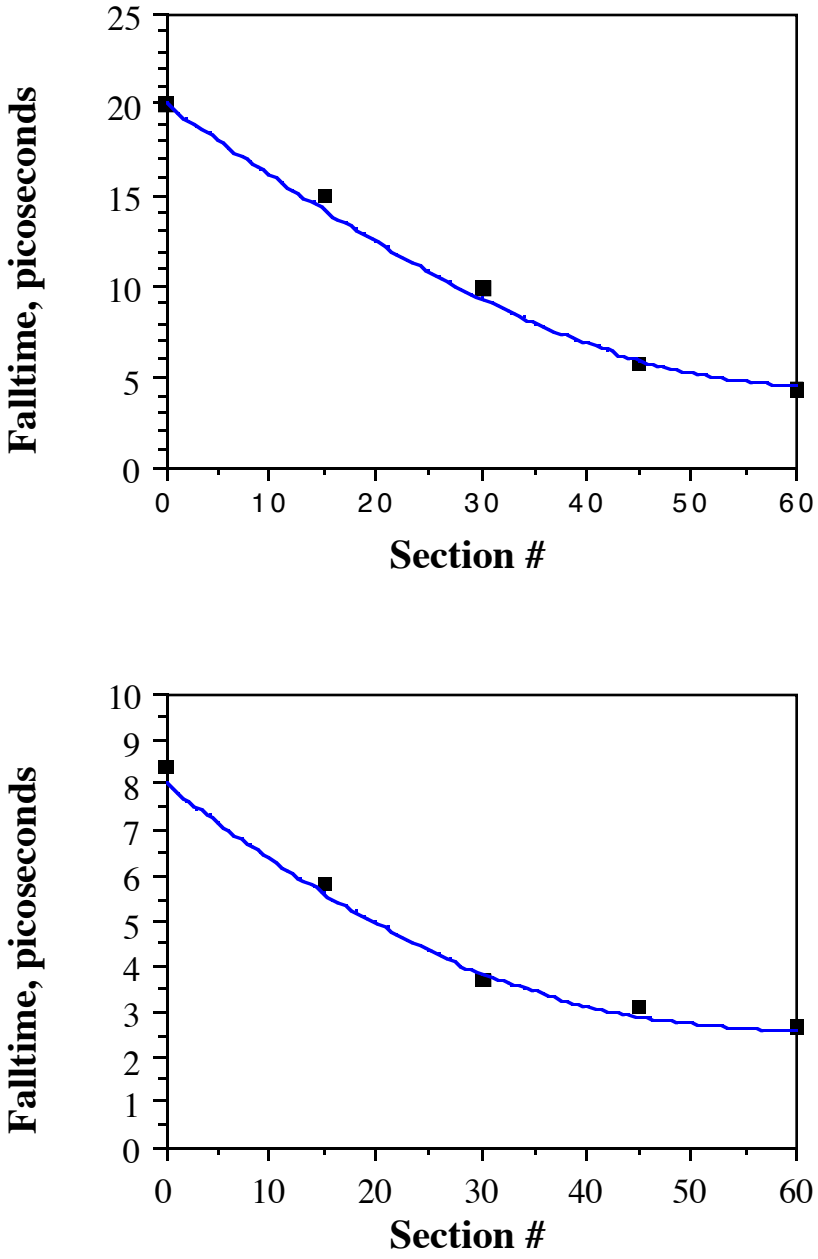


Figure 2.4: Falltimes on representative nonlinear transmission lines as calculated by the expand-compress approximation (solid lines) and nonlinear circuit simulations (points). See text for circuit parameters. In the lower graph the line cutoff frequency has been scaled 2:1 with constant ω_{diode} and constant Z_{ls}.

2.2.4 Effect of Skin Impedance and Interconnect Dispersion

The profiles of pulses propagating on coplanar waveguide or microstrip transmission lines are distorted by both the skin impedance arising from nonzero conductor resistivity, and the group velocity dispersion arising from the pseudo-TEM mode of propagation. Impulses of a few picoseconds duration can be significantly broadened after propagation over distances on the order of 1 cm on these lines [3,4]. In the periodic nonlinear transmission line, interconnect dispersion and skin effect will contribute to wavefront broadening, and can potentially increase the minimum compressed falltime.

The mathematics of skin impedance and dispersion are complex, and incorporation of these effects into the nonlinear model leads to equations which cannot be solved by analysis. Computer modeling is also inconvenient. Some microwave linear simulation packages have refined models for transmission-line loss and dispersion, but nonlinear analysis is required here. Nonlinear circuit analysis programs provide models only for ideal transmission lines and lines having frequency-independent loss. Therefore, the expand-compress approximation is used to estimate these effects. First, consider the skin effect. Following the analysis of Wigington and Nahman [5], the impulse response $h_{skin}(t)$ due to the skin impedance is

$$h_{skin}(t) \simeq \frac{1}{\tau_{skin}} \sqrt{\frac{2}{\pi}} \left(\frac{\tau_{skin}}{t}\right)^{3/2} e^{-\tau_{skin}/2t}, \tag{2.21}$$

where

$$\tau_{skin} = \frac{l^2 \mu_0 \rho_{metal}}{16 Z_{ls}^2 P^2}, \tag{2.22}$$

and l is the diode separation in units of distance, ρ_{metal} is the interconnect resistivity, and μ_0 is the magnetic permeability of vacuum. For a coplanar waveguide having a signal-ground spacing b much greater than the signal-conductor width a the effective current-carrying periphery P is approximately equal to a, while the duration of the impulse response $h_{skin}(t)$ is approximately $2.5\tau_{skin}$. An estimate of the component of the linear-system risetime T_{linear} due to the skin effect for use in the stretch- compress model is then

$$T_{skin} \simeq 0.16 \left(\frac{l^2 \mu_0 \rho_{metal}}{Z_{ls}^2 a^2} \right). \tag{2.23}$$

For the designs considered in Section 2.2.3, the effect of skin impedance is negligible in comparison with the effects of line periodicity and diode series resistance.

Interconnect dispersion will be treated very crudely. For a coplanar waveguide the phase velocity varies as the inverse square root of the line effective dielectric

constant [7]. At low frequencies the effective dielectric constant is the average of that of air and the GaAs substrate. As frequency increases, the effective dielectric constant increases, asymptotically approaching that of the substrate. The pulse dispersion must be less than the difference in delay of low-frequency ($\omega \to 0$) and very high-frequency ($\omega \to \infty$) components; for a line of electrical length τ on GaAs($\epsilon_r \simeq 13.1$), the difference in delay these component is $\Delta\tau = 0.36\tau$. This differential delay is less than T_{per} (eq. 2.18) by the ratio $1 : 0.5C_{line}/(C_{line} + C_{ls})$, which is ≈ 0.2 in typical design cases. Even with this gross upper bound, the result permits neglecting the dispersion.

2.2.5 Design Relationships

Analysis of the nonlinear transmission lines is best done by simulation, but simulation is not suitable for synthesis. For synthesis, some simple design relationships are useful. The design procedure is as follows:

In designing the transmission line, the signal voltage V_l, the input falltime $T_{f,in}$ and the line large-signal impedance Z_{ls} will be specified, the varactor diode cutoff frequency will generally be independent of the device area, and lithographic considerations will limit the minimum varactor area, setting a minimum C_{ls}. The interconnecting line impedances and lengths must then be chosen, and the necessary number of line sections determined. As ω_{diode} is fixed, any free parameters are varied to maximize ω_{per} and hence minimize the compressed falltime.

The periodic line cutoff frequency can be written as

$$\omega_{per} = \frac{2}{Z_{ls}C_{ls}}\left(1 - \frac{Z_{ls}^2}{Z_1^2}\right) , \tag{2.24}$$

hence the diode capacitance should be made as small as lithography permits, and the interconnect impedance should be made as high as possible. Constraints on interconnect impedance will be discussed in Chapter 3. To obtain the desired large-signal wave impedance, the interconnecting line lengths (in units of time) are then chosen as

$$\tau = \frac{C_{ls}Z_{ls}^2}{Z_1}\left(\frac{1}{1 - Z_{ls}^2/Z_1^2}\right) , \tag{2.25}$$

Again, high-impedance interconnections are desirable to minimize line lengths. From Eq. (2.25) the diode spacing in units of distance l are calculated by multiplication by the phase velocity of the coplanar waveguide interconnections:

$$l = \tau \times 1.15(10^8) \text{ m/s.} \tag{2.26}$$

To compute the required number of line sections, we first calculate the per-section compression

$$\Delta T_c = \sqrt{LC_T(V_h)} - \sqrt{LC_T(V_l)} \ . \tag{2.27}$$

For significant compression, V_l must be large in comparison with ϕ and the C_{ls} must dominate C_{line}, hence large interconnect impedance is again desirable. Compression to $T_{f,min}$ is approached asymptotically (Fig. 2.4); a number of sections n equal to $(T_{f,in} - T_{f,min})/\Delta T_c$ is not sufficient for the falltime to approach $T_{f,min}$. To a better approximation, the necessary number of sections is estimated as

$$n \simeq T_{f,in}/\Delta T_c \ . \tag{2.28}$$

The initial approximate circuit design can now be verified by computer simulation using SPICE.

2.3 SPICE Simulations

Simulations, by their nature, provide little material for discussion. The chosen circuit parameters are entered into the input file of the simulation program, the simulation is run, and the falltimes and waveforms at the output (and at intermediate points) of the nonlinear transmission line are predicted. The general trends of the relationship between circuit parameters and compressed falltime are consistent with those predicted by the approximate (expand-compress) model. The required line lengths for compression to near the minimum compressed falltime are consistent with Eq. 2.28. Simulations thus are used primarily for accurate computation of the minimum compressed falltime, and for computation of the shock profile: is the edge well-damped, or is it underdamped, exhibiting overshoot and ringing? Even with SPICE simulations, some errors can occur in modeling.

During the development of the nonlinear transmission line, two versions of SPICE were available, both operating on a minicomputer. Neither version provides skin effect or dispersion in its transmission line models, but these effects were considered to be negligible (Section 2.2.4). More serious were numerical problems in time-domain simulations of circuits with multiple misterminated transmission-line sections. The more modern version of SPICE, SPICE 3, was on β-test; SPICE 3 would not permit 1 picosecond time increments in its calculations and thus could not be used for simulation of a circuit whose electrical line lengths were on the order of 0.5-5 picoseconds. SPICE version 2G.4 would permit smaller time increments,

but difficulties with numerical convergence in the models for the transmission lines resulted in large run times in modeling transmission lines of significant length. A 15-section line would require approximately 10 minutes simulation time, while several submissions of 30-section circuits never resulted in a computer output.

Compression from 25 ps to ∼5 ps requires lines with approximately 45-60 sections. Initially, such lines were simulated in pieces. A 15–diode line would be simulated with an input 90%–10% falltime of 25 picoseconds. The precise shape of the output waveform would be neglected, and the output 90%–10% falltime would be used as the falltime of a linear-ramp input for simulation of a second 15–diode line. In this manner, lines having 60 sections would be simulated. From these piecewise simulations, the reduction of the falltime in the compression region was found to follow Eq. (2.10). Subsequent simulations thus concentrated on minimum compressed falltime and on wavefront profile. A 15-diode line would be simulated in which the input falltime was an estimate of the minimum compressed falltime. Unless the simulation showed substantial differences between the waveforms and falltimes at the 10th and 15th diodes, the waveform at the 15th diode would be considered to approximate the final shock profile.

Subsequent to the work of this thesis, a version of SPICE for operation on a personal computer (Microwave SPICE from EESOF, inc.) has become available, and full 60–diode lines have been simulated. The expanded capability of the program for the smaller computer is probably due to numerically efficient transmission line models. Except in the case of very small diode resistance ($R < 1\Omega$), these simulations agree well with the 15–diode simulations. With zero diode resistance, simulations on the full line reveal that an asymptotic wavefront does not exist, and the wavefront propagates on the line with wavefront damping decreasing progressively with distance. This is consistent with the predictions of the energy-charge methods of 2.2.1. If the asymptotic wavefront is approximated as a second-order system response, then the wavefront damping must decrease towards zero as the diode resistance approaches zero. Nonzero diode resistance results in nonzero damping, and an asymptotic wavefront can then exist.

What remains to be said concerning simulations? Unlike the results of approximate methods, simulated solutions can predict shock-wave profiles. Figure 2.5 shows the results of self-consistent 15–diode simulations of nonlinear transmission lines. Note the increase in both damping and falltime as the diode resistance is increased. These simulations use the initial design values for the first monolithic device.

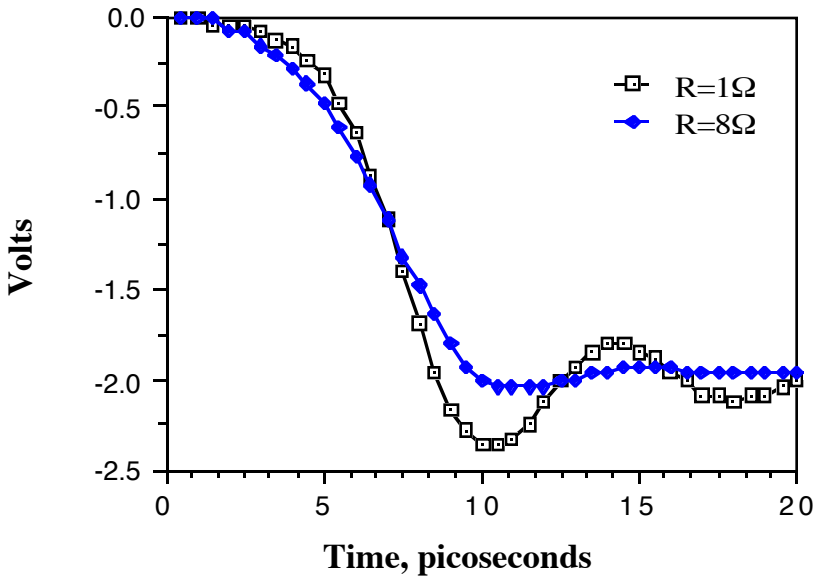


Figure 2.5: SPICE simulations for calculation of minimum compressed falltime. For both simulations the interconnecting line impedance is 90Ω, $\tau = 1.5$ps, $V_l = -2$ volts, $\phi = 0.8V$, and $C_{j0} = 50$fF. The falltime is 4.0 ps at $R = 1\Omega$ ($\omega_{diode} \simeq 2\pi \times 3$ THz) and increases to 5.7 ps at $R = 8\Omega$ ($\omega_{diode} = 2\pi \times 400$ GHz). The periodic-line cutoff frequency ω_{per} is approximately $2\pi \times 140$ GHz.

2.4 Scale-Model Experiments [6]

Wavefront compression on the nonlinear transmission line appears to be consistently modeled by simulations and by the approximate methods of the preceding sections, but are the results correct?

The approximate method contains several weak assumptions, and separation of each line section into decoupled compressive and dispersive section has not been justified by any analytical method. SPICE simulations are more convincing, but even these can be in error. Are the transmission line models reasonable, and are the lines of $\tau = 1.5$ps adequately modeled by a finite-difference algorithm using time increments only a factor of 3–5 smaller than the line electrical length (very large computation times result if smaller increments are used)? The sporadic convergence difficulties are also worrying: given zero initial conditions, and no perturbation to initiate a characteristic response, SPICE simulations of unstable systems often fail to show oscillation (several AT&T high-speed communications IC's are know to have fallen into this design trap). Often the circuit instability manifests itself as

occasional poor numerical convergence during simulation. Are the convergence problems with simulations of the nonlinear transmission line due to difficulties with line models, or do they reflect possible parametric instabilities in the circuit?

To demonstrate the feasibility of falltime compression before investing six months in process development, the design of the monolithic device was verified through experiments with easily-fabricated scale models. Preliminary designs for the monolithic device were first developed using the procedures of Section 2.2.5. The design was then simulated using SPICE; while the process had not been designed, it was expected that diode cutoff frequencies in the range of 400 GHz to 3 THz would be feasible on the first design iteration. The preliminary design is as given in the previous section; 50fF diodes (C_{j0}) were assumed as the minimum feasible size for the first process iteration, 90Ω interconnecting lines were selected as the highest feasible characteristic impedance (see Chapter 3), the line large-signal impedance must be 50Ω, and the input falltime was assumed to be 25 picoseconds. The step magnitude was selected somewhat arbitrarily as 2 volts, and Schottky contact potentials ϕ on GaAs are invariably 0.7-0.9 volts. The interconnect lengths are then calculated as $\tau = 1.5$ps, and approximately 55 sections are required for compression to the minimum compressed falltimes (Fig. 2.5).

The scale model was based upon this design. If the input and output falltimes are to be scaled with constant waveshapes, with terminal impedances held constant, then diode capacitance and interconnect electrical lengths τ must be increased in proportion, and interconnect impedance and diode series resistance must be held constant. To produce the shortest possible wavefronts from the scale model (and hence produce the most impressive results) the scaling parameter was made as small as possible. Commercial silicon abrupt-junction diode die (Eaton AL-35136) were selected; with 1pf zero-bias capacitance, the smallest available, the scaling parameter is 20:1. Interconnect electrical length must scale to 30 picoseconds. For these diodes ϕ was specified as 0.5–0.7 volts, but measured 0.8 V. From quality-factor specification, the series resistance was estimated: $R \simeq 1\Omega$.

The scale model was constructed as a series of 3 hybrid assemblies (Fig. 2.6), each consisting of a 90Ω microstrip transmission line on a 0.6 mm alumina substrate ($\epsilon_r = 10$) loaded to ground by 15 diodes spaced at intervals of 3mm ($\tau = 26$ ps). The input and output of each line consisted of a short section of 50Ω microstrip and an SMA-family microstrip launcher, thus providing low-reflection 50Ω interfaces. The diodes were die-attached to a wrap-around ground bus adjacent to the microstripline and the diode anodes were bonded to the line by short (0.2 mm) sections of 500 line/inch gold mesh.

Several parameters of the scaled device deviated from 20:1 scaling. Dispersion

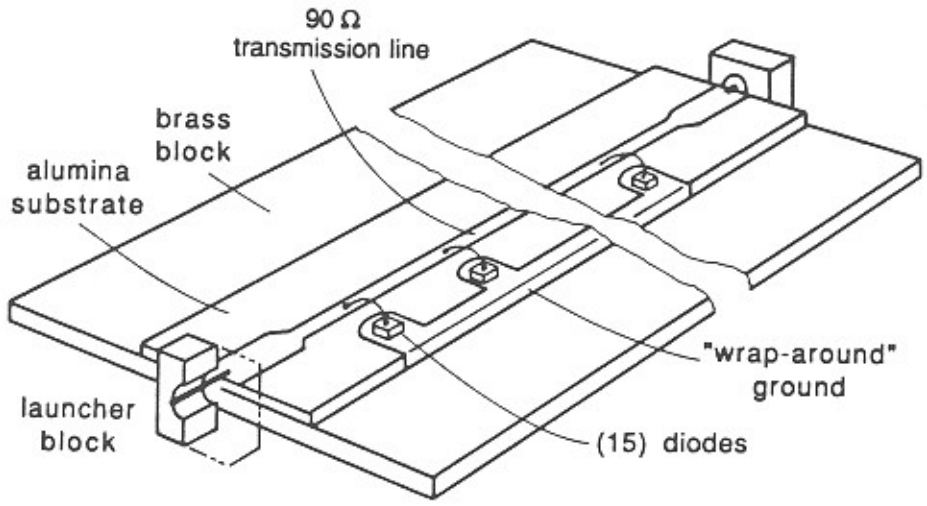


Figure 2.6: Single 15–diode section of 20:1 scale model nonlinear transmission line.

and skin effect were not well-scaled: scale-modeling of the skin impedance would have required transmission-line conductors having scaled resistivity, while the dispersion of a 90Ω microstrip on a 600 μm alumina ($\epsilon_r \approx 12.9$) substrate is not a scaling of the dispersion of a 90Ω coplanar waveguide having a 120μm ground-ground separation on a GaAs ($\epsilon_r \approx 13.1$) substrate. These two neglected effects are of lesser importance. The interconnecting line lengths were approximately 10% short, due to some quantization difficulties in the graphic layout program used to cut the rubylith masks. Measurements of the transmission lines prior to and after diode attachment indicated that the diode zero-bias capacitance, averaging 0.86 pf, was 14% low. The diodes did, however, conform well to step-junction C–V characteristics over the range of zero to -15 volts. The circuit thus corresponded better to a 18:1 scale model. Finally, the substrate size limited each hybrid to 15 diodes; the total line consisted of 45 sections. For an 18:1 scale model having 45 diodes, an input falltime of less than 360 ps is required to obtain compression to the minimum compressed falltime of 72 ps.

The small-signal parameters of the 45-diode line were measured by a time-domain-reflectometer and a microwave vector network analyzer. The measured small-signal characteristic impedance varied from 45 to 58 Ω as the line bias was varied from 0 to -3 volts. The group delay (Fig. 2.7) varies by 500 ps over the same range of voltages. Note also in this figure the increase in group delay at frequencies approaching the (bias- dependent) periodic-line cutoff frequency in the vicinity of 7 GHz. The inflections in the group delay curve for zero bias voltage are due to source and load reflections from the ends of the 3 lines; the group delay of a 45Ω line measured in a 50Ω system differs from the delay measured in a 45Ω system by terms in the source and load reflections.

Wavefront compression measurements were performed using a 25 ps sampling oscilloscope. As no standard pulse generators having appropriate risetimes or falltimes were available, a 2.5 volt, 525 ps negative-going step-function was generated by amplifying the oscilloscope's 25 ps tunnel- diode pulse generator with a ∼800 MHz inverting amplifier. These edges were longer than the required 360 ps; to obtain compression to a minimum compressed falltime, the input bias was increased 0.5 V to increase the line's compression ($T(v_h) - T(v_l)$). The bias change increased C_{ls} and hence decreased ω_{per} ; the resulting compressed output falltime was 100 ps (Fig. 2.8). A step-function input is not necessary for fast edge generation; driving the line with a 15 dBm, 300 MHz sinewave with -1.25 V bias generated a sawtooth output waveform with 100 ps falltime (fig 2.9).

The nonlinear line is also a broadband voltage-controlled linear phase- shifter; at frequencies from DC to $\sim \omega_{per}/4$ the line provides constant voltage-controlled

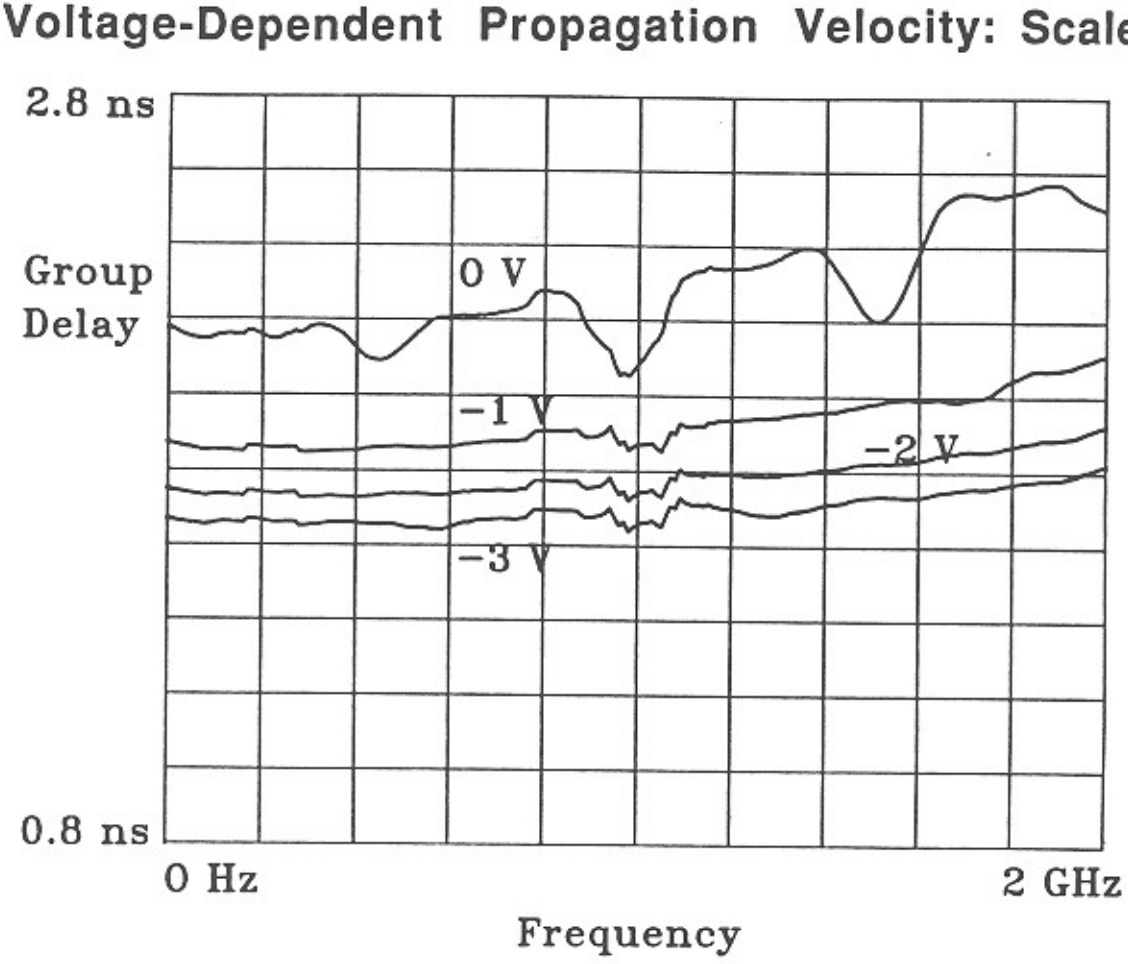


Figure 2.7: Group delay (ie $-d\phi/d\omega$ of S_{21}) vs. bias voltage and frequency for the 45-diode scale-model nonlinear transmission lines. The delay includes that of the 50Ω input and output interfaces of the 3 15–diode line sections.

Compression of 500 ps Risetime to 100 ps: Scale Model

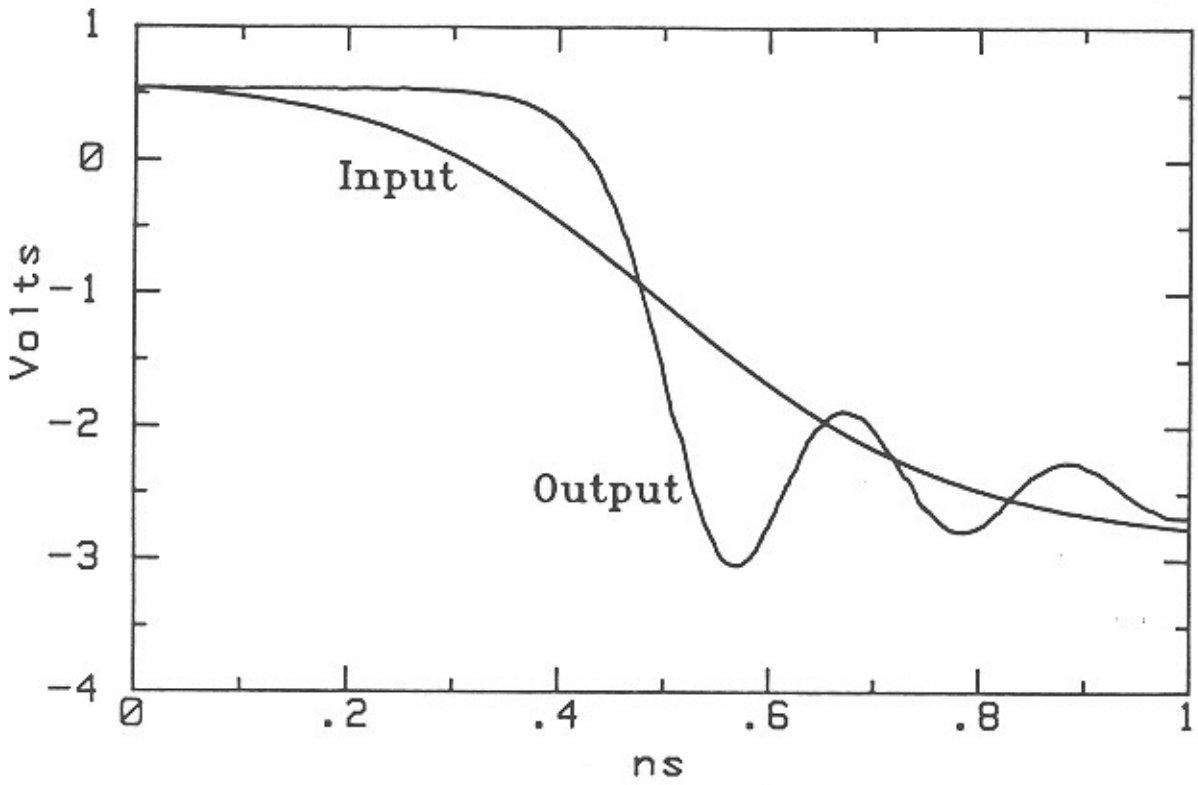


Figure 2.8: Measured compression of 525 ps input falltime to 100 ps output falltime on scale model nonlinear transmission line.

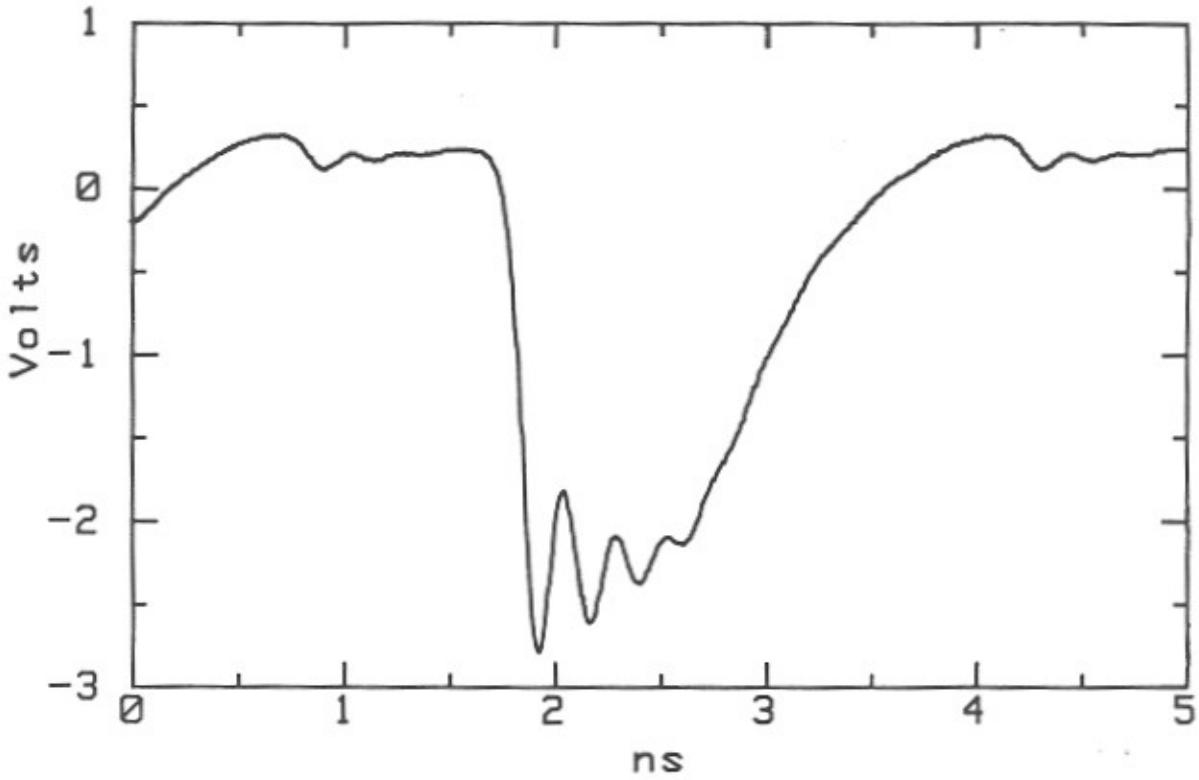


Figure 2.9: Output of scale model nonlinear transmission line with input driven with 300 MHz sinusoid. Fall time is 100 ps.

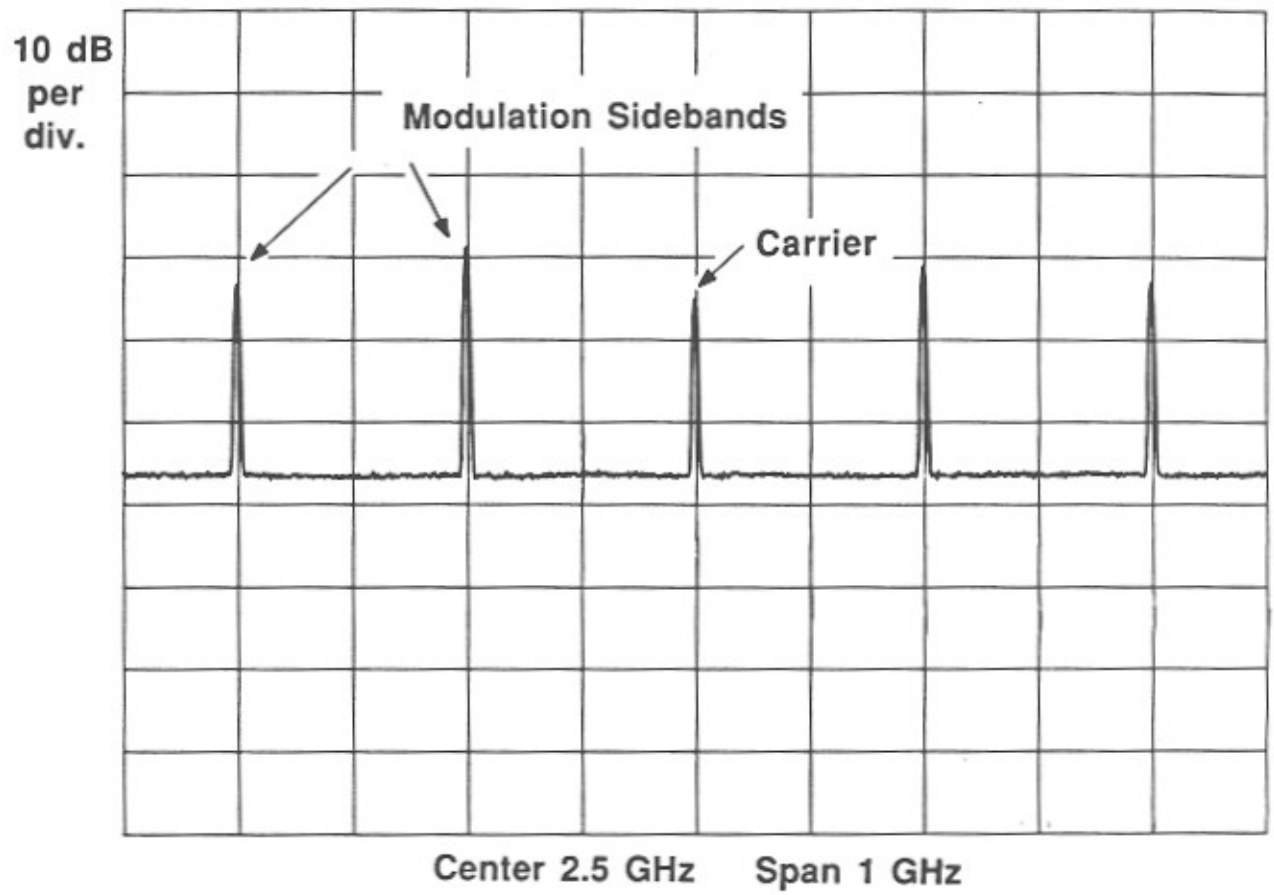


Figure 2.10: 2.5 GHz carrier phase-modulated at 200 MHz by scaled nonlinear transmission line.

group delay and small insertion loss, while Z_0 shows only small variation with voltage. Superimposing on the line a DC bias plus carrier and modulation frequencies results in phase modulation. The carrier ω_0, modulation ω_m, and generated sidebands $\omega_0 \pm n\omega_m$ are phase-matched provided that all these frequencies much less than ω_{per}, thus the phase modulation is broadband in both the carrier and modulation frequencies. By Eq. 1.6, spurious harmonics of ω_0 and ω_m are generated; second-order and third-order intercepts for the scale model were +12 and +10 dBm at 1 GHz and -1 volt bias. On the scale model, carriers from 100 MHz to 2.5 GHz were phase-modulated at frequencies from 10 MHz to 1 GHz (Fig. 2.10). As constructed, the line is a poor phase-modulator due to spurious outputs, including the modulation signal and harmonics of the carrier; a balanced structure might eliminate these outputs. The phase modulation experiment of Fig. 2.10 does demonstrate the broad bandwidths of periodic networks, and suggests their possible application to a variety of microwave signal-processing operations.

Since the scale-model results are in reasonable agreement with the SPICE simulations and approximate analysis, the design can now progress from analytical considerations to the constraints imposed in fabricating a monolithic device.

References.

[1] R. Landauer: "Shock Wave Structure in Nonlinear Dielectrics", *Ferroelectrics*, 1976, Vol. 10, pp. 237-240.

[2] R.H. Freeman and A. E. Karbowiak: "An investigation of nonlinear transmission lines and shock waves", J. Phys. D: Appl. Phys. **10** 633- 643, 1977.

[3] Hasnain, G., Dienes, A., and Whinnery, J.R.: "Dispersion of Picosecond Pulses in Coplanar Transmission Lines" IEEE Trans. on Microwave Theory and Tech., 1986, MTT-34, No.6, pp. 738-741.

[4] Ketchen, M.B., Grischkowski, D., Chen, T.C., Chi, C-C., Duling, I.N., Halas, N.J., Halbout, J-M., Kash, J.A., and Li, G.P.: "Generation of subpicosecond electrical pulses on coplanar transmission lines", Appl. Phys. lett., 1986, 48, pp.751-753.

[5] Wigington, R.L., and Naham, N.S.: "Transient Analysis of Coaxial Cables Considering Skin Effect", Proc. IRE, February 1956, pp. 166-174.

[6] Rodwell, M.J.W., Bloom, D.M., and Auld, B.A.: "Nonlinear Transmission Line for Picosecond Pulse Compression and Broadband Phase Modulation" , Elect. Lett., 1987, Vol. 23, No. 3.

[7] K.C. Gupta, Garg, Ramesh , and I.J. Bahl:*Microstrip Lines and Slotlines*, Artech House, 1979.

Chapter 3: Design of the Monolithic Device

In the previous two chapters we have studied nonlinear transmission lines from the viewpoint of network theory. Given some knowledge of feasible circuit parameters, we have concluded that wavefronts with transition times on the order of a few picoseconds can be generated. Some constraints imposed by fabrication were considered: diode cutoff frequency was bounded, interconnect impedance could not be picked arbitrarily, decreased Schottky contact capacitance per unit area (useful in the continuous lines) comes at the expense of degraded diode cutoff frequency.

In designing the monolithic device, we must define the structure, dimensions, and means of fabrication of both the diodes and the interconnecting transmission lines. Design of the fabrication sequence itself will be deferred until Chapter 4. Here we will consider the physical design of the monolithic nonlinear transmission line on GaAs given predefined constraints on dimensions and dopings.

The diode's doping and dimensions, constrained by capabilities of epitaxy, implantation, and lithography, determine the maximum wavefront voltage, the diode cutoff frequency, and (through limits on junction capacitance) the periodic-line cutoff frequency. The dimensions of the interconnections, constrained by lithography, determine their loss and their characteristic impedance. Choice of the line dimensions is also governed by their effect on dispersion and line radiation. Finally, the diodes must be connected to the lines; the parasitic reactances arising from these connections are a function of the diode and line dimensions, and the design is further constrained.

The design of the nonlinear transmission line and its fabrication processes evolved over the past two years. At that time, no electronic devices had been designed by our research group, nor had any electronic devices (discrete or monolithic) been fabricated in the Ginzton laboratory microstructures facility. Decisions were based on intelligent choice (occasionally), on imitation of successful devices and processes, on ignorance, and on iteration: device designs and fabrication sequences frequently failed, and parameters were then adjusted. Slowly, we converged to a design which could be fabricated with our limited processing facilities.

Given the current resolution of our fabrication processes, the first device design is not optimum, and we are currently developing improved devices. While the current process resolution allows ~ 3 μm design rules, the current design is close to an optimum physical design if the minimum feature size allowed by process resolution were 10 μm. Abandoning the reality of a long and iterative design history

in favor of a clearer and more concise (but fictitious) development, we present the design development as constrained by 10 μm design rules.

3.1 Diode Design

We start by defining the structure of the Schottky diode. Schottky diodes can be fabricated on both silicon and GaAs, and the doped layers can be formed by diffusion, ion implantation, or epitaxial growth. While millimeter-wave diodes can be made in silicon, the high intrinsic carrier concentration results in high intrinsic (undoped) conductivity; silicon is a poor, lossy dielectric. Transmission lines fabricated on silicon substrates suffer from high dielectric losses and are unsuitable for microwave and picosecond propagation.

For gallium arsenide, ion implantation and epitaxy are the most refined process technologies. While Schottky diodes are frequently fabricated on ion-implanted material, their characteristics are unsuitable for the nonlinear transmission line. High-conductivity buried layers beneath the Schottky contact area are hard to attain, and the diode current must instead pass laterally from beneath the Schottky contact to the ohmic contact, through a shallow implanted region of only moderate doping. The resistance of this path, termed "spreading resistance", results in a poor diode cutoff frequency.

Schottky diodes fabricated on epitaxial layers of GaAs have been reported with cutoff frequencies in the range of 10–20 THz for discrete diodes [1– 5], while planar epitaxial diodes with structures suitable for monolithic integration (Fig. 3.1) can be realized with cutoff frequencies in the range of 1–5 THz. While the diode can be fabricated using several epitaxial technologies, molecular beam epitaxy (MBE) is prevalent in the research community, and is used for our devices. In the structure of Fig. 3.1, a Schottky contact is made by evaporation of some nonreactive metal onto a lightly-doped N- layer (Fig. 3.2). The depletion layer, from which arises the diode's voltage-variable capacitance, extends a distance Y_d into the N- layer. The diode (displacement) current passes vertically through the depletion capacitance and the undepleted portion of the N- layer, and subsequently passes laterally through a thick, heavily-doped, low- resistance buried N+ layer to the ohmic contacts.

Access to the N+ layer for the ohmic contacts is typically provided by a recess etch. Ohmic contacts are formed by evaporation of a metal-dopant alloy into the contact regions. After metal deposition, the wafer is heated, the dopant diffuses from the ohmic metal into the N+ layer, and a shallow, very heavily doped region is

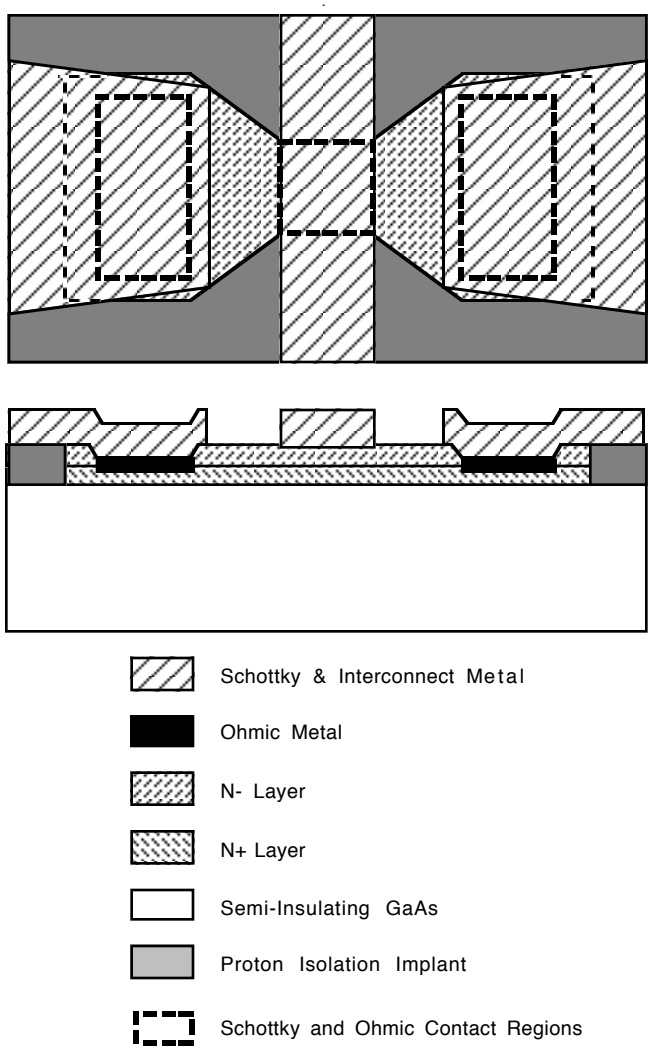


Figure 3.1: Plan view and cross-section of planar epitaxial Schottky diode on GaAs. A Schottky contact is formed in the region where metal intersects unimplanted areas on the N- layer.

formed immediately below the contact. The resulting metal-semiconductor junction has a very narrow potential barrier, and current readily flows by tunneling [6]. To minimize series resistance, two parallel cathode connections (ohmic contacts) are provided on opposing sides of the Schottky contact.

The doped epitaxial layers must be removed or rendered insulating in undesired locations. In many circuits, electrical isolation between the cathodes of distinct diodes is necessary. In the nonlinear transmission line, the semiconducting layers must be removed below the interconnecting transmission lines both to prevent the formation of an undesired Schottky contact between the line and the layer beneath, and to provide a low loss (i.e. high resistivity) dielectric for the transmission lines. These disparate functions, all termed "isolation", are achieved by removal of the layers by chemical or ion-beam etching, or conversion of the doped layers to a semi-insulating state by implantation of appropriate ions. Isolation by chemical etching is easy to implement, but results in nonplanar surfaces which cause difficulties in subsequent photolithography and metal deposition. Difficulties also arise from the formation of undesired contacts between buried conducting layers and any metal connections that cross the vertical or sloping edge of the isolation etch. Isolation by ion implantation results in planar surfaces, but limits to the ion penetration depth and implanted particle density further constrain the diode design. Our devices use isolation by ion implantation.

3.1.1 N- layer Doping and Thickness.

The choice of N- layer doping and thickness, together with the mask minimum design rules, sets the minimum zero-bias capacitance and the maximum signal voltage, and influences the diode cutoff frequency. The parameters of the N- layer are chosen to maximize the diode cutoff frequency; the optimum parameters are a function of the diode dimensions as constrained by the minimum dimensions allowed by the fabrication processes. If the minimum dimensions are sufficiently large, the N- layer thickness selected by maximization of ω_{diode} will be larger than the feasible isolation implantation depth. In this case, the diode is designed to maximize ω_{diode} given a maximum allowable N- layer thickness.

A simplified cross-section of the diode, including its dimensions and its equivalent circuit, is shown in Fig. 3.2. The Schottky contact has dimensions $a \times w$, where w is the extent of the contact perpendicular to the cross-section. Following the development of Section 1.3, the diode capacitance $C_{depl}(V_c)$ and the undepleted N- layer resistance R_{N-} are given by:

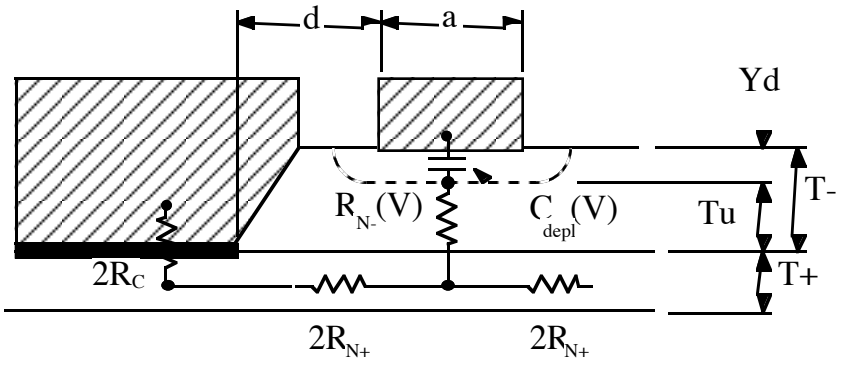


Figure 3.2: Vertical dimensions and equivalent circuit elements for the planar Schottky varactor diode.

$$C_{depl}(V_c) = aw\sqrt{\frac{q\epsilon N_{d-}}{2(\phi - V_c)}} \qquad R_{N-}(V_c) = \frac{1}{aw\sigma_-}\left(T_- - \sqrt{\frac{2\epsilon(\phi - V_c)}{qN_{d-}}}\right), \quad (3.1)$$

where $\epsilon \simeq 13.1\epsilon_0$ is the permittivity of GaAs, $\sigma_- \simeq q\mu_- N_{d-}$ is the conductivity of the N- layer, and μ_- its electron mobility. Following standard notation, q is the magnitude of the electron charge and ϵ_0 is the permittivity of vacuum.

To minimize the resistance R_{N-} of the undepleted N- layer, the N- layer thickness T_- should be small. As in Section 1.3.1, the depletion-layer capacitance decreases with voltage only until the depletion edge reaches the N+/N- interface, and thereafter remains constant. The N- layer thickness T_- is selected so that the layer is fully depleted at the peak negative voltage V_l:

$$T_- = \sqrt{2\epsilon(\phi - V_l)/qN_{d-}} \quad (3.2)$$

Increased N- layer doping (given a minimum diode area set by minimum design rules) increases C_{j0} but decreases the resistance of the undepleted portion of the N- layer R_{N-}. The N- layer doping can thus be chosen to maximize the diode cutoff frequency. The N+ layer resistance and contact resistance are independent of the N- layer doping but vary inversely with the contact length w, $R_{N+} = r_{N+}/w$, $R_c = r_c/2w$ (neglecting flaring of the ohmic contact), hence the zero-bias diode RC time-constant,

$$(R_{N-}+R_c+R_{N+})C_{j0} = \frac{\epsilon}{q\mu_n N_{d-}}\left(\sqrt{1 - V_l/\phi}-1\right) + (r_c/2+r_{N+})a\sqrt{\frac{q\epsilon N_{d-}}{2\phi}}\ , \quad (3.3)$$

can be minimized by appropriate choice of N_{d-}. The N- layer thickness is then calculated from Eq. 3.2. Since both the contact width a and the N+ layer resistance r_{N+} are determined by the minimum mask design rules, the optimum doping N_{d-} increases and optimum layer thickness T_- decreases as the minimum feature size decreases. Adjustment of the N- layer doping to increase the diode cutoff frequency will also change the minimum diode capacitance and hence the maximum achievable periodic-line cutoff frequency ω_{per}; the minimum compressed falltime is a function of both cutoff frequencies.

With 10 μm design rules, the optimum N- layer thickness determined by Eqs. (3.2) and (3.3) is greater than that permitted by the isolation implantation process. Allowing margins for errors in published data, the available energy for ion implantation limits the combined thicknesses of the N- and N+ layers to 1.4 μm, of which $T_- = 0.6\mu$m is allocated for the N- layer. If the design is constrained by this depth limitation, the N- layer doping N_{d-} is selected so that the layer is (with some margin for variation in layer thickness) fully depleted at the peak negative voltage V_l. The minimum allowable doping is then

$$N_{d-} \geq \frac{2\epsilon(\phi - V_l)}{qT_-^2} \ . \tag{3.4}$$

Since larger signal voltages result in greater compression and hence more ready experimental demonstration, the parameters of the initial design were somewhat inconsistently chosen so that, while $Z_{ls} = 50\ \Omega$ at $V_l = -2$ volts, signals as large as $V_l = -6$ volts could be compressed. With 0.1 μm allocated for variations in $T-$ (finite N- to N+ transition region, variations in MBE growth), a doping $N_{d-} = 3(10^{16})/cm^3$ was selected; at 6 volts bias, the depletion edge extends to 0.53 μm depth. With minimum design rules setting a $10\mu\text{m} \times 10\mu\text{m}$ contact area, the zero-bias capacitance is 50 fF and the resistance of the undepleted N- layer at zero-bias is 2.5 Ω. With increased process resolution, thinner and more highly-doped N- layers will be desirable.

3.1.2 N+ layer Doping and Thickness.

The N+ layer serves as a low-resistance conduction path from beneath the diode region to the ohmic contacts. To minimize the series resistance, the path length d should be minimized, and the path width w, path depth T_+, and conductivity σ_- should be maximized. The geometry of the N+ layer is shown in Fig. 3.3. The increased width of the unimplanted N+ layer near the ohmic contacts permits a small reduction in the N+ layer resistance. If the angle of the flaring is small, then

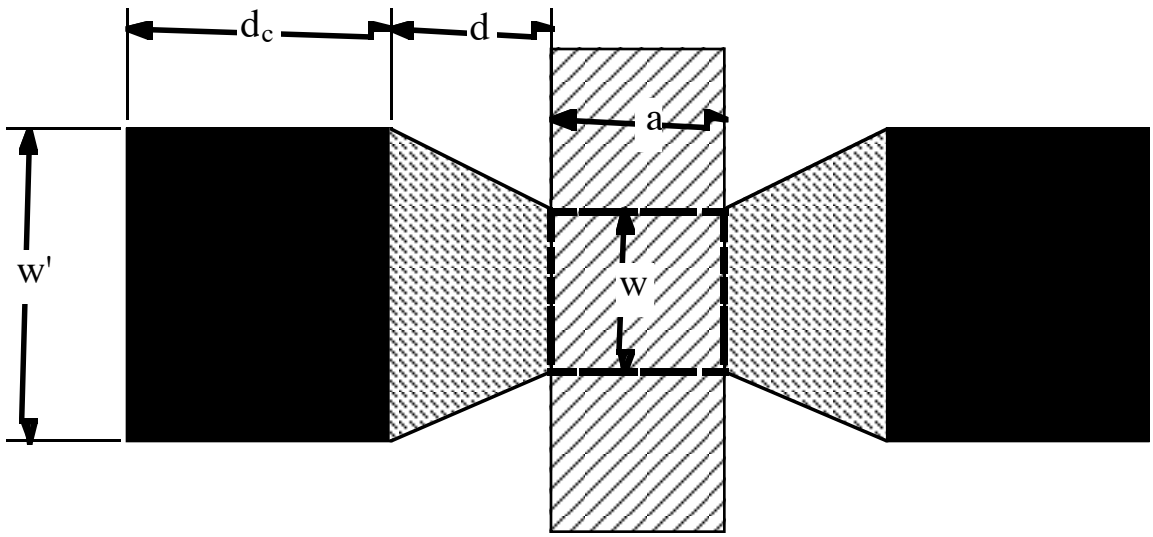


Figure 3.3: Dimensions of the N+ layer and the ohmic contacts.

$$R_{N+} \simeq \frac{1}{2\sigma_+ T_+}\left(\frac{a}{6w} + \frac{2d}{w+w'}\right) . \tag{3.5}$$

The factor of 1/2 arises from the pair of ohmic contacts, and the term $(a/6w)$ is the spreading resistance under the Schottky contact. The conductivity σ_+, a nonlinear function of doping at typical doping levels, can be found in Sze [6]. The dimensions a and w, are, in this case, set at the minimum design dimension of 10 μm, while $d = 12\mu$m. The contact width was chosen as $w' = 16\mu$m; with greater contact widths, the currently flow is strongly two-dimensional, Eq. (3.5) does not hold, and resistance decreases only slowly with w'. Large flare angles also violate the intent of the specified minimum feature size; with a large flare angle, a small misalignment of the the Schottky/interconnect mask relative to the isolation implant mask will result in a substantial increase in the Schottky contact area.

The remaining parameters are constrained by the capabilities of isolation implantation and epitaxial growth. With a maximum implant depth of 1.4 μm and with an N- layer thickness of 0.6 μm, the N+ layer thickness T_+ is 0.8 μm. The limits to doping have not been clearly established. Dopant concentrations N_d much larger than the conduction- band effective density of states, $N_c = 4.7(10^{17})/cm^3$ at 300 K, result in a Fermi level above the donor level (~ -6 meV), and a partially ionized donor population. The free electron concentration and material conductivity continue to increase with increases in doping for $N_d < 10^{20}/cm^3$ [6,7].

With some MBE growth systems, high dopings can require either dopant fluxes beyond the capabilities of the source, or correspondingly reduced growth rates. At the time of this project, the MBE system at Stanford was limited to $\sim 10^{18}/cm^3$

doping for reasonable growth rate and acceptable defect density; this doping was too low for our requirements. Instead, we were provided MBE material through collaboration with Varian III-V device center, who could provide material with doping exceeding $10^{19}/cm^3$. The N+ layer doping may then be limited by isolation implantation. The required implant dose is proportional to the pre- implantation doping. Implantation was available through a commercial service. For the particles used (protons), the service **estimated** a maximum implant dose of $5-7(10^{14})/cm^3$, although dopings up to $\sim 3(10^{15})/cm^3$ could presumably be attained through multiple or extended-time implantations at added cost. Even with the availability of high implant doses, we had low confidence in isolation of layers with doping exceeding $\sim 3(10^{18})/cm^3$; a search of the literature revealed no reports on proton isolation of dopings greater than $4(10^{18})/cm^3$, and no reports we found on millimeter-wave diodes had N+ doping greater than $3(10^{18})/cm^3$. Based on limited knowledge of the process technology, the N+ layer doping was selected as $N_{d+} = 3(10^{18})/cm^3$. The N+ layer resistance can now be calculated. At $3(10^{18})/cm^3$ doping, $\rho \simeq 8(10^{-4})$ Ω-cm, and the resistance is $R_{N+} \simeq 6.3$ Ω. This high resistance arises from the large 10μm minimum dimensions.

Because the diode current flows laterally from the ohmic contacts, the contact resistance is a function of both ohmic contact metallurgy and N+ layer sheet conductivity [8]. If the resistance per unit area for currently flowing through and perpendicular to the ohmic contact is ρ_c, then the contact resistance of current flowing across the ohmic contact and then parallel to the N+ layer is

$$R_c = r_c/2w' \ , \tag{3.6}$$

where

$$r_c = \sqrt{\rho_c/\sigma_+ T_+} \quad \text{if} \quad d_c \gg L_c \equiv \sqrt{\rho_c \sigma_+ T_+} \ , \tag{3.7}$$

and L_c is the $1/e$ penetration depth of current beyond the edge of the ohmic contact. While ρ_c is set by processing, an N+ layer having a large sheet conductivity ($\sigma_+ T_+$) reduces the contact resistance. In Chapter 4, it is seen that a contact resistance per unit length of $r_c = 0.03$ Ω-mm was achieved, and the contact resistance was 0.9 Ω.

The constraints of 10 μm minimum feature size and 1.4 μm maximum implantation depth result in the design of a varactor diode whose capacitance varies as that of a step-junction for voltages in the range of 0 to -6 volts, whose zero-bias capacitance is 50 fF, whose series resistance is 9.7 Ω, and whose zero-bias cutoff frequency is $\omega_{diode} = 2\pi \times 330$ GHz. With improved design rules the diode time constant can be decreased (i.e. the cutoff frequency extended). Assuming that no flaring is used ($w = w'$) , the diode zero-bias time constant is

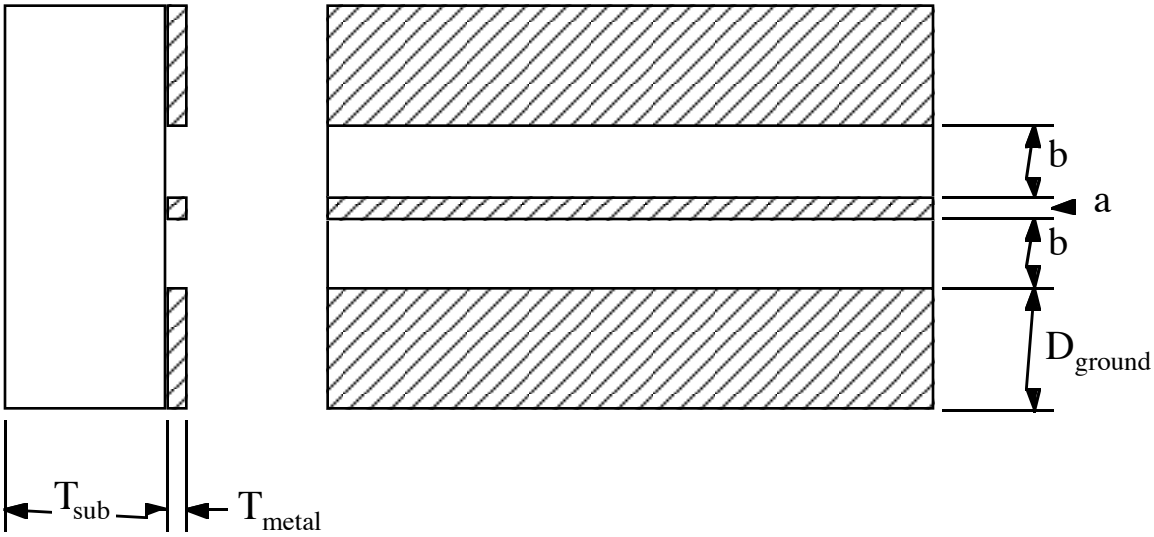


Figure 3.4: Coplanar waveguide transmission line.

$$(R_{N-} + R_c + R_{N+})C_{j0} = \frac{\epsilon}{q\mu_n N_{d-}}\left(\sqrt{1 - V_l/\phi} - 1\right) + a\left(\frac{\sqrt{\rho_c/\sigma_+ T_+}}{2} + \frac{d + a/6}{2\sigma_+ T_+}\right)\sqrt{\frac{q\epsilon N_{d-}}{2\phi}}\ , \tag{3.8}$$

To maximize ω_{diode}, the diode width a and the ohmic-Schottky separation d should be the minimum allowed by process resolution.

3.2 Interconnecting Transmission-Line Design

After the diode doping and geometry have been selected, the inductive interconnects are then designed. Inductance can be implemented in monolithic form using planar spiral inductors or short sections of high- impedance microstrip or coplanar waveguide transmission lines. From the viewpoint of circuit theory (Chapter 2), interconnecting lines of very high impedance (i.e. inductors) are preferable. Planar spiral inductors provide large inductance per unit area, reducing the required die area, but they require processing with two levels of metal, exhibit resonances and high skin losses, and are poorly modeled [9].

Short sections of high-impedance microstrip transmission lines provide a high-Q series inductance, and the parasitic capacitance is the well-modeled transmission-line shunt capacitance. Fabrication is complex, requiring wafer thinning, backside

metallization, and via etching and plating. The inductance of the vias connecting circuit elements to the backside ground plane is a significant circuit parasitic; scaling of the line's dimensions to reduce the via inductance and to decrease the line dispersion requires wafer lapping to thicknesses below 100 μm. Fabrication of coplanar waveguide transmission lines (Fig. 3.4) is simpler, requiring only a single liftoff step to define the conductor geometry. No wafer thinning is required. If the substrate thickness T_{sub} is significantly greater than the ground-plane spacing $2b$ and the metal thickness T_{metal} is much smaller than the center-conductor width a, the characteristic impedance is a function only of b/a,. The waveguide parameters are thus controlled exclusively by lithography.

Given the choice of a coplanar line with thin metal and a thick substrate, and using the quasi-TEM (i.e. low-frequency) approximation, the group velocity is [10]

$$v_{group} \simeq c_0/\sqrt{\epsilon_{re}} \ , \tag{3.9}$$

where the effective dielectric constant ϵ_{re} is the average of the dielectric constants of GaAs ($\epsilon_r \simeq 13.1$) and of air:

$$\epsilon_{re} = \ \frac{\epsilon_r + 1}{2} \ . \tag{3.10}$$

Rearranging Gupta's [10] expressions for the characteristic impedance Z_1 of the high-impedance interconnects:

$$Z_1 = \frac{30}{\sqrt{\epsilon}_{re}} \ \ln\left[2\frac{1+\sqrt[4]{1-k^2}}{1-\sqrt[4]{1-k^2}}\right] \quad 0 \le k \le 1/\sqrt{2}$$
$$\text{where} \quad k \equiv \frac{a}{a+2b} \ . \tag{3.11}$$

Equation (3.11) is plotted in Fig. 3.5. The expression is approximate, but with a claimed error of less than 3 parts per million. A more easily evaluated expression, accurate to within 1% for $b > a$, is obtained by expanding 3.11:

$$Z_1 \simeq \frac{60}{\sqrt{\epsilon_{re}}} \ \ln(4/k) = 21.6\Omega \ \times \ln(4/k) \quad \text{for} \quad k < 1/3 \ . \tag{3.12}$$

Line impedance varies only as $\sim \ln(b/a)$, so that impedances above 100 Ω require either very large ground-plane separations b or very narrow center-conductor widths a. Because these dimensions are constrained interconnect impedances above 100 Ω are unattractive.

A narrow center conductor results in high DC resistance. The metallization thickness $T_{metal} = 1.6$ μm is set by process limits, and gold is selected for low resistivity ($\rho = 2.4 \times 10^{-8}\Omega$-m). The DC loss of the nonlinear transmission line

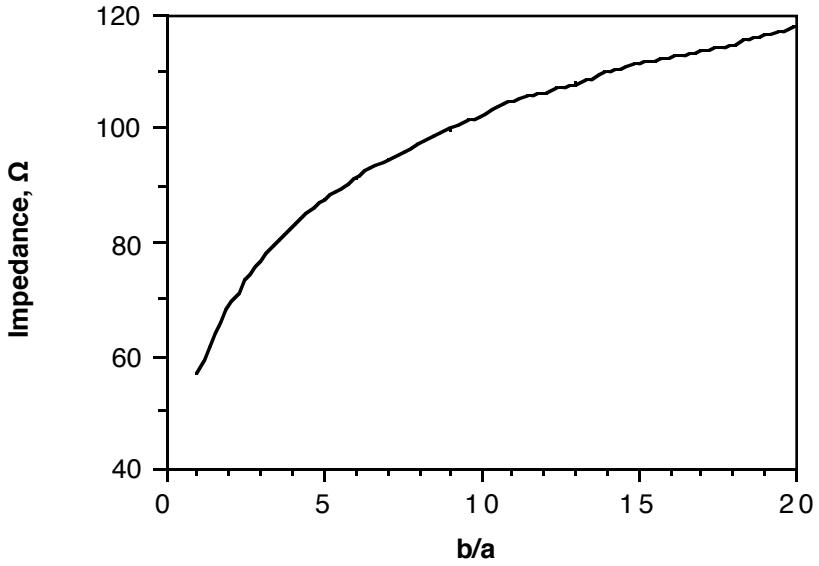


Figure 3.5: Characteristic impedance of a coplanar waveguide transmission line on GaAs, assuming a thick substrate and thin metallization.

determines the attenuation of a step-function input propagating on the line. The line series resistance is

$$R_{line} = nl\rho_{metal}/aT_{metal} \ , \tag{3.13}$$

which results in a DC transmission of $2Z_0/(2Z_0+R_{line})$. As before, n is the number of line sections and l is the interconnect length. Both the DC resistance and the mask minimum dimensions set a lower bound on the linewidth. In Chapter 2, we considered and dispensed with the effect of skin impedance (Eq. 2.23). The pulse broadening due to skin impedance varies as $1/a^2$, and may be significant if picosecond pulse compression is attempted on $\sim 1\mu$m width lines.

Large conductor-ground spacings b are also undesirable. As the lateral dimensions of the line $(a+2b)$ approach the substrate thickness T_{sub}, the transmission-line field region extends through the substrate to the back surface. The line impedance and phase velocity diverge from Eqs. (3.9)- (3.11) and become sensitive to the boundary conditions at the substrate back surface [9,10]. The circuit characteristics then vary as the substrate is placed on different surfaces, and if the substrate backside or its supporting surface are rough, losses will result from scattering into high-order radiative modes. With a standard 400 μm substrate, the lateral dimensions $(a + 2b)$ of the line should be kept below $\sim$ 200 μm to avoid these effects.

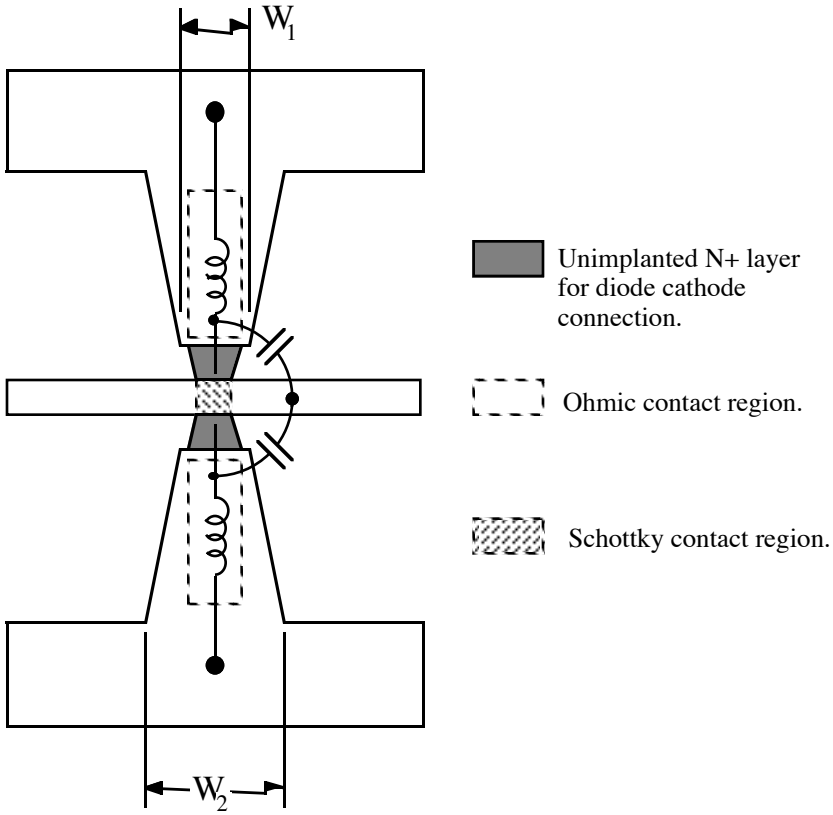


Figure 3.6: Integration of Schottky diode with coplanar transmission line, and resulting circuit parasitics.

Before this limit is reached, connection of the diodes to the lines becomes difficult.

3.3 Connection of the Diodes to the Line.

Figure 3.6 shows the integration of the diodes with the coplanar transmission lines. A stripe of unimplanted semiconductor runs perpendicular to the transmission line, and the Schottky contact is formed in the region of overlap between the transmission line signal conductor and the unimplanted semiconductor. The ohmic contacts, placed close to the Schottky contact to minimize resistance, are connected to the distant ground planes by metal fins. The fins and the buried N+ layer also provide connection between the two ground planes of the coplanar waveguide, suppressing propagation of the unbalanced ("slotline") mode.

The fins introduce parasitic shunt capacitance and parasitic series inductance. Neither parasitic is readily calculated. Fin inductance can be estimated by treating the fins as short sections of a coplanar transmission line. From Eqs. (3.9) and

(3.12), the inductance of a coplanar transmission line of center-conductor width w, ground-plane separation β and length b is approximately

$$L = b \times 2(10^{-7}) \ \ln(4 + 8\beta/w) \ \ \text{H/m}. \tag{3.14}$$

But the fins are not true sections of coplanar waveguide, and the distance β of the return current path is undefined. Clearly, β is not dependent on the separation between diodes, as the inductance L would then be unbounded for a line shunted by a single diode. The effective distance β of the ground current path should be on the order of the fin length b. With the additional approximation $w \approx (w_1 + w_2)/2$, the inductance of the parallel pair of fins is estimated as

$$L_{fins} \sim b \times 10^{-7} \ \ln\left(4 + \frac{16b}{w_1 + w_2}\right) \ \ \text{H/m}. \tag{3.15}$$

Clearly, the signal-ground spacing b must be minimized. Some reduction in inductance is also achieved through using wide fins, but the fin width w_1 adjacent to the line must be kept small to minimize the capacitive loading of the line.

The ground fins extend into the transmission-line field region and increase the line shunt capacitance in the vicinity of the diodes. The effect is strongly two-dimensional, with the fields perturbed over a length on the order of the transmission line width b, and the fin capacitance cannot be modeled as that of a low-impedance line section of length w_1. The effect of fin capacitance was determined by scale modeling. Fin capacitance adds an additional periodic shunt capacitance to the nonlinear transmission line. The unloaded (i.e. fins but no diodes) transmission line, becomes a periodic structure, with a decreased impedance Z_1' and increased delay τ', and with a periodic-line cutoff frequency. The loaded line consists of line sections of impedance Z_1 and length τ loaded by both the diode capacitance $C_{depl}(V)$ and the fin capacitance C_{fin}. If the fin capacitance is small in comparison with the diode capacitance, an equivalent model for the loaded line consists of a series of line sections of impedance Z_1' and length τ' loaded by the diode capacitance $C_{depl}(V)$. The effect of the fin capacitance can thus be assessed by measurement of the impedance and delay of an unloaded line having fins.

With $V_l = -2$ volts, $Z_{ls} = 50\ \Omega$, and a $C_{j0} = 50$ fF diode capacitance determined in Section 3.1, the required line length between diodes (Section 2.2.5) is 130 μm for $Z_1 = 100\ \Omega$ and 160 μm for $Z_1 = 90\ \Omega$. With the signal conductor width a set at the 10μm minimum dimension, $b = 100\mu$m at $Z_1 = 100\Omega$ and $b = 62\mu$m at $Z_1 = 90\Omega$. Lines of these dimensions were constructed at 100:1 scaling on Stiecast (TM), a commercial dielectric material available in a variety of permittivities. The delay and impedance of these lines was measured on time-domain reflectometer

(TDR) both before and after the addition of fins ($w_1 = 20\mu$m, $w_2 = 30\mu$m) at the above spacings. Addition of fins to the 100Ω line resulted in $\sim 7\%$ changes in both Z_1 and τ, while the change in parameters of the 90 Ω line could not be measured within the resolution of the reflectometer. The fin series inductance and shunt capacitance decrease as the signal-ground spacing b is decreased. With a fixed by minimum dimensions, a 90Ω line impedance was selected.

The design is now complete: diode parameters are as in Section 3.1, and a 90Ω line with $a = 10\mu$m sets $b = 60\mu$m, $\tau = 1.4$ps, and $l = 160\mu$m. The circuit values are as in the SPICE simulation, except that the simulation uses 8 Ω diode resistance rather than 9.6 Ω. More seriously, the fin inductance is not included in the model; the extreme imprecision of the estimate (Eq. 3.15) prevents its inclusion in the simulation. The estimated fin inductance is ~ 20 pH, which will resonate with C_{ls} at approximately 200 GHz, while $\omega_{per} = 140$ GHz. If the estimate is correct, the fin inductance will have little effect on compressed falltime at $R = 10\Omega$, but will have a significant but nondominant effect as the diode resistance is decreased. Finally, while approximate methods suggest that 45 sections are required for compression of a 25 ps input, and SPICE simulations indicate that ~ 55 sections are required, the mask design limited us to 42 sections. With 42 sections, the maximum input falltime is approximately 20 ps. From Eq. (3.13), the line series resistance is 10Ω, giving 10% attenuation during compression of step-functions. The simulations predict falltimes of approximately 6 picoseconds. Given the constraint of 10 μm minimum feature size, the device should compress of 20 ps, ∼6 volt step- functions to 6 ps output falltime, with 10% insertion loss.

References.

[1] Lundien, K., Mattauch, R.J., Archer, J., and Malik, R. : "Hyperabrupt Junction Varactor Diodes for Millimeter-Wavelength Harmonic Generators", IEEE Trans. MTT-31, 1983, pp. 235-238.

[2] Eric R. Carlson, Martin V. Schneider, and Thomas F. McMaster: "Subharmonically Pumped Millimeter-Wave Mixers" IEEE Trans. on MTT, Vol. MTT-26, No. 10, pp.706-715, October 1978.

[3] Keith S. Champlin and Gadi Eisenstein: "Cutoff Frequency of Sub- millimeter Schottky-Barrier Diodes", IEEE Trans on MTT, Vol. MTT-26, No. 1, pp. 31-34, January 1978.

[4] Joseph A. Calviello, John L. Wallace, and Paul R. Bie: "High Performance

GaAs Quasi-Planar Varactors for Millimeter Waves", IEEE Trans. Electron Devices, Vol. ED-21, No. 10, pp. 624-630, October 1974.

[5] John W. Archer and Marek T. Faber: "High-Output, Single- and Dual- Diode Millimeter-Wave Frequency Doublers", IEEE Trans on MTT, Vol. MTT-33, No. 6, pp. 533-538, June 1985.

[6] S.M. Sze: *Physics of Semiconductor Devices*, Wiley- Interscience,1981.

[7] S.M. Sze and J.C. Irvin, " Resistivity, Mobility, and Impurity Levels in GaAs, Ge, and Si at 300 K", Solid State Electron., 11, 599, 1968

[8] R.E. Williams, "GaAs Processing Techniques", Artech House, 1985.

[9] Robert A. Pucel: "Technology and Design Considerations for Monolithic Microwave Integrated Circuits" in *Gallium Arsenide Technology*, David K. Ferry, Editor, Howard Sams and co., 1985.

[10] K.C. Gupta, Garg, Ramesh , and I.J. Bahl:*Microstrip Lines and Slotlines*, Artech House, 1979.

Chapter 4: Device Fabrication and Evaluation.

In this chapter we will briefly review eight months of experimental work: the fabrication of the monolithic device and its evaluation using both electrooptic sampling and more conventional measurement techniques.

Fabrication of the devices outlined in Chapter 3 requires a gallium arsenide wafer with epitaxial N+ and N- layers, isolation, ohmic and Schottky contacts, and thick metalization for the interconnecting transmission line. The diodes and transmission lines are similar to those used in monolithic microwave integrated circuits (MMIC's); fabrication of the nonlinear transmission line generally follows well-established processes used in the fabrication of microwave GaAs IC's. Fabrication technologies for GaAs monolithic circuits are discussed in Williams [2] and in Howes and Morgan [3]. The process developed for the nonlinear transmission line differs slightly from that of MMIC's: the thickness and doping of the N+ layer cause difficulties with the isolation implant, Schottky contacts and interconnecting metal are formed in a single process step, and thick metal layers are produced by a single liftoff without subsequent electroplating.

The completed devices are tested by DC measurements to determine diode parameters and by microwave scattering parameter measurements to determine the circuit small-signal characteristics. In functional testing the line is driven by a pulse generator or microwave source, and the transition time of the shock wavefronts is measured at points internal to the line by direct electrooptic sampling.

4.1 Device Processing

Figure 4.1 shows fabrication of a diode and a section of line, starting with a [100]-cut GaAs wafer MBE-grown N+ and N- layers. Ohmic contacts are formed by a recess etch to the N+ layer, a self-aligned liftoff of ohmic metal (e.g. an alloy of metal and dopant), and an alloying step in which the dopant is driven into the N+ layer (Fig. 4.1a). For isolation, the ohmic and Schottky contact regions and the semiconductor between them are then masked, and the wafer is bombarded with hydrogen ions (Fig. 4.1b). Finally, Ti/Pt/Au is deposited on the wafer and patterned by liftoff, forming the interconnecting transmission lines, connecting to the ohmic contacts, and forming Schottky contacts in the region where metal overlaps unimplanted semiconductor (Fig. 4.1c).

The fabrication sequence, then, consists of three mask steps: ohmic contacts, isolation, and Schottky contact and interconnect metalization . All three process steps require liftoff photolithography for patterning of thick metal layers.

4.1.1 Liftoff Lithography.

Layers of metal must be deposited and patterned in all three process steps; the ohmic contacts require $\sim 0.3\mu$m metal, the implant mask requires approximately 1.6μm Au for effective protection of the Schottky contact, and the interconnect metal must be as thick as possible to minimize line series resistance. Thick layers of metal can be produced by electroplating a thinner deposited layer, but the electroplating requires the addition of several process steps, and is unattractive. Isolation implantation, in particular, requires seven steps, even without the use of electroplating. Because gallium arsenide is readily attacked by many acids, metal conductors cannot be patterned by acid etching, as is done in the fabrication of silicon integrated circuits. Instead, the metalization is patterned by the liftoff technique (Fig. 4.2). While liftoff is a common processing technique for GaAs [2, 3], liftoff of thick (1.6μm) layers is unusual and requires significant process development.

In liftoff lithography, some special photoresist treatment is used to produce resist edges with negative slopes (fig. 4.2a), i.e. overhanging edges. Metal deposition from some point source (an electron-beam evaporator) will then result in a film which is discontinuous over resist edges (fig. 4.2b). Dissolution of the resist in acetone will then remove all metal over the resist, leaving only metal which was deposited on the semiconductor substrate (fig 4.2c). Without the negative edge slope, a continuous film will be deposited on the wafer and the photoresist, and the unwanted metal lying upon the photoresist will not be removed by resist dissolution.

To attain edges with negative slope, the resist surface is hardened with chlorobenzene [4] and underlying resist is softened by incomplete softbaking. While an idealized positive photoresist will not develop in unexposed areas, all practical resists have some small development rate even if unexposed. If the normal pre-exposure resist-hardening bake ("softbake") is reduced in its temperature or duration, significant dissolved solvents and water will remain in the resist, increasing the development rate of unexposed resist. Underbaking introduces some difficulties, as development times are reduced and become more sensitive to the particular bake conditions. In combination with a resist surface which has been hardened by chlorobenzene , development will occur more rapidly in the depth of the resist than on the surface, generating an undercut edge profile. The edges of the resist then will have an overhanging lip, plus some recessed sidewall with a small positive slope (Fig. 4.3)

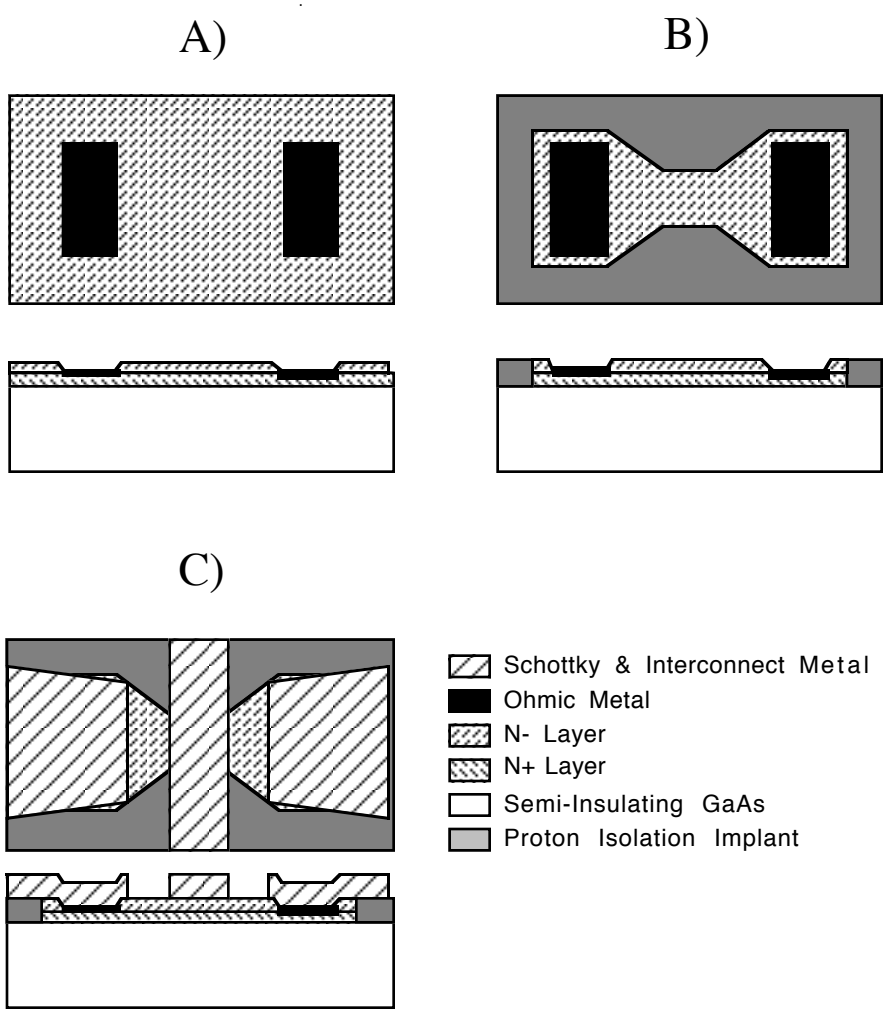


Figure 4.1: Process sequence for nonlinear transmission lines: a) MBE wafer after self-aligned recess etch and ohmic deposition, b) after proton isolation implantation, and c) completed device after Schottky/interconnect metal liftoff.

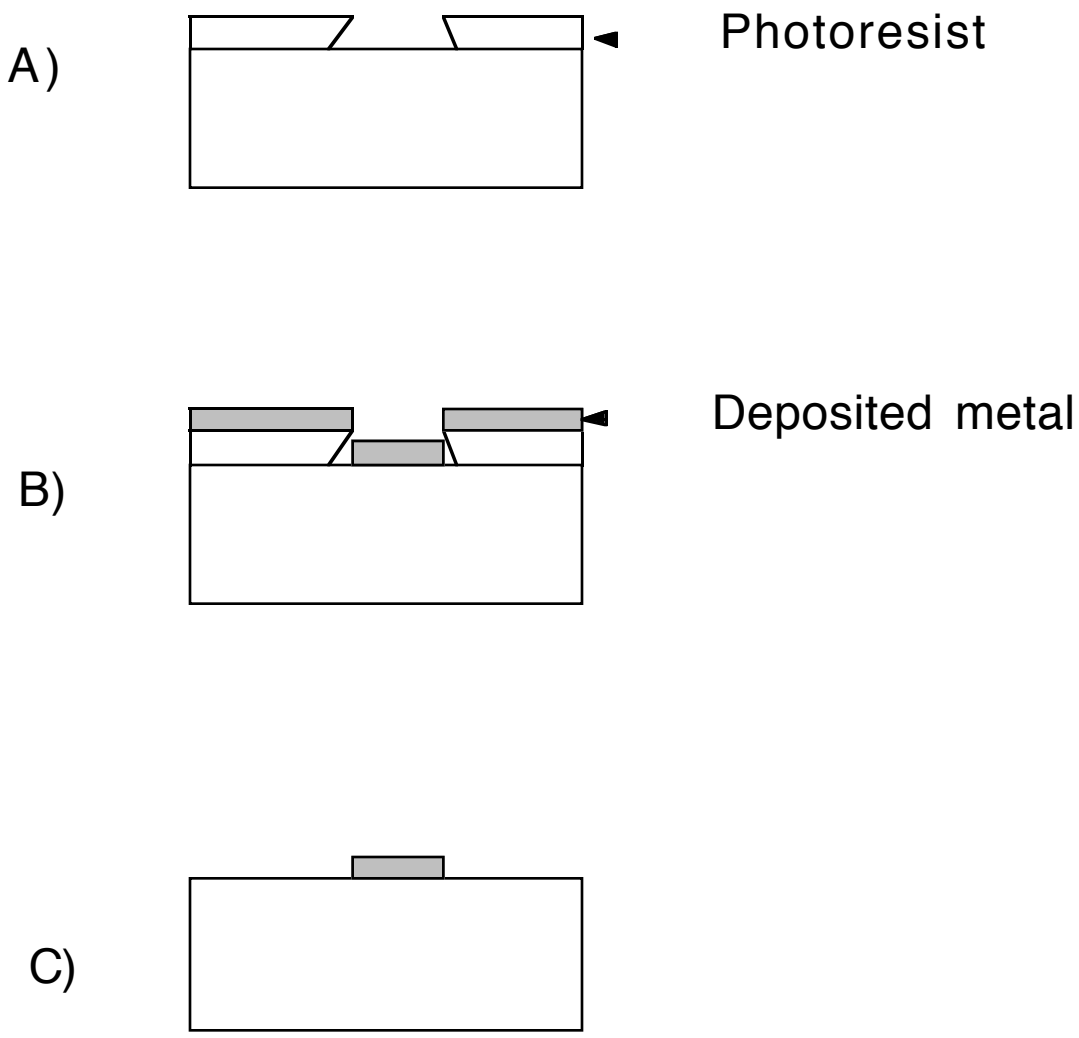


Figure 4.2: Liftoff photolithography for patterning of metals. Resist is treated to attain negative edge slope upon development, a). Evaporation of metal then results in discontinuities at the resist edges, b). Unwanted metal is removed by dissolution of resist in acetone, c).

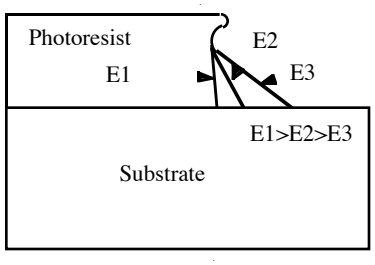


Figure 4.3: Resist edge profiles with the chlorobenzene process. Sidewall slope is a function of exposure energy.

Surface hardening is produced by a chlorobenzene soak immediately prior to development. Chlorobenzene selectively removes lighter (low molecular weight) polymers from the resist, leaving a resist surface which then dissolves more slowly in developer. As longer soaks permit progressively deeper penetration of the chlorobenzene solvent, the vertical thickness of the resist lip is a function of the soak time. A fifteen minute soak results in a $\sim 0.2\mu$m lip which appears resistant to deformation resulting from exposure to high temperatures during metal evaporation.

The recessed sidewall must have near-vertical slope. The desired profile is a slight outward slope, so that the overhanging lip of the resist is directly above the edge of the resist at the substrate. This results in accurate registration of the recess etch and the deposited metal in the self-aligned ohmic contact process. If the bottom edge of the resist projects far beyond the resist lip, the metal-semiconductor contact area will be reduced, reducing the area of Schottky contacts in the Schottky/interconnect step. Finally, with severe sidewall slope, a continuous metal film will be deposited on the sidewall, and liftoff will fail.

Edge slope is determined by resist optical attenuation and exposure time. Optical attenuation within the resist results in greater exposure at the resist surface than at depth. Unless the exposure energy is large, so that the photoactive compound is fully exposed at the base of the resist, the gradient in exposure intensity will result in a gradient in development rate with depth and a sloping sidewall (fig. 4.3). During process development, small adjustments are made in the exposure energy to adjust the sidewall slope.

Our exposure tool is a Karl Suss aligner operating at 300 nm. The most common positive photoresists, Hoescht 1400– and 1300– series, exhibit ~ 1.5 nepers/μm attenuation at 300 nm; the resulting sidewalls slope is approximately 45 degrees, and liftoff of 1.6μm metal films is unreliable. AZ-5214e resist (Hoescht) has much better transparency at 300 nm, and permits near-vertical sidewalls at 300 nm exposure. With resist spun on small GaAs pieces at 3000 rpm, the film thickness

after softbake is 1.65-1.75 μm, permitting reliable liftoff of 1.60 μm films (we have recently found that resist is thinner when spun onto full 2 inch wafers). The process sequence is listed in the appendix to this chapter.

4.1.2 Ohmic Contacts.

Ohmic contacts to the buried N+ layer form the diode cathode connections. The process sequence is as in Fig. 4.4. Photoresist is applied, softbaked, exposed, soaked in chlorobenzene, and developed, producing edges with negative slope (fig. 4.4a). The photoresist is then partially hardbaked to increase its stability during acid etching. The wafer is then etched (fig. 4.4c) to a depth of $0.75 \pm 0.05\mu$m using a NH_4OH:H_2O_2:H_2O solution in the proportions 14 : 2.4 : 200. Due to variations in the etch rate, ascribed to temperature variations and evolution of the H_2O_2, the etch is performed as a series of partial etches, with the etch depth monitored between etches by a surface profilometer. The chemistry of the etchant is discussed in Howes and Morgan [3]; strong anisotropic effects should be observed, and deep etch holes should have sloping sides determined by the (1 1 1), (1 $\bar{1}$ $\bar{1}$), ($\bar{1}$ $\bar{1}$ 1), and ($\bar{1}$ 1 $\bar{1}$) planes. The ($\bar{1}$ 1 $\bar{1}$) and ($\bar{1}$ $\bar{1}$ 1) planes have negative slope and can potentially cause breaks in the interconnect metal evaporation as it crosses the edges of the etch well, but the well depth after ohmic metal evaporation, 0.5μm, is small in comparison with the 1.6μm interconnect metal thickness, and breaks are unlikely.

Ohmic metal is then deposited by electron-beam evaporation (fig. 4.4c), and the excess metal removed by liftoff (fig. 4.4d). The ohmic metal contains a dopant which is subsequently diffused into the GaAs surface by a high- temperature treatment, termed "alloying". In the development of the nonlinear transmission line two similar ohmic metalizations were employed. The final devices used 0.5 μm eutectic Au/Ge (88%-12%), a 100 $\mathring{A}$ Ni barrier, and a 0.2 μm Au cap, all applied by resistive evaporation at Varian III-V device center. The ohmic contact is alloyed by heating to 450 C, above the eutectic melting point. The metallurgy of (Au– Ge)/Ni/Au contacts is complex, with a variety of phases formed at the interface [3]. With eutectic Au/Ge contacts, contact resistances of $r_c \simeq 0.03$ Ω–mm were attained; from Section 3.1.2, this low resistance reflects the low sheet resistivity of the N+ layer.

Ginzton Lab has only an electron-beam evaporator, and eutectic Au/Ge cannot be directly evaporated. To approximate the alloy, a series of thin layers of Ge and Au can be evaporated. Because the Ge and Au are distinct phases, the melting point of the contact is higher than for the eutectic. Thermal treatment of these contacts was investigated in our lab by Scott Diamond, who found a 450 C, 12 second alloying optimum. Contact resistances of $r_c \simeq 0.05$ Ω–mm are typical for

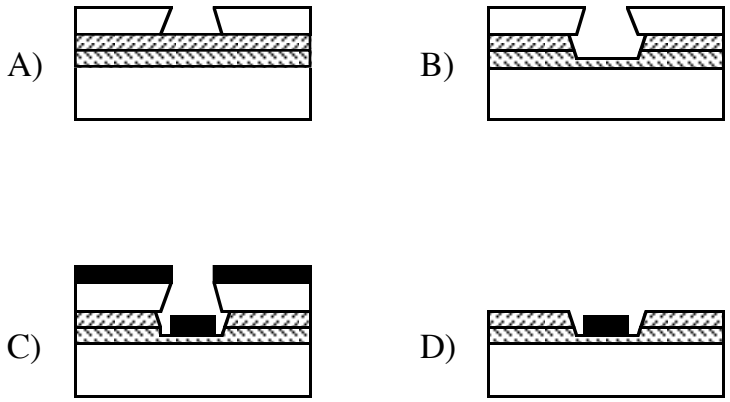


Figure 4.4: Ohmic contact sequence. Resist having a negative edge slope is produced by the chlorobenzene process, a), access is provided to the N+ layer by a recess etch, b), metal is evaporated, c), and lifted off, d).

contacts to similar N+ material.

4.1.3 Isolation Implantation.

Isolation can be performed by chemical etching or by implantation. Chemical etches result in sloped or vertical edges in the vicinity of the Schottky contact, edges which the interconnect metal must cross without either breaking or forming electrical contact to the exposed semiconductor layers. While isolation by etching (mesa isolation) is often used for experimental devices, step-coverage, resolution, and reliability problems are frequently encountered. If feasible, implant isolation and planar processing are preferred.

While a variety of ions are used for implant isolation, the combination of available implant energy (200 keV) and the required implant depth (1.4 μm) demands hydrogen ions (protons); for fixed implant energy, the projected range of hydrogen is the largest of any atomic species. Implant isolation renders the semiconductor semi-insulating by creation of defect states. The implanted ions, dissipating their energy through collisions with the crystal lattice, create dislocation defects within the crystal. The defects generate electron energy states within the bandgap; if the density of defects with energy near mid-gap is large in comparison with the doping density, the Fermi level will be pinned near mid-band, and the density of mobile electrons in the conduction band will be low [3]. Because band pinning results from defect states with a wide range of energies, the isolated material is not truly semi-insulating; conduction can occur at low frequencies through charging and dis-

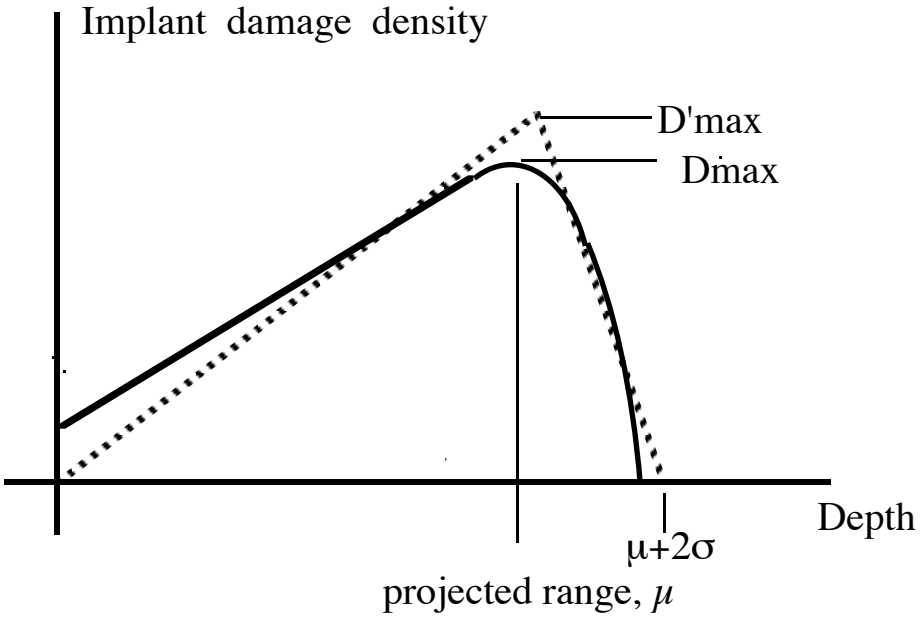


Figure 4.5: Distribution of effective defect density vs. depth (solid line), and piecewise-linear approximation.

charging of the compensating levels [5]. For this reason, an optimum implant dose exists, with low-frequency resistivity decreasing beyond this dose due to conduction through these deep levels. Given some N+ layer doping, both the implant energy and dose thus must be selected.

Proton implant isolation is not used extensively for GaAs processing, and the literature provides few references. Because the defects generated by proton implantation have low activation energy for annealing, proton- isolated circuits exposed to elevated temperatures show degraded reliability; in the fabrication process, isolation must follow ohmic processing due to implant defect annealing at 450 C. Notable references on proton implant isolation include a general discussion in Chapter 5 of Howes and Morgan [3], which draws heavily upon the results of Donnelly and Leonberger [5]. From their data, the optimum density of implanted hydrogen can be determined for dopings in the range of 10^{18} to $10^{19}/cm^3$. Gibbons [6] gives the calculated range statistics of a wide variety of ions in silicon and germanium. While projected ranges for GaAs are similar to the range statistics in Ge, the distribution of crystal damage can differ significantly from the implanted ion density distribution [3].

In a more recent publication by D'Avanzo [7], the damage distribution and the effect on free electron concentration are measured. Through capacitance- volt-

age measurements of layers implanted at densities well below the doping density, D'Avanso determined that the distribution approximates a linear distribution at depths below the damage peak, and approximates a Gaussian distribution at greater depths (Fig. 4.5). The peak of the damage distribution is at a depth μ of 0.65–0.68 μm/100 keV; the depth of the damage peak is ~20% less than the peak of the ion distribution given by Gibbons. Through integration of the distribution, D'Avanso determined that the electrons are trapped in proportion to the implant density, with 3 electrons trapped per implanted proton. In partial confirmation of this data, TRIM simulations (thanks to Robert Anholt, Solid State Lab), give a projected range of $\approx 0.65\mu$m/100 keV, and ≈ 10 defect states per implanted proton. As the defect states have a distribution of energies and thus have varying effectiveness in trapping electrons, D'Avanso's data seems credible.

The data from these references are used to estimate the implant damage distribution and the required damage density (Fig. 4.5). An effective density of defect states equal to three times the implated proton is assumed, and the peak of the damage distribution is assumed to be at a depth μ of 0.65μm/100 keV. D'Avanso does not give the standard deviation σ of the half-gaussian distribution; Gibbons' range data for protons in Ge is used as an estimate. For ease in calculation, the linear-Gaussian distribution is approximated as a linear-linear distribution peaking at μ, and reaching zero at the surface and at a depth $(\mu + 2\sigma)$; the peak density D'_{max} of the fitted damage distribution is chosen to be consistent with 3 trapped electrons per implanted proton. Fitting the data from Donnelly and Leonberger to this distribution gives an optimum defect density in the range of 2–8 times the doping density.

To attain this defect desity over the full 0.8 μm thickness of the N+ layer, implantation at two energies is required. Figure 4.6 compares the N+ layer doping to the effective defect density for a dual implantation of $6(10^{14})$ protons/cm^2 at 160 keV and $1.5(10^{14})$ protons/cm^2 at 110 keV. The surface defect density (Fig. 4.5) will be well in excess of the N- layer doping.

The implant mask must now be designed. Thick layers of photoresist, often used for implant masking, are unsuitable becuase of the high implant energy, the large implant dose, and the long projected range of protons in most materials. All masking materials have a finite nonzero probability of blocking an implanted ion. The probability of passage can be found by reference to tables of the range statistics of the ions within the masking material [6]. To prevent substantial changes in the N- doping profile, the effective ion-induced defect density of masked material should be less than $3(10^{15})/cm^3$. Assuming a half-gaussian distribution and the implant schedule of Fig. 4.6, the mask thickness must be at least $(\mu_{mask} + 4\sigma_{mask})$,

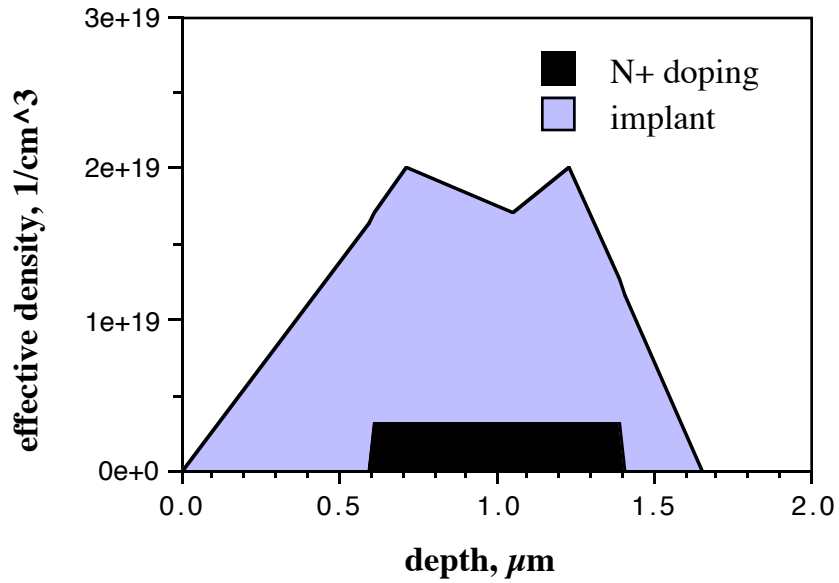


Figure 4.6: N+ layer doping vs. effective defect density for proton implantation at $6(10^{14})/cm^2$,160 keV and $1.5(10^{14})/cm^2$, 110 keV.

where μ_{mask} and σ_{mask} are the projected range and range standard deviation of the protons within the masking material. Thick layers of a dense metal are required– gold was chosen.

If directly placed upon the substrate, the masking metal would adhere to the GaAs; to facilitate mask removal, the Au masking metal is placed above a 1.4 μm layer of polyimide. Range statistics of protons in polyimide have not been published; energy loss in photoresist, which is less dense, is approximately 30 keV/μm [6]. If 40 keV is assumed lost in the polyimide, the gold mask must be at least 1.5 μm thick. A gold masking layer thickness of 1.6μm is selected, at the limit of reliable liftoff with our current procedure. The margin for error in range statistics is small, and thicker (1.7–1.8 μm) masks would be preferable.

The isolation implantation sequence is shown in Fig. 4.7. Polyimide is spun onto the wafer and partially cured by baking; to facilitate subsequent removal, the bake temperature is kept below that required for significant imidization. Photoresist is then applied to the wafer, exposed, soaked in chlorobenzene, and developed, following the liftoff photolithography process schedule. The mask, 1.6 μm Au with a 100$\mathring{A}$ adhesion layer, is then evaporated and lifted off. The wafer is then immersed in commercial polyimide etchant to remove exposed polyimide. The mask, consisting of a Ti/Au film supported on polyimide, is complete and the wafer is now implanted

with hydrogen. Subsequent to implantation, the mask is removed by immesion of the wafer in polyimide thinner or etchant.

4.1.4 Schottky Contacts and Interconnect Metal.

Device fabrication is completed by deposition of interconnect metal and formation of Schottky contacts. While these two functions are frequently performed separately, a single deposition of metal suitable for Schottky contacts will also serve for interconnections if the metal thickness is sufficient for low resistivity. Schottky contacts are formed only in those areas in which the deposited metal lies on unimplanted GaAs. Most metals deposited on GaAs form a Schottky contact; selection of the metal system is guided by considerations of stable stochiometry at elevated temperatures, adhesion, and compatibility of thermal expansion coefficients [3]. Despite some difficulties with thermal stability, the Ti/Pt/Au system is the predominant metalization for Schottky contacts and MESFET gates in commercial GaAs circuits, and so was selected. Due to chemical changes in the metal-semiconductor interface at elevated temperatures, Schottky contacts must be deposited after both ohmic processing and isolation implantation. Metal deposition follows the liftoff photolithography procedure; in evaporation, 0.1 μm of titanium forms the Schottky contact, 750 $\mathring{A}$ of platinum serves as a diffusion barrier, and 1.4 μm of gold reduces the sheet resistance of the interconnections. The photoresist is dissolved in acetone, removing the excess metal, and the nonlinear transmission line is complete. Photomicrographs of the device are shown in Fig. 4.8.

4.2 Evaluation

Device evaluation consists of process verification measurements to determine the success of various process steps, measurements of the individual circuit elements to determine the variation from the design values, and functional testing of the full circuit. Because few test structures were designed into the mask set, several circuit element characteristics (e.g. diode resistance and capacitance) were determined through measurements of the small-signal microwave propagation characteristics of the full nonlinear transmission line. Because the anticipated transition times were shorter than the risetime of both sampling oscilloscopes, microwave connectors, and microwave wafer probes, wavefront compression (functional testing) was measured by direct electrooptic sampling (Chapter 5), a nearly noninvasive internal-node probing technique providing $\sim$100 GHz measurement bandwidth.

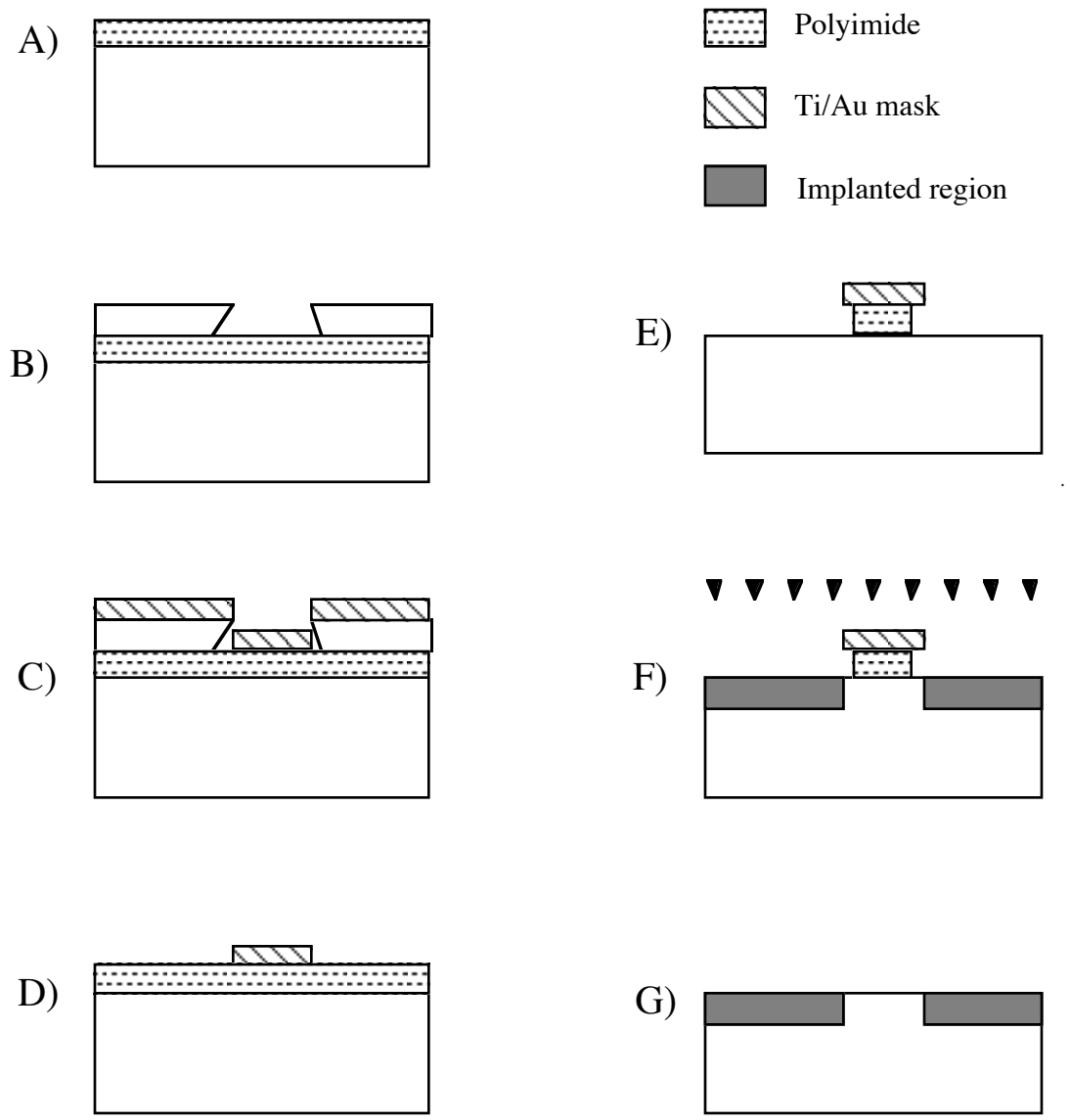


Figure 4.7: Isolation implantation process sequence. Polyimide is applied to the wafer, a), and cured. Liftoff photolithography is then used to pattern a 100$\mathring{A}$ Ti/1.6 μm Au mask, b), c), and d). Excess polyimide is removed by wet chemical etching, e). The wafer is then implanted with hydrogen, f), and the mask is removed by immesion in polyimide thinner, g).

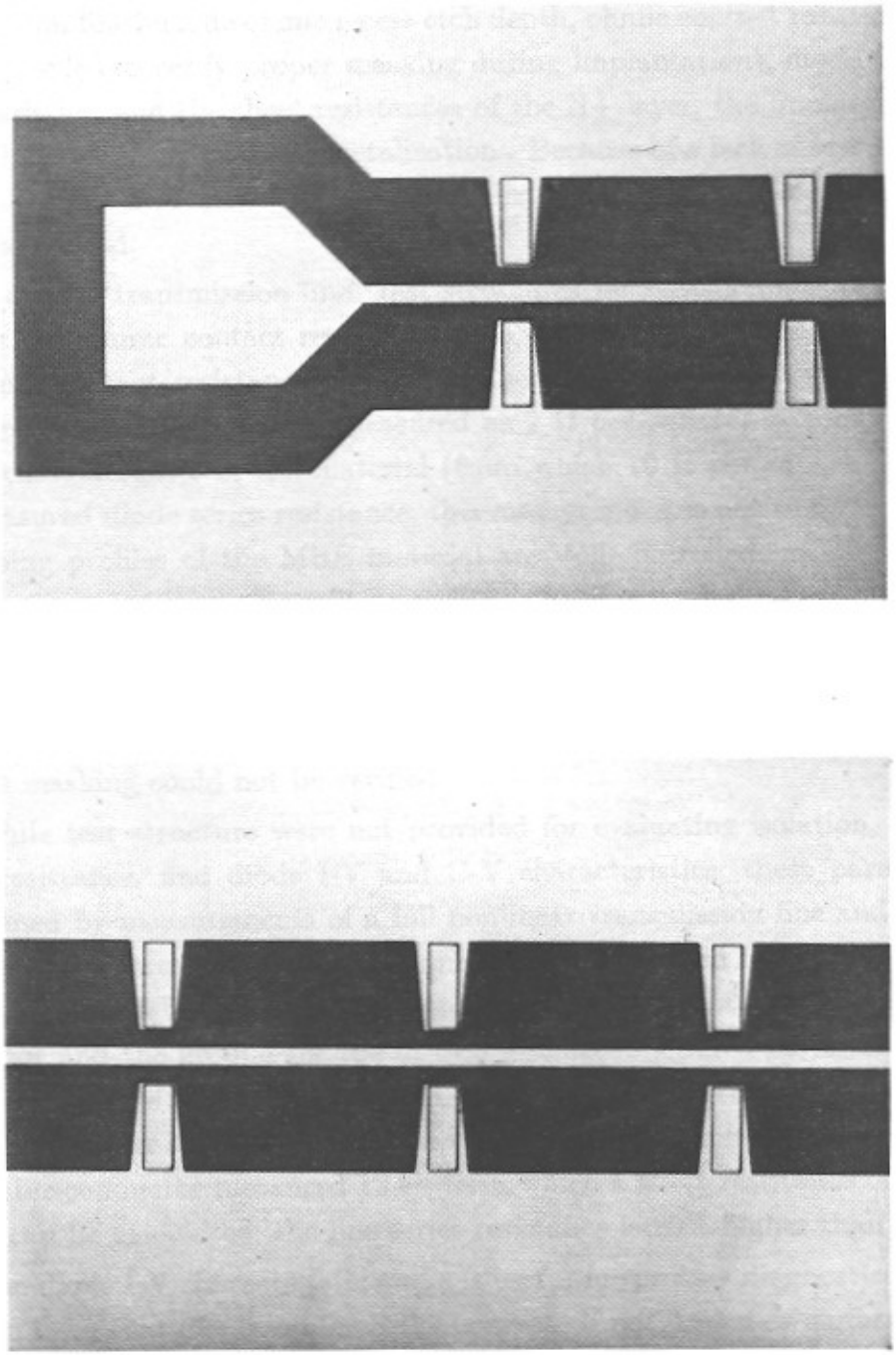


Figure 4.8: Photomicrographs of the nonlinear transmission line. The ohmic etch wells are visible as ridges in the interconnect metalization . The (100 μm$\times$100 μm) bond pads at the ends of the line interface to 40 GHz microwave wafer probes.

4.2.1 Process Diagnostics

Process parameters which should be verified during fabrication of the nonlinear transmission line include ohmic recess etch depth, ohmic contact resistance, N- layer doping profile (to verify proper masking during implantation), diode C-V and I-V characteristics, and the sheet resistances of the N+ layer, the implanted (isolated) material, and the interconnect metalization . Because of a lack of test structures on the mask, some measurements were made indirectly, while other parameters could not be measured.

So-called "transmission line" test structures for measurement of N+ layer resistivity and ohmic contact resistance were present on the mask; from these, the 0.03Ω-mm contact resistance was determined. Through these test patterns, the N+ layer sheet resistance was measured as 7 Ω per square, in conflict both with known characteristics of the material (from which 10 Ω per square is calculated) and measured diode series resistance; this measurement is not to be trusted. While the doping profiles of the MBE material are well-controlled, insufficient masking during proton implantation will degrade the carrier concentration at the surface, decreasing the change in capacitance with voltage. Doping profiles of the N- layer can be readily determined through capacitance- voltage measurements on a large-area Schottky contact, but no such contact was designed into the mask, and proper implant masking could not be verified.

While test structure were not provided for evaluating isolation, interconnect metal resistance, and diode I-V and C-V characteristics, these parameters were determined by measurements of a full nonlinear transmission line and a test wafer in which the interconnect metalization layer was deposited on a blanket-implanted (e.g. implant isolated over the entire area) wafer. The resistance between the signal conductor and the ground for the line on blanket-implanted material was 180 kΩ; from knowledge of the line geometry, the implanted sheet resistance was calculated as over 20 MΩ per square. The series resistance of the 6.7mm × 10μm transmission-line center-conductor measured 12 Ω, from which a sheet resistance of 0.018 Ω per square can be calculated; the line series resistance is 20% higher than calculated.

The diode I-V characteristic curve is useful for process diagnostics, and should conform from that of a good Schottky contact. If not, metal or surface contamination should be suspected. Measured reverse-bias current for the 42 parallel diodes was 10 μA at -3 volts, increasing to 30 μA at -10 volts. This current includes the resistive current through the implant-isolated material. The I-V characteristics of the blanket- implanted line show 15 μA leakage at -3 volts and 50 μA leakage at -10 volts, and so the reverse I-V characteristics of the diodes themselves could not be determined. As the diodes incorporate no guard ring or other edge passivation,

leakage at -10 volts may also arise from soft breakdown in the high-field regions at the contact edges [1]. Drawn as a log- linear plot, the forward-bias I-V characteristics reveal generation- recombination current [8] at low forward bias ($V < 0.3$ volts), with a slope of approximately $q/2kT$, and thermionic injection for bias above 350 mV, with a slope of approximately q/kT. Current for the 42 parallel diodes is 1 mA at 550 mV forward bias.

Subtracting the C-V curves of the nonlinear transmission line and the blanket-implanted line, and dividing by the number of diodes (42) determines the diode C-V characteristics. The capacitance of the blanket- implanted line includes the line capacitance, the bond pad capacitance, and the fin capacitance, but does not include stray capacitance from the signal conductor to the segments of buried N+ layer passing under the line. The diode C-V characteristics determined from the two measurements conformed to that of a 14 fF constant capacitance in parallel with a step- junction with $C_{j0} = 50$ fF and $\phi = 0.8$ volts. The net C-V characteristics could be the result of either a stray shunt capacitance or a nonuniform doping profile due to inadequate implant masking, but the cause cannot be determined without a more comprehensive set of test structures.

4.2.2 Circuit Characterization by Microwave Measurements.

Process diagnostic measurement provided diode capacitance and the line series resistance, but the interconnect impedance and electrical delay, and the diode series resistance were still unknown. Before testing the nonlinear transmission line for compression, these circuit parameters needed to be verified. Further, wavefront compression occurs on the line as a result of a variation in group delay with voltage, and sucessful interfacing of the device to sources is determined by the circuit's voltage-dependent characteristic impedance. Success in wavefront compression requires appropriate values of these two fundamental parameters. The interconnect impedance and delay, the diode series resistance, and the nonlinear transmission line delay and impedance were all determined through measurements of two-port microwave scattering parameters.

The ends of the nonlinear transmission line (Fig. 4.8) flare to the dimensions of bond pads compatible with Cascade microwave wafer probes. These probes exhibit repeatable microwave connections to frequencies from DC to approaching 50 GHz [9,10], and are equipped with on-wafer calibration standards which permit calibration of the probes to a reference plane located at the probe tips. Combined with an automatic microwave network analyzer (Hewlett-Packard model 8510), the probes permit calibrated measurements of the small-signal 2-port parameters of the

nonlinear transmission line.

From two-port measurements of the lines on blanket-implanted material, the effective interconnect impedance Z_1' and effective interconnect delay τ' can be determined. Recall from Chapter 3 that these effective parameters include perturbations to the impedance and delay arising from the fin capacitance. Line impedance Z_1', determined through measurement of S_{11}, was approximately 80 Ω, while the effective interconnect delay τ', determined from measurement of S_{21}, was approximately 1.4 to 1.6 picoseconds. The interconnect attenuation ($|S_{21}|$), which was 1 dB at low frequencies (due to the 12Ω line series resistance), showed resonances due to the mismatched impedance, and (neglecting the resonances) increased to 2 dB at 26 GHz—the increased loss is due to skin resistance.

The diode series resistance was determined through measurements of the small-signal attenuation ($|S_{21}|$) of the nonlinear transmission line (Fig. 4.9). The expression for attenuation given on Fig. 4.9 is correct for frequencies well below both the diode and periodic-line cutoff frequencies. As the capacitance, line series resistance, and impedance were known from other measurements, the diode series resistance, 9.2 to 10.5 Ω, was found by curve-fitting to the data of Fig. 4.9. The measured series resistance is in good agreement with the calculated 9.6 Ω. For determination of the diode series resistance, the line attenuation measurement is superior to an impedance measurement on a single diode within a test structure: the impedance of a single diode does not differ measurably from its capacitive reactance for frequencies below the 300 GHz cutoff frequency.

The small-signal wave impedance $Z_0(V)$ was determined through measurement of S_{11}. The line impedance, varying from 39 Ω at zero volts bias to 50 Ω at -3 volts bias, is 12% lower than the designed impedance due to the the variations in effective interconnect impedance Z_1' and diode zero-bias capacitance. The large-signal impedance, designed to be 50 Ω for a zero to -2 volt signal, will approach 50 Ω for input signals switching from zero to approximately -5 volts.

Finally, the group delay ($-d\theta/d\omega$ of S_{21}) was measured (Fig. 4.10). The spikes are artifacts in the measurement: to keep within the linear region, the network analyzer test signals were kept below -20 dBm, and the resulting impulsive measurement noise in the calibration sequence was smoothed by the analyzer video filter. This led to rectangular perturbations in the delay data. Within the noise limits of the measurement, the group delay is independent of frequency, and varies by 20 ps as the bias is varied between zero and -3 volts. Over this range of voltages, the change in delay is 30% smaller than the 28 ps design value. From this data, it is expected that the input falltime must be less than 20 ps if the line output is to reach the minimum compressed falltime. New circuit simulations were performed using

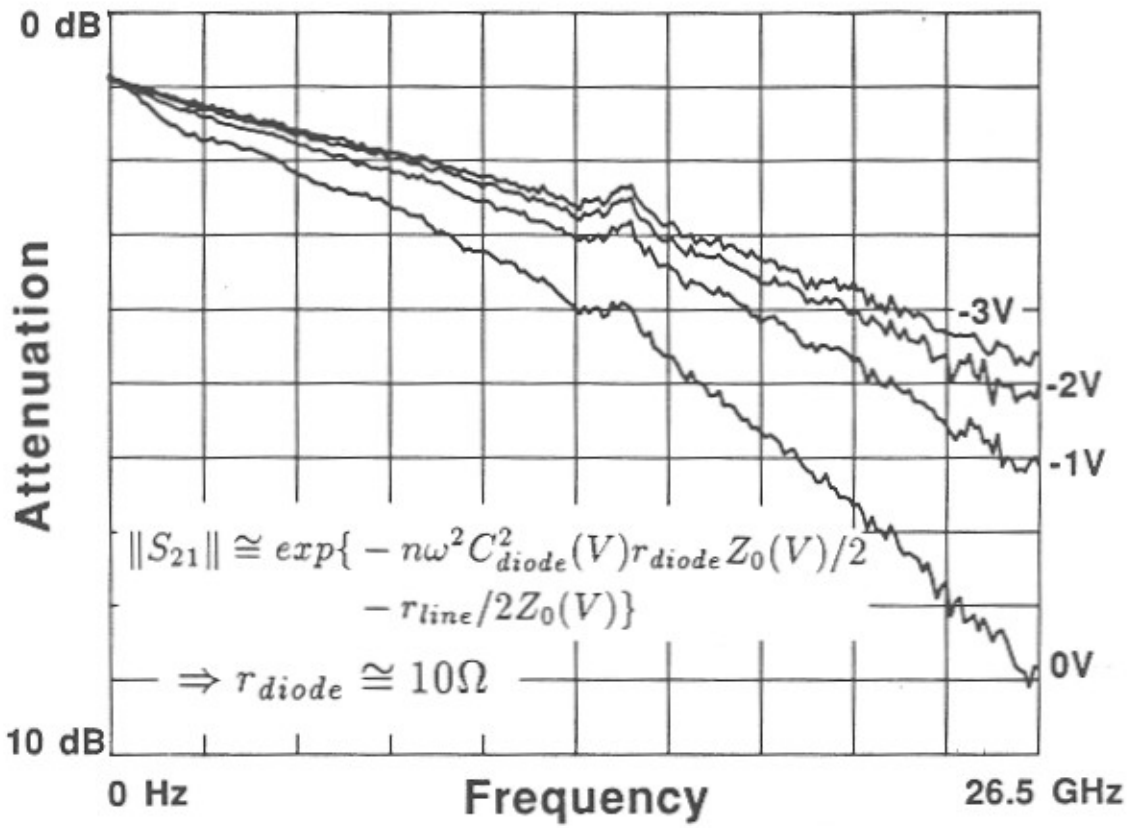


Figure 4.9: Small-signal attenuation vs. frequency and bias voltage. The high loss is evidence of high diode series resistance; small- signal loaded line loss is a good test method for determining the diode series resistance.

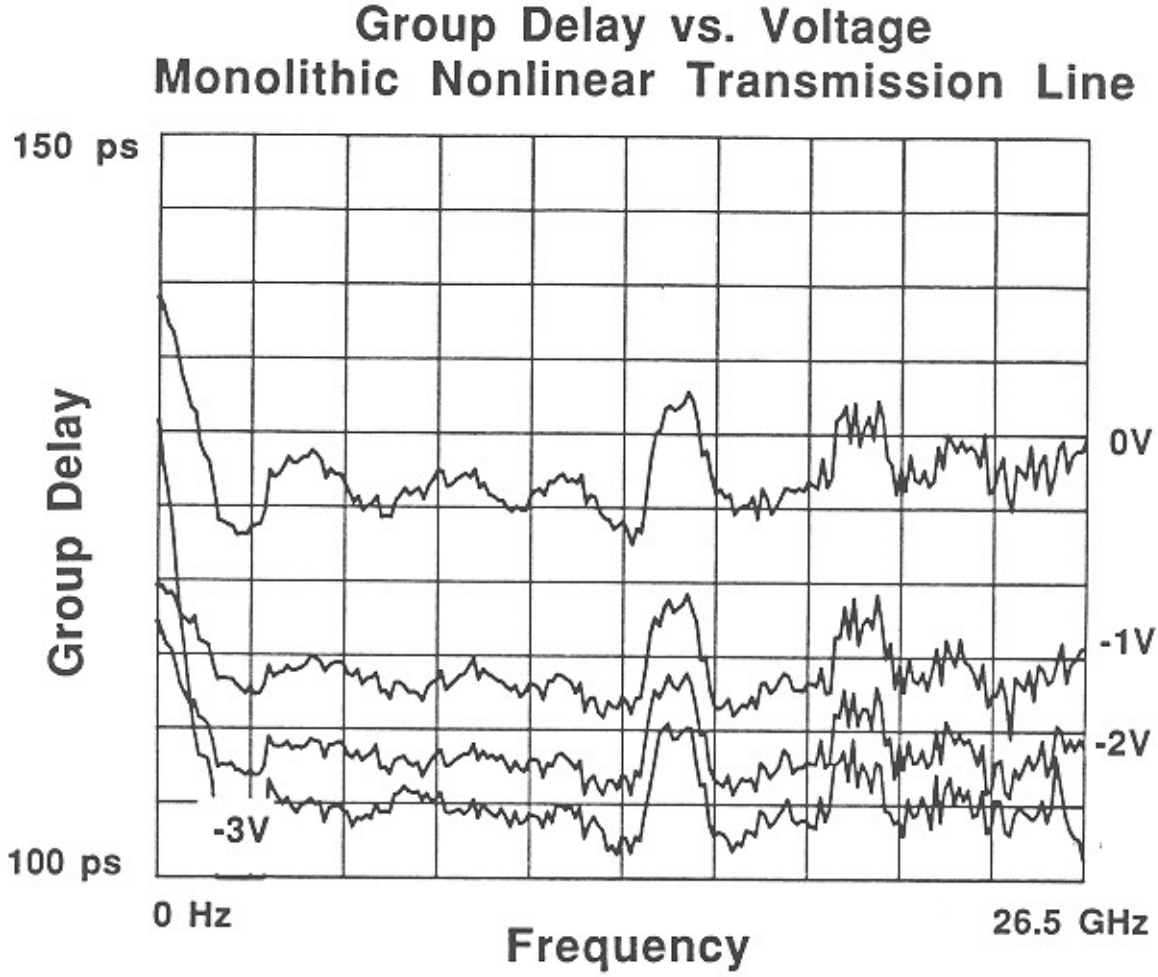


Figure 4.10: Small-signal group delay vs. frequency and bias voltage. The delay is relatively constant with frequency, and varies $\sim$ 25 ps as the bias voltage is changed from 0 to -3 volts.

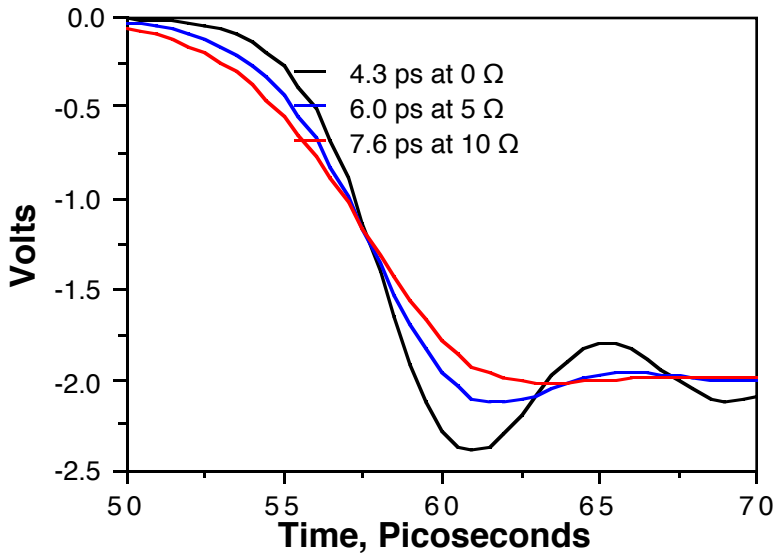


Figure 4.11: Simulations of the minimum compressed falltime with the measured circuit parameters and a zero to -2 volt step- function input. The measured diode series resistance is 10Ω.

the measured circuit element values; the results of these simulations are shown in Fig. 4.11. An output falltime of approximately 7.5 picoseconds is expected. Also shown are simulations for decreased diode series resistance.

4.2.3 Measurements of Falltime Compression.

In functional testing of the nonlinear transmission line, the line should be driven with an step-function input varying from zero to a negative voltage in the range of -2 to -6 volts (-6 volts is the voltage required to fully deplete the N- layer, and hence the maximum signal voltage for which the falltime will be compressed), with larger signal voltages yielding both larger compression ratios and smaller output falltimes.

With the current design, even a cascade of 2 nonlinear transmission lines requires an input falltime of less than 40 ps for compression to the minimum compressed falltime. While we had commercial step-recovery- diodes with 40 ps transition times, the diodes had not been fixtured at the time of these experiments. Instead, the lines were tested by using sinusiodal inputs with a superimposed negative DC bias. The formation of shock wavefronts with sinusiodal excitation is similar to that occuring with excitation with a step-function. For the case where

$\omega_{diode} \ll \omega_{per}$, the analyses of Khochlov [11] and Peng and Landauer [12] show that the same final wavefront transition time is reached with step-function and sinusiodal inputs, *but only for the same wavefront magnitude*. For the case where ω_{diode} approaches or exceeds ω_{per}, no analytical solutions exist. Appeal must then be made to the experimental results with the scale model nonlinear transmission line (Figs. 2.7 and 2.8), where identical wavefront transition times were attained for step-function and sinusiodal excitation.

In the scale model, the diode series resistance (1Ω) was negligible. The 10 Ω diode series resistance of the prototype monolithic devices results in substantially different behaviour for sine waves and step-functions propagating on the nonlinear transmission line. For step- function excitation, both elementary DC circuit arguments and the analysis of Section 1.2 show that the diode series resistance contributes no attenuation. The sole influence of the diode resistance is in determining the final profile of the shock wavefront. In marked contrast to this, with large sinusoids propagating on the line, the diode series resistance introduces attenuation similar to the small-signal attenuation shown in Fig. 4.9. As a sinusiod propagates along the line, its magnitude decreases. The input signal may swing from zero volts to the negative voltage V_l, but the output voltage swing will be smaller and the peak positive voltage will be less than zero volts. For maximum compression and minimum falltime, the peak positive voltage must be near zero volts, where capacitance and delay vary most strongly with voltage. Both the reduced signal magnitude and the reduced peak positive voltage reduce the change in delay ΔT across the wavefront, and increase the wavefront transition time.

The anticipated output falltime, 7.5 ps, is below the time resolution of commercial sampling oscilloscopes. Even 40 GHz microwave wafer probes and 40 GHz microwave connectors will severly degrade the falltime, so the formation of shock wavefronts was measured within the line itself, using electrooptic sampling. Electrooptic sampling is discussed in more detail in Chapter 5: the technique permits measurements of the voltages within GaAs integrated circuits. The spatial resolution of the instrument is approximately 3 μm, permitting probing directly within the $10 \times 10\mu$m diode area to measure the voltage across the diode depletion region. The time resolution of the instrument is approximately 2–3 picoseconds, sufficient for accurate measurement of the falltimes generated on these first-generation nonlinear transmission lines. Finally, the instrument is nearly noninvasive; the sole circuit perturbation is a 17 μA photocurrent induced by absorbtion of the probe beam; this photocurrent is three to four orders of magnitude below the signal currents.

The circuits were driven by sine waves from a microwave synthesizer and probed using an electrooptic sampling system. To maximize the signal compression and to

bring the peak signal excursion at the line output close to zero volts, the line was biased so that the input diodes were at the onset of forward conduction at the peak positive excursion of the input waveform. It was observed that the falltime degraded as the signal approached the output bond pad, presumably as a result of parasitics from the bond pad and the microwave wafer probes; waveforms were measured at the third diode from the line output.

Figure 4.12 shows the waveform at the 39^{th} diode of the transmission line, with sinusiodal excitation at 15 GHz. The input falltime, i.e. the time between the 90% and 10% points on the sinewave, was 20 ps. The output 90% –10% falltime was 7.8 ps. The input signal magnitude was approximately 20 dBm (6 volts peak–peak), although some undetermined losses from cables, a bias network, and a microwave wafer probe were present between the 20 dBm source and driven circuit. The output falltime varied significantly with the bias voltage, changing from 7.8 to 9.5 ps as the bias voltage was decreased 300 mV. To attain a greater compression ratio, two lines were cascaded. Due to the significant line attenuation, each line was seperately biased, with each line biased to bring the input diodes to the onset of forward conduction at the peak positive excursion of the waveform. Compression of an 8 GHz input signal (37 ps falltime) to 10.1 ps output falltime was attained (Fig. 4.13). The results are in rough agreement with the SPICE simulations of Fig. 4.11. Better results could be attained with a step-function excitation, and with improved devices with reduced series resistance.

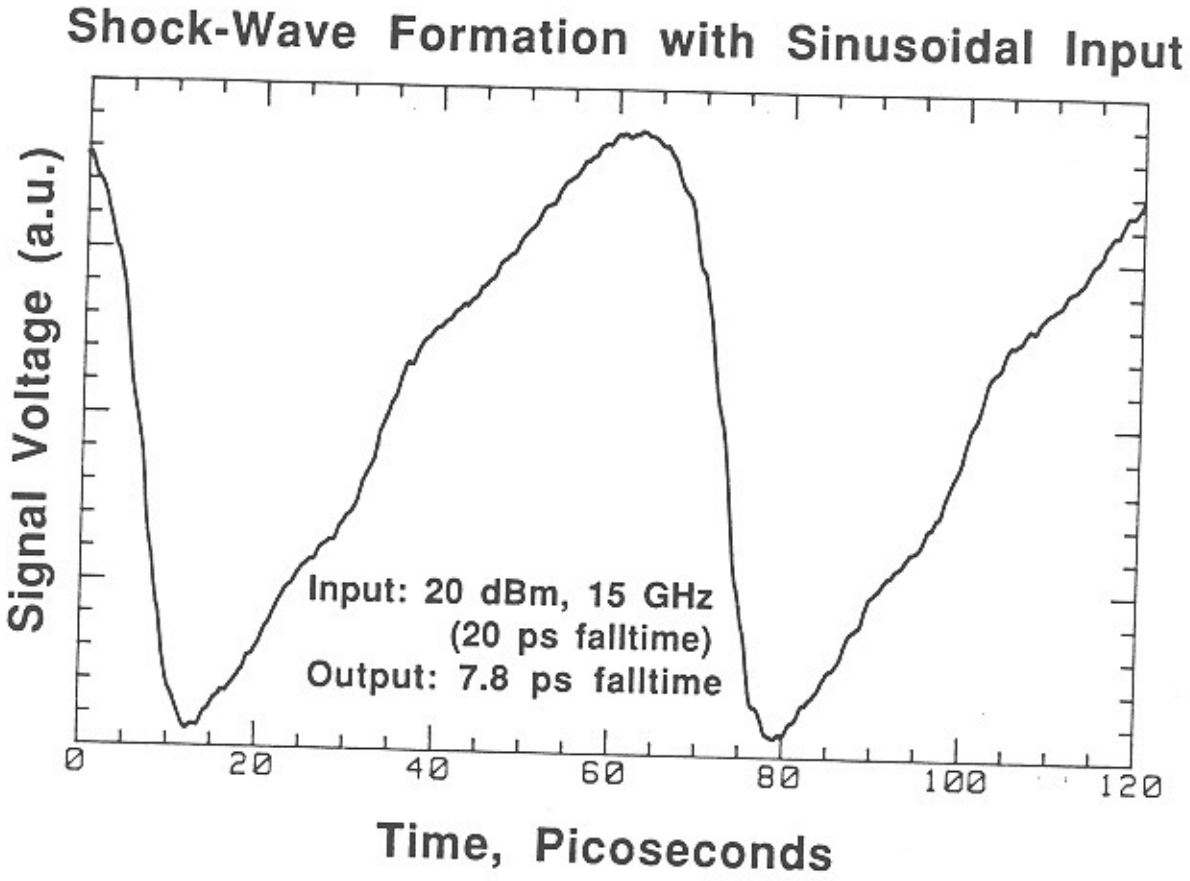


Figure 4.12: Shock-wave formation with sinusoidal input.

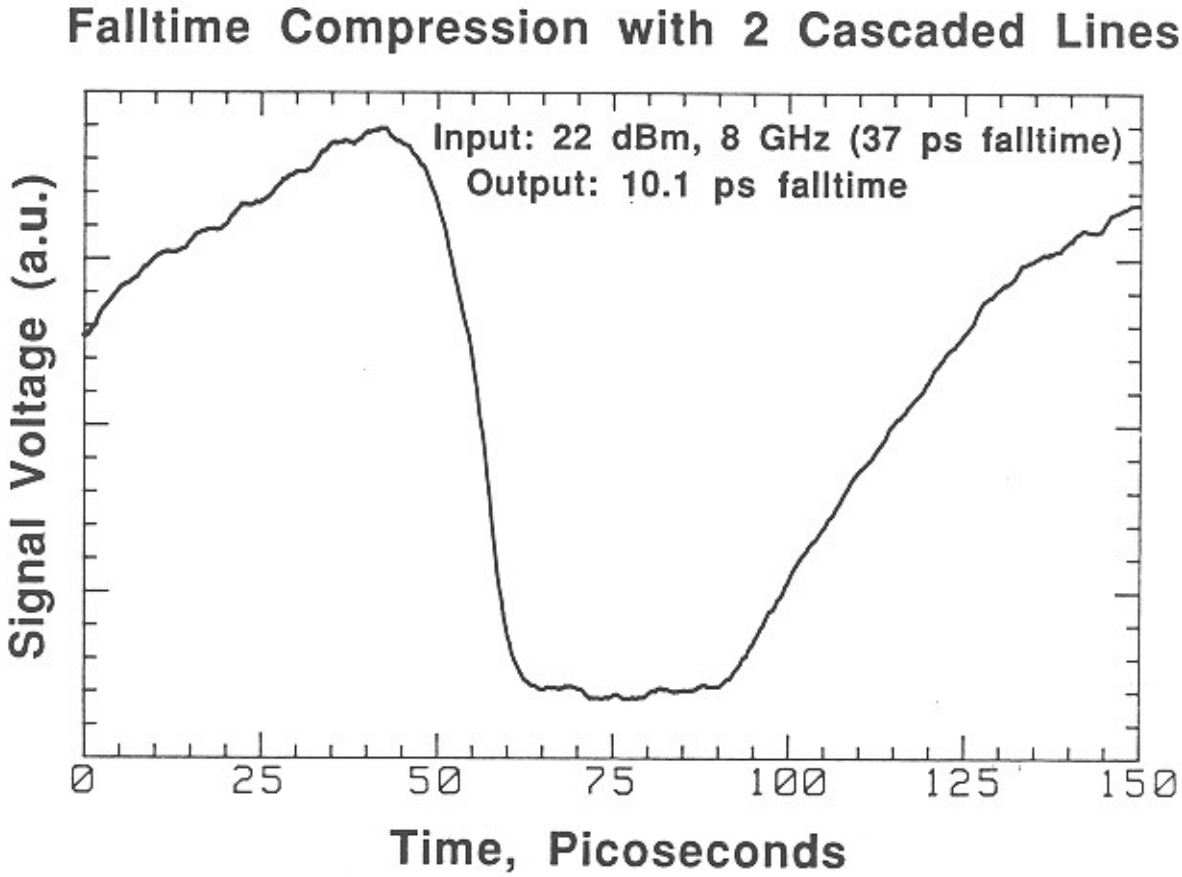


Figure 4.13: Falltime compression with two cascaded lines.

Appendix: Abbreviated Process Sequence

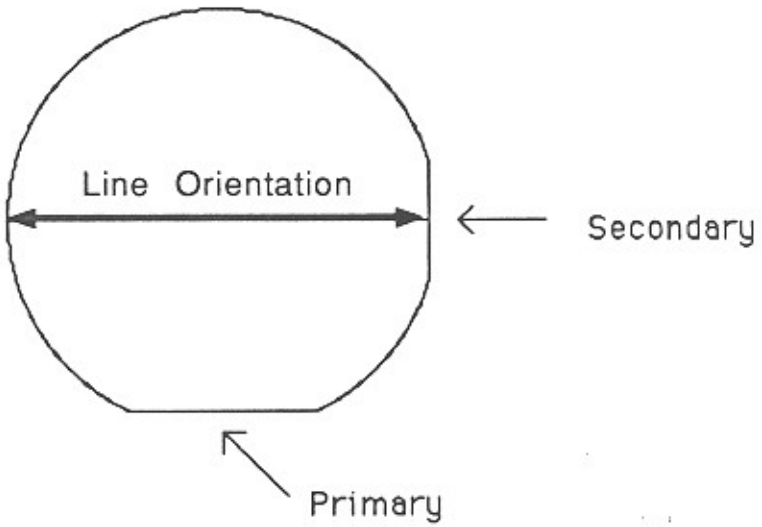
Etch times and ion implant energies given are for the standard material for the nonlinear transmission line, ie a .6 μm N- layer (3X10^16) above a .8 μm N + layer (3X10^18) on a [100] semi-insulating GaAs substrate.

1: Self-Aligned Ohmic Contacts

Clean wafer

Wafer Orientation

By american standard, the wafer is right side up when the primary flat is towards you and the secondary flat is to your right.



Photoresist

Use AZ 5214 and follow liftoff photoresist procedure
but postbake 90°C, 30 min, and USE HMDS adhesion promoter
After postbake, use profilometer to find resist thickness
Verify resist thickness before etching!

Wet Etch

Use NH40H:H2O2:H2O (14 : 2.4 : 200)
Etch in H2O bath for stable temperature.
Etch rate estimated at 56 A/sec (1.15 microns in 3.8 min)
Etch first piece for 1 min to establish etch rate
Stop etch with 2 min DI H2O rinse, Nitrogen dry
Monitor Etch depth with surface profilometer
Perform additional wet etches to 0.75 μm depth

Evaporate Metal

Etch wafer for 60 seconds in 6:1 BOE immediatly prior to evap.
Evaporate ohmic metal
use either Au/Ge/Au... multilayers

108 A Ge
102 A Au
63 A Ge
236 A Au
100 A Ni
2000 A Au

or eutectic Au/Ge mixture

500 A (88% Au-12% Ge)
100 A Ni
2000 A Au

Liftoff metal
3-solvent clean, DI rinse, N2 dry
Alloy metal at 450°C for 12 seconds in Rapid Thermal Annealer
Measure Contact resistance

2: Proton Implant

Hot Solvent Clean
Apply Ciba-Geigy polyimide
Spin adhesion promoter at 5000 rpm for 30 sec
Spin Probimide 286 at 5000 rpm for 30 sec
Bake polyimide
Bake at 100°C for 30 min
Ramp to 180°C, Hold for 15 min at 180 +/- 5°C
Ramp to 240°C, Hold for 15 min at 240+/- 5°C
Let oven cool down to 150°C before removing wafers
Photoresist
Use liftoff photolithography procedures
Increase exposure time by about 10% to compensate for the low reflectivity of polyimide, and omit TCA in wafer cleaning
Evaporate Metal

100A Ti
1.6um Au

Liftoff Metal
Etch Polyimide
Etch in polyimide Etcher, QZ 3296, for 150 sec
Rinse in Etch Rinse, QZ 3297, for 30 sec
Rinse in DI H2O for at least 30 sec
N2 dry and inspect
Use gentle ultrasonic to remove polyimide residues
Implant Hydrogen
6E14/cm^2 at 160 keV
1.5E14/cm^2 at 110 keV
Strip polyimide
Soak in hot polyimide thinner to remove gold
Use ultrasonic bath as necessary to aid in liftoff
Ash polyimide residue in oxygen plasma
Clean wafer

3: Schottky and Interconnect Metal

Photoresist
Use liftoff photolithography procedures, but...
...increase exposure ≈20% to fully expose ohmic etch holes
... and develop in the dark.
Evaporate
Etch wafer in BOE for 60 secs prior to evaporation
Evaporate

1000 A Ti
750 A Pt
1.4 µm Au

Liftoff metal

I: Wafer Cleaning

2 minutes in boiling TCA (Trichloroethane)
2 minutes in hot Acetone
2 minutes in boiling Isopropyl
DI water rinse
N2 dry
dehydration bake 120°C 30 minutes
Let wafer cool 5-10 minutes

II: Liftoff Photoresist (Metallization)

Solvent Clean Wafer
    Trichlorethane
    Acetone
    Isopropyl
    Rinse in cool DI water
    Nitrogen blow dry
Spin AZ 5214e photoresist for 40 seconds at 3000 rpm
Soft Bake in convection oven for 25 minutes at 75 °C
        Bake in an open petri dish on the top shelf of a Blue-M oven.
Align and expose
        13 Seconds @ 14.8 mW/cm^2 only if through borosilicate mask. Silica Masks should give times around 7 seconds.

Chlorobenzene soak
    Soak 15 minutes in undiluted chlorobenzene
    Nitrogen dry
Develop
    Develop 1.5 to 1.75 minutes in 1:1 AZ solution developer:H2O
            Watch the development. When large (visible) areas on the wafer have cleared, transfer to the "clean" beaker for the final 15 seconds.
    two-beaker method: last 15 seconds in "clean" beaker of developer
    Rinse in DI water for 2 minutes
    Nitrogen dry
Pre-evaporation clean
        Just before placing wafers in evaporator, do a 60 second buffered oxide etch in a teflon beaker.
Evaporate desired metal(s)
    up to 1.6 micron thickness
Lift Off
    Soak wafers for at least 30 min in cool acetone
    spray with acetone to remove metal
    If absolutely necessary, Ultrasonic clean
    Acetone, isopropyl, and DI water rinse
    N2 dry wafers
    Inspect in optical microscope

References.

[1] S.M. Sze: *Physics of Semiconductor Devices*, Wiley- Interscience,1981.

[2] R.E. Williams, *GaAs Processing Techniques*, Artech House, 1985.

[3] M.J. Howes and D.V. Morgan, editors, *Gallium Arsenide: Materials, Devices, and Circuits*, Wiley-Interscience, 1985.

[4] M. Hatzakis, B.J. Canavello, and J.M. Shaw, "Single-step optical lift-off process", IBM J. Res. Dev., 24(4), 452-460, July 1980.

[5] J.P. Donnelly and F.J. Leonberger, "Multiple energy proton borbardment in n+ GaAs", Solid State Electronics, 20, 183-189 (1977).

[6] J.F. Gibbons, W.S. Johnson, and S.W. Mylroie, *Projected Range Statistics: Semiconductors and Related Materials*, Dowden, Hutchinson, and Ross, Stroudsburg, PA, 1975.

[7] D.C. D'Avanzo, "Proton isolation for GaAs integrated circuits", IEEE Trans. Electron Devices, ED-29, 1051-1059, 1982.

[8] E.S. Yang: *Fundamentals of Semiconductor Devices*, McGraw- Hill, 1978.

[9] E.W. Strid, K.R. Gleason, and T.M. Reeder, "On-wafer measurement of gigahertz integrated circuits", *VLSI Electronics: Microstructure Science, Vol 11*, New York, Academic Press, 1985.

[10] Cascade Microtech, Inc., P.O. Box 2015, Beaverton, OR 97075

[11] Khokhlov, R.V. :"On the Theory of Shock Radio Waves in Non-Linear Lines", *Radiotekhnika i elektronica*, 1961, 6, No.6, pp. 917-925.

[12] S.T. Peng and R. Landauer: "Effects of Dispersion on Steady State Electromagnetic Shock Profiles", IBM Journal of Research and Development, vol. 17, no. 4, July 1973.

.

Part 2:
Picosecond
Optoelectronic Instrumentation

Chapter 5: Electrooptic Sampling

The development of advanced GaAs devices and integrated circuits has been spurred by a number of applications, including microwave and millimeter-wave radar and communication systems, fiber optic digital data transmission at gigahertz rates, high-speed data acquisition, and the constant push for faster digital logic in high-speed computers and signal processors; the IC's developed for these applications are creating new demands upon high-speed electronic instrumentation.

One demand is for increased instrument bandwidth. GaAs MESFET's have been demonstrated with maximum frequency of oscillation, f_{max}, in excess of 110 GHz [1], while pseudomorphic InGaAs/AlGaAs modulation-doped field-effect transistors [2] have shown power- gain bandwidth products which extrapolate to give $f_{max} \sim 200$ GHz, resonant tunnelling diodes have exhibited oscillation at 56 GHz [3], and heterojunction bipolar transistors are expected to show similar performance. Because the maximum frequency of oscillation of these devices is often greater than the 100 GHz bandwidth of commercial millimeter-wave network analyzers, f_{max} is estimated by extrapolation from measurements at lower frequencies. Used as switching elements, propagation delays and transition times of 1-10 ps should be expected for these devices, times well below the resolution of commercial sampling oscilloscopes. In either case the device bandwidth exceeds that of the measurement instrument.

A second demand is for noninvasive access to the internal signals within high-speed integrated circuits. GaAs digital integrated circuits of MSI (medium-scale integration) complexity and 1-5 GHz clock rates are now available commercially, as are GaAs monolithic microwave integrated circuits (MMIC's) of SSI (small-scale integration) complexity and 1-26 GHz bandwidths. More complex LSI (large-scale integration) digital circuits are under development, and experimental SSI digital circuits operating with 18GHz clock rates [4] have been demonstrated. In contrast to silicon LSI integrated circuits operating at clock rates in the tens and hundreds of megahertz, the development of GaAs high-speed circuits is hampered both by poorly refined device models and by layout-dependent circuit parasitics associated with the high frequencies of operation. A test instrument providing noninvasive measurements within the integrated circuit would permit full characterization of complex high-speed IC's.

Conventional electrical test methods are limited by both the instrument bandwidth and by the bandwidth, invasiveness, and spatial resolution of the probes used

to connect circuit to instrument. Sampling oscilloscopes, used for time-domain measurements, have risetimes of 20- 25 ps, while network analyzers, used for small-signal frequency-domain 2-port characterization, are available for microwave waveguide bands as high as 60-90 GHz. The utility of network analyzers at the higher microwave bands is impaired by difficulties in characterization of the network parameters of the fixture connecting the tested device to the instrument's waveguide ports. Microwave wafer probes use tapered sections of coplanar waveguide transmission line to connect the instrument to small circuit bond pads; the low characteristic impedance of such probes (typically 50 Ω) limits their use to input/output connections. High-impedance probes suitable for probing intermediate circuit nodes have significant parasitic impedances at microwave frequencies, severely perturbing the circuit operation and affecting the measurement accuracy. In both cases, the probe size is large compared to IC interconnect size, limiting their use to test points the size of bond pads.

This chapter reviews direct electrooptic sampling, a measurement technique developed in Ginzton Lab which allows for internal-node voltage measurements in GaAs IC's with picosecond time resolution, corresponding to bandwidths in excess of 100 GHz. Electrooptic sampling using external electrooptic elements was first proposed by Gunn [5], and refined by Valdmanis and Mourou [6,7] . Direct electrooptic sampling, i.e. probing directly within the substrate of a GaAs circuit, was first developed in our group by Brian Kolner [8,9]. More recently, Kurt Weingarten [10] extended the technique to full characterization of GaAs integrated circuits, while Jim Freeman and Scott Diamond developed the back-side probing technique [11].

Here, work with the electrooptic sampling system has focused on laser timing stabilization [12], system noise analysis and characterization, and demonstration of the system's capabilities by characterization of GaAs integrated circuits [13,14], primarily monolithic travelling-wave amplifiers. This work cannot be presented independent of a description of the principles and limitations of direct electrooptic sampling, and the material presented will thus include the many contributions of Kurt, Brian, Jim, and Scott. Laser timing stabilization is a more self-contained topic, and will be discussed separately in Chapter 6.

5.1 Electrooptic Voltage Probing in a GaAs Crystal

The electrooptic effect is an anisotropic variation in a material's dielectric constant (and hence index of refraction) occurring in proportion to an applied electric

field. The effect, whose origin lies in small quadratic terms in the relationship between an applied field and the resulting material polarization, is present in a variety of non-centrosymmetric crystals. Among these are GaAs, InP and AlAs, used for high-speed semiconductor devices, and lithium niobate ($LiNbO_3$), lithium tantalate ($LiTaO_3$) and potassium dihydrogen phosphate (KH_2PO_4), used for nonlinear optical devices. Centrosymmetric crystals do not exhibit the electrooptic effect; notable among these materials are silicon and germanium.

The change in refractive index of these materials with electric field can be used for optical phase- modulation, and, from this, polarization-modulation or intensity-modulation [15,16]. Lithium niobate and lithium tantalate electrooptic modulators are used in both lasers, and fiber-optic systems; direct electrooptic sampling uses the electrooptic effect in GaAs to obtain voltage-dependent intensity modulation of a probe beam.

Gallium arsenide is an electrooptic material. Through the electrooptic effect, the electric fields associated with conductor voltages within a GaAs circuit induce small anisotropic changes in the optical index of refraction of the substrate [17]. With the use of a suitably oriented and polarized probing beam passing through the circuit's substrate in the vicinity of a conductor, the polarization and subsequently the intensity of the probe beam will be modulated in proportion to the conductor's potential. Picosecond time resolution is achieved by using a pulsed optical probe beam with pulse durations on the order of one picosecond, permitting instrument bandwidths greater than 100 GHz.

The principal axes of a GaAs IC fabricated on standard (100)-cut material are shown in Fig. 5.1. The **X**, **Y**, and **Z** axes are aligned with the $\langle 100 \rangle$ directions of the GaAs cubic Bravais lattice, while the $\mathbf{Y}'$ and $\mathbf{Z}'$ axes are aligned with the $[01\bar{1}]$ and [011] directions, parallel to the cleave planes along which a GaAs wafer is scribed into individual IC's. Because the $[01\bar{1}]$ and [011] directions, parallel to the IC edges, are also the eigenvectors of the electrooptic effect, the **X**, $\mathbf{Y}'$ and $\mathbf{Z}'$ axes are the natural coordinate system for describing electrooptic sampling in GaAs.

In GaAs, the difference in the refractive indices $n_{y'}$ and $n_{z'}$ in the $\mathbf{Y}'$ and $\mathbf{Z}'$ directions is [18]

$$n_{z'} - n_{y'} = n_0^3 r_{41} E_x \ , \tag{5.1}$$

where E_x is the component of the circuit electric field in the **X**-direction, n_0 is the zero-field refractive index, and r_{41} is the electrooptic coefficient for GaAs. Given a beam propagating in the **X**-direction, these field-dependent refractive indices will result in differential phase modulation of the beam components having electric fields $e_{y'}$ and $e_{z'}$ polarized in the $\mathbf{Y}'$ and $\mathbf{Z}'$ directions.

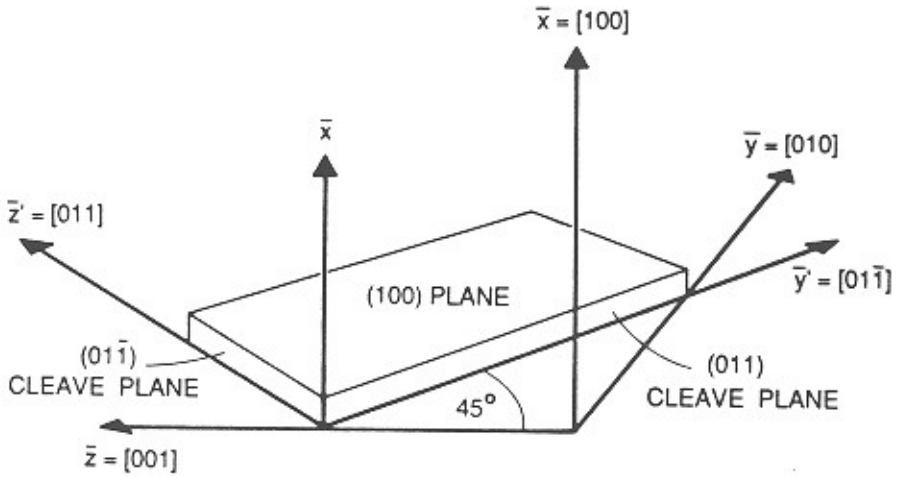


Figure 5.1: Principal axes and cleave planes in (100)-cut Gallium Arsenide.

The wave equations for propagation in the **X**-direction are

$$\frac{\partial^2 e_{y'}}{\partial x^2} = \left(\frac{n_{y'}^2}{c^2}\right)\frac{\partial^2 e_{y'}}{\partial t^2} \ ,$$

$$\frac{\partial^2 e_{z'}}{\partial x^2} = \left(\frac{n_{z'}^2}{c^2}\right)\frac{\partial^2 e_{z'}}{\partial t^2} \ , \tag{5.2}$$

where c is the speed of light in vacuum. Sinusoidal optical fields thus propagate with phase velocities $c/n_{y'}$ and $c/n_{z'}$, and are given by

$$e_{y'}(x,t) = A_{y'}\cos(\omega t - k_{y'}x + \varphi_{y'}) \ ,$$

$$e_{z'}(x,t) = A_{z'}\cos(\omega t - k_{z'}x + \varphi_{z'}) \ , \tag{5.3}$$

where $k_{y'} = \omega n_{y'}/c$ and $k_{z'} = \omega n_{z'}/c$ are the wavenumbers along the $\mathbf{Y'}$ and $\mathbf{Z'}$ axes respectively. Because of the field-dependent refractive indices (Eq. 5.1) the propagating e-fields experience a differentail phase-modulation in proportion to the electric field,

$$e_{y'}(x,t) = A_{y'}\cos\left(\omega t - kxn_{y'} + \varphi_{y'}\right) \ ,$$

$$e_{z'}(x,t) = A_{z'}\cos\left(\omega t - kx\left(n_{y'} + n_0^3 r_{41}E_x\right) + \varphi_{z'}\right) \ , \tag{5.4}$$

where now $k = 2\pi/\lambda_0 = \omega/c$ is the wavenumber for λ_0, the free space wavelength. Thus as the two polarization components propagate, they undergo a differential phase shift proportional to the x-component of the electric field, resulting in a change in the beam's polarization.

Consider the electrooptic amplitude modulator shown in Fig. 5.2. At the point of entry to the GaAs wafer, at $x = 0$, a circularly-polarized probe beam has electric fields given by

$$e_{y'}(0,t) = A\cos(\omega t)$$

$$e_{z'}(0,t) = A\cos(\omega t - \pi/2) \ . \tag{5.5}$$

After propagating through a distance W, the thickness of the substrate, the relative phase of the two field components is shifted in proportion to the electric field

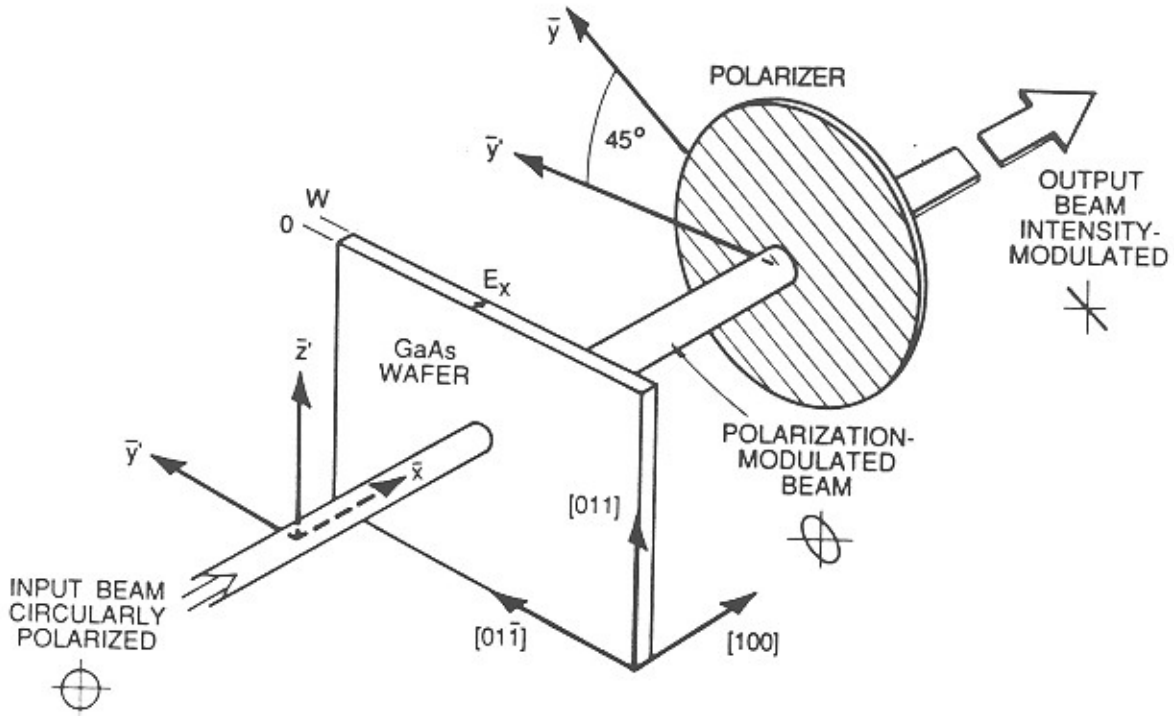


Figure 5.2: Gallium Arsenide electrooptic intensity modulator.

$$e_{y'}(W,t) = A\cos(\omega t - \varphi_0(E_y,E_z) + \Delta\varphi/2)$$

$$e_{z'}(W,t) = A\cos(\omega t - \varphi_0(E_y,E_z) - \pi/2 - \Delta\varphi/2)\ , \tag{5.6}$$

where $\varphi_0(E_y,E_z)$ is the phase shift through the substrate for $E_x = 0$ and

$$\Delta\varphi = \frac{2\pi}{\lambda_0}\int_0^W n_0^3 r_{41} E_x(x)\ dx = \frac{2\pi}{\lambda_0} n_0^3 r_{41} V \tag{5.7}$$

is the change in phase between the $\mathbf{Y}'$ and $\mathbf{Z}'$ polarizations due to the electrooptic effect. The electric field expressions of Eq. 5.5 then no longer represent circularly polarized light because of this additional phase shift $\Delta\varphi$; the polarization emerging from the substrate has changed from circular to slightly elliptical. The line integral $\int_0^W E_x(x)\ dx$ through the wafer of the x-component of the electric field is the potential difference V between the front and back surfaces of the GaAs wafer, where the probing beam enters and exits the wafer. The change in beam polarization is thus a function only of the potential difference V across the wafer at the probed point, and is independent of the particular field direction and distribution giving rise to V.

To measure this voltage-induced polarization change, the beam emerging from the GaAs is passed through a polarizer oriented parallel to the [010] ($\mathbf{Y}$) direction. With the polarizer oriented at 45 degrees to the $\mathbf{Y}'$ and $\mathbf{Z}'$ axes, its output field $e_y(t,z)$ is the difference of $e_{z'}$ and $e_{y'}$:

$$e_y = \frac{e_{z'} - e_{y'}}{\sqrt{2}} \tag{5.8}$$

Substituting in Eq. (5.6), we find that e_y is amplitude modulated by $\Delta\varphi$:

$$\begin{aligned} e_y(t, z = \text{at polarizer}) &\propto \cos(\omega t - \varphi_0(E_y,E_z) + \Delta\varphi/2) \\ &\quad - \cos(\omega t - \varphi_0(E_y,E_z) - \pi/2 - \Delta\varphi/2) \\ &\propto \cos(\omega t - \varphi_0(E_y,E_z) + \pi/4)\cos(\Delta\varphi/2 - \pi/4)\ . \end{aligned} \tag{5.9}$$

The intensity of the output beam, detected by a photodiode, is proportional to the square of e_y:

$$
\begin{aligned}
P_{out} &\propto \text{time average of} e_y^2 \\
&= 2P_0 \cos^2(\pi/4 - \Delta\varphi/2) \\
&= P_0 \left(1 + \sin\left(\frac{2\pi}{\lambda_0} n_o^3 r_{41} V \right) \right) \ , \\
&= P_0 \left(1 + \sin\left(\frac{\pi V}{V_\pi} \right) \right)
\end{aligned}
\tag{5.10}
$$

where P_0 is the output intensity with zero field in the substrate. The quantity V_π , called the half-wave-voltage, and given by

$$
V_\pi = \frac{\lambda_0}{2 n_o^3 r_{41}}
$$

is the voltage required for 180^0 phase shift between the $\mathbf{Y}'$ and $\mathbf{Z}'$ polarizations. For GaAs, $V_\pi \cong 10$ kV at a wavelength of 1.064 μm for $n_0 = 3.6$, and $r_{41} \cong 1.4 \times 10^{-12}$ m/V [19]; the argument of the sine expression is thus small for typical voltages V encountered on integrated circuits, and Eq. (5.10) can be approximated by:

$$
P_{out} \simeq P_0 \left(1 + \frac{\pi V}{V_\pi} \right) . \tag{5.11}
$$

Thus for substrate voltages up to several hundred volts, the output beam intensity is very nearly a linear function of the voltage across the substrate. The intensity of the output beam, detected by a photodiode, is thus a measure of the voltage across the substrate of the IC. To make useful measurements of the voltages in microwave and high-density digital GaAs IC's, the simplified probing geometry of Fig. 5.2 must be adapted to the conductor geometries found on these circuits.

5.2 Probing Geometries in GaAs IC's

While the simplified probing geometry of Fig. 5.2 provides modulation of the probe beam intensity in proportion to the voltage across the wafer, the arrangement is not readily applied to circuit measurements. This transmission type arrangement would require separate lenses for focusing and collecting the probe beam, precisely aligned on opposite sides of the wafer. Also, high-density interconnections on the circuit side of digital IC's and backside metallization on many microwave IC's would obstruct passage of the beam through the wafer. Reflection-type probing geometries, as shown in Fig. 5.3, provide better access to the wafer, using only a single

lens and using the IC metallization for reflection. The frontside geometry is suitable for probing microstrip transmission lines of MMIC's. The backside geometry, used with digital circuits and microwave circuits using planar transmission media, permits very tight focusing of the probing beam to a diameter limited by the numerical aperture of the focusing lens. For this probing geometry, the measured signal is proportional to the potential difference across the substrate at the probed point; if the spacings of the many conductors on the wafer frontside are small in comparison with the substrate thickness, the backside potential will be uniform, and the probed potential is then that of the probed conductor, independent of the particular field configuration.In backside probing, the probe beam modulation is sensitive to the probed conductor's voltage but is independent of nearby signal conductors, a necessity for testing high- density IC's.

In this reflection-mode probing, the incident and reflected beams, centered on the microscope lens for optimum focusing, are separated by manipulation of their polarizations by a pair of waveplates. The details of this arrangement are discussed in reference [10]; the probe beam passing through the circuit is elliptically polarized, with equal intensity in each of the two eigenpolarizations of the substrate. As with the simplified probing example using circularly polarized light, the relative phase shift of the two eigenpolarizations is proportional to the electric field (Eq. 5.7). With appropriate interference of these two polarization states, the output intensity will again vary sinusoidally with the relative phase shift:

$$I \simeq I_0 \left(1 + \frac{\pi V}{V_\pi} \right) ,$$

where I_0 is the photocurrent with $V = 0$ and where V_π is now given by

$$V_\pi = \frac{\lambda_0}{4 n_o^3 r_{41}} \simeq 5 \text{ kV} .$$

The doubled sensitivity (i.e. decreased V_π) is due to the double passage of the probe beam through the substrate.

5.3 Electrooptic Sampling

The longitudinal reflection-mode geometries provide intensity modulation proportional to voltage. With a continuous optical probe beam, the output intensity incident upon the photodiode will be a large steady-state intensity I_0 plus a small intensity perturbation following the voltage of the probed conductor; microwave-frequency or picosecond-risetime signals on the probed conductor will result in

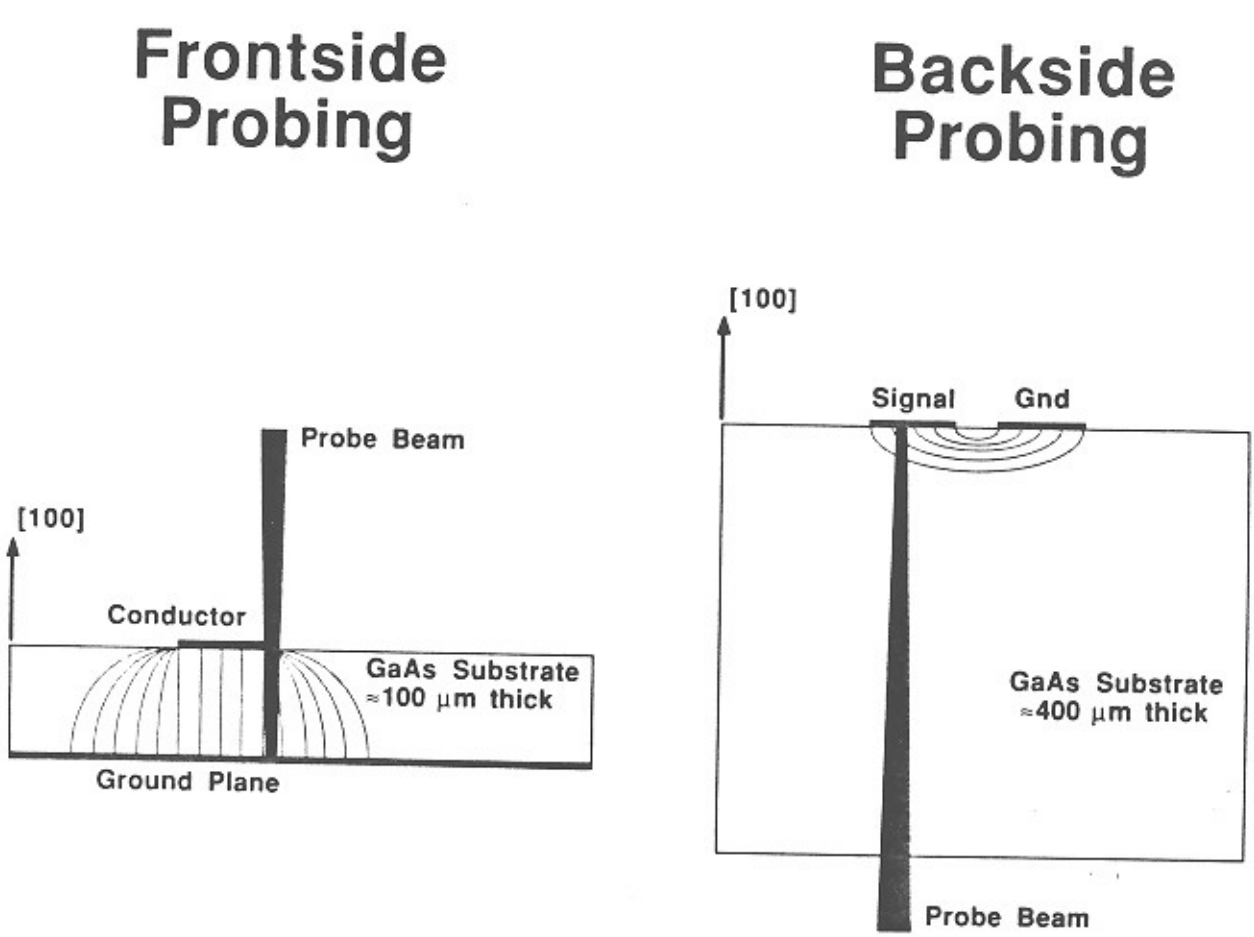


Figure 5.3: Reflection-mode probing geometries for electrooptic sampling of GaAs integrated circuits. The frontside geometry is used for probing microstrip transmission lines on MMIC's, while the backside geometry is used for probing planar transmission lines on MMIC's and wire interconnects on digital IC's.

microwave-frequency or picosecond-risetime modulation of the probe beam. Detection of this modulation would require a photodiode/receiver system with bandwidth comparable to that of the detected signal; the bandwidth of the electrooptic probing system would be limited to that of the photodiode, the receiver system, and the oscilloscope displaying the signal. With the bandwidth of commercial sampling oscilloscopes and efficient infrared photodiodes currently limited to $\simeq$20 GHz, the probing system would be limited to a bandwidth of $\simeq$14 GHz, insufficient for probing many high-speed and microwave GaAs circuits. In addition, because of the very small modulation provided by the electrooptic effect, direct detection of a probe beam intensity-modulated at microwave bandwidths would result in extremely low signal-to-noise ratio, and thus very poor instrument sensitivity.

Mode-locked laser systems in conjunction with optical pulse compressors can generate extremely short optical pulses; pulses as short as 8 fs [20] have been generated at visible wavelengths, while subpicosecond pulsewidths have been generated at the infrared wavelengths [21,22] where GaAs is transparent. Sampling techniques, using a pulsed optical probe to achieve time resolution set by the optical pulse duration and the circuit-probe interaction time, permit instrument bandwidths exceeding 100 GHz. Two related methods, synchronous sampling and harmonic mixing, are used in electrooptic sampling. In synchronous sampling, equivalent-time measurements of the voltage waveforms on probed conductors are generated in a manner similar to the operation of a sampling oscilloscope. In harmonic mixing, the electrooptic sampler measures the amplitude and phase of sinusoidal voltages on probed conductors, thus emulating a microwave network analyzer.

In equivalent-time sampling, an optical probe pulse with a repetition rate f_0 samples a repetitive voltage waveform. If the waveform repeats at Nf_0, an integer multiple of the probe repetition rate, an optical pulse interacts with the waveform every N^{th} period at a fixed point within its repetition period. Thus, over many optical pulse and voltage waveform repetitions, these pulses sample the voltage waveform at a single point in time within the cycle. Each optical pulse thus undergoes an equal modulation in its intensity; the resulting change in the average intensity of the probe beam is proportional to the signal and detected by a photodiode receiver whose frequency response can be much less than the optical pulse repetition frequency.

To detect the entire time waveform, the waveform frequency is increased by a small amount Δf (Fig. 5.4.) The probe pulses are thus slowly delayed with respect to the waveform, sampling successively delayed points, so that the average intensity at the photodiode changes in proportion to the waveform, but repeating at a rate Δf, as shown below.

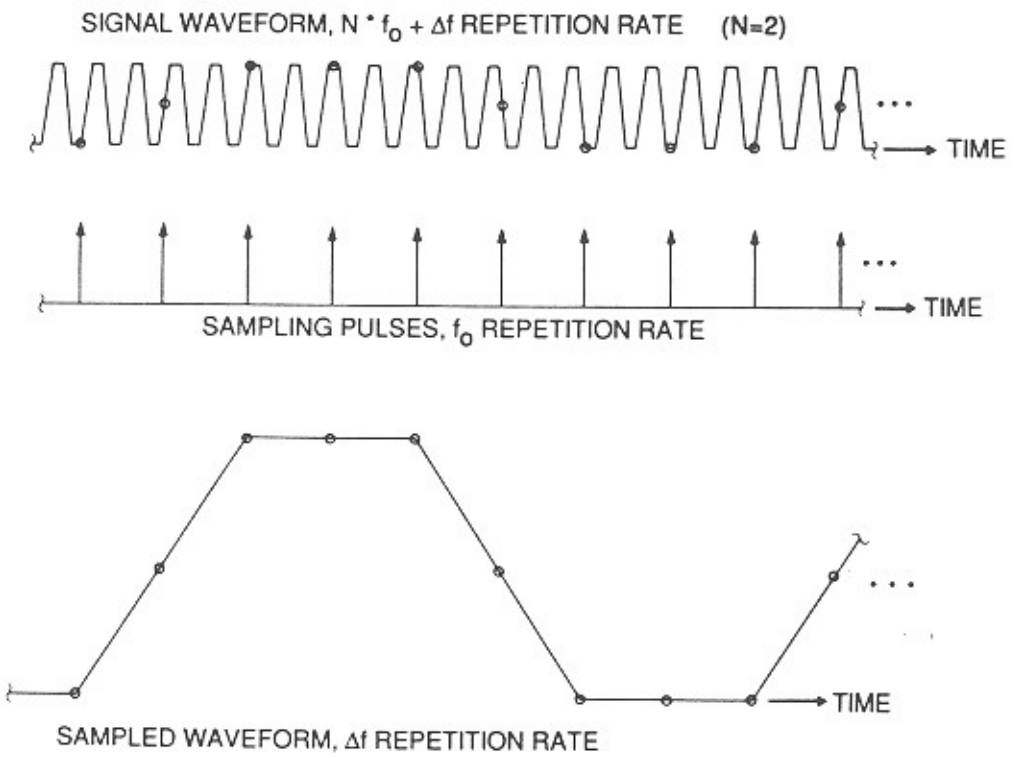


Figure 5.4: Equivalent-time sampling. Typically, $f_0 \simeq 82$ MHz, $0 \leq N \leq 500$, and $\Delta f \approx$ 10-100 Hz.

After Eq. 5.11, the photodiode current $i(t)$ is

$$i(t) \simeq I_0 \left(1 + \frac{\pi V}{V_\pi} \right) ,$$

where $I(t) = rP(t)$ is the photocurrent with $V(t) = 0$, r is the photodiode responsivity, $P(t)$ is the laser intensity, and $V(t)$ is the signal on the probed conductor. The laser intensity $P(t)$, and hence also $I(t)$ are a series of impulses:

$$I(t) = \frac{I_0}{f_0} \sum_{m=-\infty}^{m=+\infty} \delta(t - m/f_0) \ . \tag{5.12}$$

Thus $i(t)$ is

$$i(t) = I(t) + \frac{\pi}{V_\pi} \frac{I_0}{f_0} \sum_{m=-\infty}^{m=+\infty} V(m/f_0) \delta(t - m/f_0) \ ,$$

but the probed voltage $V(t)$ is periodic at frequency $(Nf_0 + \Delta f)$, and hence

$$\begin{aligned} i(t) &= I(t) + \frac{\pi}{V_\pi} \frac{I_0}{f_0} \sum_{m=-\infty}^{m=+\infty} V \left(\frac{m}{f_0} - \frac{mN}{Nf_0 + \Delta f} \right) \delta(t - m/f_0) \\ &= I(t) + \frac{\pi}{V_\pi} \frac{I_0}{f_0} \sum_{m=-\infty}^{m=+\infty} V \left(\frac{m\Delta f}{f_0(Nf_0 + \Delta f)} \right) \delta(t - m/f_0) \\ &= I(t) + \frac{\pi}{V_\pi} \frac{I_0}{f_0} \ V \left(\frac{t \ \Delta f}{Nf_0 + \Delta f} \right) \sum_{m=-\infty}^{m=+\infty} \delta(t - m/f_0) \\ &= I(t) \left(1 + \frac{\pi}{V_\pi} \ V \left(\frac{t \ \Delta f}{Nf_0 + \Delta f} \right) \right) \ . \end{aligned} \tag{5.13}$$

Equation 5.13 shows that the photodiode current is a pulse train whose amplitude varies at a rate Δf . The receiver then averages (low-pass filters) the photocurrent over a period much longer than 1/fo, eliminating the individual pulses. The averaged photocurrent $i_{out}(t)$ is then continuous and varies with $V(t)$, but at a slow repetition rate Δf:

$$i_{out}(t) = I_0 \left(1 + \frac{\pi}{V_\pi} \ V \left(\frac{t \ \Delta f}{Nf_0 + \Delta f} \right) \right) \ . \tag{5.14}$$

Typically f_0 is $\simeq$82 MHz, N varies from 1 to 500 for circuit drive frequencies to 40 GHz, and Δf is 10 to 100 Hertz.

Equivalent-time sampling can also be described as harmonic mixing. The signal detected by the photodiode receiver is proportional to the product of the laser signal and the measured signal.The fourier spectrum of the detected signal is thus the convolution of the spectra of the laser pulse train and the probed signal:

$$F(i(t)) \simeq F(I(t)) * \left(\delta(f) + \frac{\pi}{V_\pi} F(V(t)) \right) \ , \tag{5.15}$$

where F is the Fourier transform operator, $\delta(f)$ is the delta function, and $*$ represents the convolution operation. Figure 5.5 shows a schematic representation of this convolution for a mode-locked laser spectrum and a single microwave frequency signal. Scaled replicas of the signal appear in the laser intensity spectrum as amplitude-modulation sidebands around each laser harmonic. To measure the magnitude and phase of a sinusoidal signal, the circuit signal frequency is offset by some convenient offset Δf from the nearest laser harmonic, resulting in a signal at Δf at the sampler output whose magnitude and phase follow that of the circuit signal. The magnitude and phase of output signal at Δf are then measured with quadrature synchronous detectors (i.e. a 2-phase "lock-in" amplifier), or for scalar measurements, a spectrum analyzer.

5.4 Electrooptic Sampling System

The sampling system, shown schematically in Fig. 5.6, can be grouped into three sections; the laser system for optical pulse generation, the microwave instrumentation for driving the IC under test, and the receiver system for signal processing and data acquisition. The laser system consists of a mode-locked, Nd:YAG laser, a fiber-grating pulse compressor, and a timing stabilizer feedback system. The Nd:YAG laser, a commercially available system, produces 1.06μm, 90 ps pulses at an 82 MHz rate. The laser has free-running pulse-to-pulse timing fluctuations of 4 ps rms, reduced to less than 300 fs rms by a phase-lock-loop feedback system which synchronizes and stabilizes the laser pulse timing with respect to the microwave synthesizer (Chapter 6). The fiber-grating pulse compressor shortens the pulses to 1.5 ps FWHM (full width at half maximum) [23]. The beam passes through a polarizing beamsplitter and two waveplates to adjust its polarization, then is focused through the IC substrate with a microscope objective to a 3μm spot on the probed conductor (backside probing) or a 10 μm spot on the ground plane adjacent to the probed conductor (frontside probing). The reflected light is analyzed by the polarizing beamsplitter; the change in intensity, proportional to the voltage across the

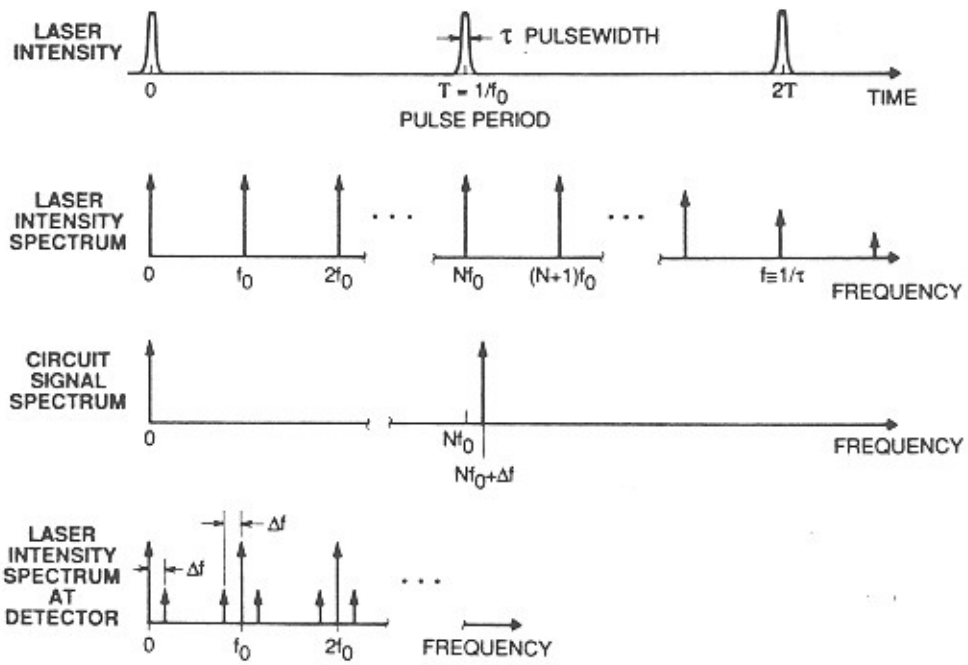


Figure 5.5: Electrooptic harmonic mixing.

GaAs substrate, is detected by a photodiode connected to a vector receiver. The circuit under test is driven with a microwave synthesizer.

Equivalent-time sampling is used to view time waveforms: the synthesizer is set to an exact multiple of the laser repetition rate (82 MHz), plus a small frequency offset Δf (1-100 Hz). To suppress laser intensity noise the synthesizer is pulse or phase modulated at 1-10 MHz; the resulting intensity modulation at the photodiode varies at the slow offset rate Δf in proportion to the measured signal and is detected by a narrowband synchronous receiver. Harmonic mixing is used for vector voltage measurements; the synthesizer is set to an exact multiple of the laser repetition rate plus a 1-10 MHz frequency offset , and the receiver is configured as a vector voltmeter to measure the magnitude and phase of the received signal.

5.5 Bandwidth

The system's bandwidth or time resolution is determined by the optical pulsewidth, the pulse-to- pulse timing jitter of the laser with respect to the microwave synthesizer driving the circuit, and the optical transit time of the pulse through region of fields within the GaAs substrate. The overall time resolution is the root-mean-square sum of these values

$$\Delta\tau_{total} = \sqrt{\Delta\tau_{pulse}^2 + \Delta\tau_{jitter}^2 + \Delta\tau_{OTT}^2} \ , \tag{5.16}$$

where $\Delta\tau_{pulse}$is the rms optical pulsewidth, $\Delta\tau_{jitter}$is the rms pulse-to-pulse jitter, and $\Delta\tau_{OTT}$ is the rms optical transit time of the pulse through substrate.

The relation between the time resolution and the frequency bandwidth for the optical pulse is given by

$$\Delta\tau_{FWHM} = \frac{0.312}{f_{\mathrm{3dB}}} \ , \tag{5.17}$$

where for a Gaussian pulseshape $\Delta\tau_{FWHM} = 2.35\Delta\tau_{rms}$ is the full width at half-maximum and f_{3dB} is the half power frequency.

The pulse compressor and the laser timing stabilizer serve to reduce $\Delta\tau_{pulse}$ and $\Delta\tau_{jitter}$, increasing the system bandwidth. Timing jitter influences both bandwidth and sensitivity; the impulse response of the sampling system is the convolution of the optical pulseshape with the probability distribution of its arrival time (neglecting optical transit time), while those Fourier components of the jitter lying within the detection bandwidth of the receiver introduce noise proportional to the time

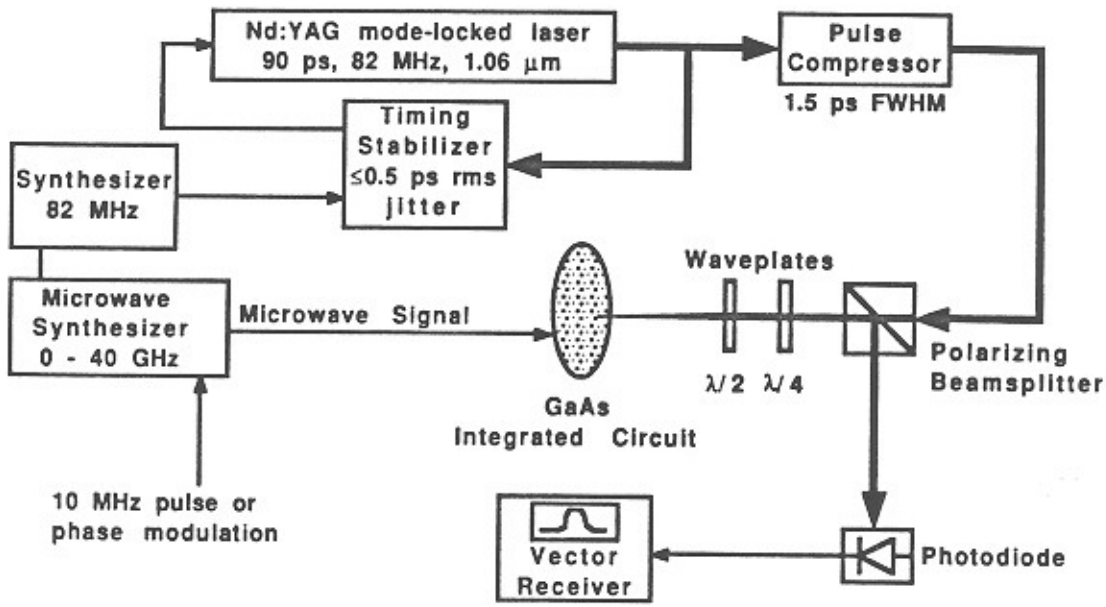


Figure 5.6: System for direct electrooptic sampling.

derivative of the measured waveform. Stabilization of the laser timing is thus imperative for low-noise measurements of microwave or picosecond signals. With 300 fs rms timing jitter and a 1.5 ps FWHM optical pulse, the 3 dB system bandwidth should be approximately 200 GHz; due to wings (long-duration substructure) on the pulse emerging from the compressor, the 3 dB system response is reduced to approximately 80-100 GHz.

In general the optical transit time of the pulse in the GaAs substrate can be neglected for microwave IC's. Because the optical and microwave dielectric constant in GaAs are nearly equal, microwave transmission lines have a cutoff frequency for higher-order modes roughly equal the inverse of the optical transit time. Well-designed microwave circuits operate at frequencies well below the multimode cutoff frequency. Only when measuring interconnects near or above the cutoff frequency (where dispersive characteristics are of interest) must the optical transit time be considered. For example, the optical transit time for a 125 μm thick substrate, typical of MMIC's operating at frequencies below 40 GHz, is 3 ps, corresponding to a 3 dB response rolloff of >100 GHz.

5.6 Sensitivity

If the high measurement bandwidth provided by the electrooptic sampler is to be useful, the instrument must also provide sufficient sensitivity to readily observe the comparatively small voltages typical in high-speed GaAs circuits. As in any system, sensitivity is determined by the signal to noise ratio; the instrument's sensitivity, or minimum detectable voltage, is the probed voltage which results in a measured signal equal to the measurement system's noise voltage. Most noise sources have power spectral densities which are independent of frequency ("white" noise), resulting in a noise voltage proportional to the square root of the signal acquisition bandwidth. The acquisition bandwidth is the bandwidth or integration time of the low-frequency photodiode/receiver system and sets the maximum rate Δf at which the sampler can scan a voltage waveform. The minimum detectable voltage is thus proportional to the square root of the acquisition bandwidth, and is expressed in units of volts per root Hertz; smaller minimum detectable voltages, in units of $V/\sqrt{\text{Hz}}$, permit more rapid measurement acquisition for fixed measurement accuracy. With appropriate system design and signal processing, the various sources of noise in the electrooptic sampler can be reduced or their effect eliminated, permitting low-noise voltage measurements at scan rates up to ~100 Hertz. Noise sources in the electrooptic sampling system include probe beam shot noise

(observed as shot noise of the photodiode quiescent current), receiver noise, laser phase noise and low frequency $1/f$ intensity noise, and intensity noise from the pulse compressor.

Including these noise terms, and dropping the constant term I_0, the output of the electrooptic sampler receiver $i_{out}(t)$ is

$$i_{out}(t) = \frac{I_0 \pi}{V_\pi} V\left(\frac{t\ \Delta f}{N f_0 + \Delta f}\right) + i_{SN} + i_{phase} + i_{receiver} + i_{laser} + i_{compressor}\ .$$

The shot noise, i_{SN}, associated with the DC component of the photodiode current, has a variance given by

$$\overline{i_{SN}^2} = 2qI_0B\ ,$$

where q is the electron charge, the horizontal bar denotes the statistical expectation, and B is the measurement acquisition bandwidth. The receiver equivalent input noise current, $i_{receiver}$, is

$$\overline{i_{receiver}^2} = \frac{4kTB}{R_L} + \overline{i_{amp}^2} + \frac{\overline{v_{amp}^2}}{R_L^2}\ ,$$

where k is Boltzmann's constant, T is the absolute temperature, and R_L is the photodiode load resistor. The first term is the Johnson or thermal noise of the load resistor, and $\overline{i_{amp}^2}$ and $\overline{v_{amp}^2}$ are the equivalent input noise current and equivalent input noise voltage of the amplifier following the photodiode. Noise (i_{phase}) arises in proportion to the product of the laser timing jitter and the time derivative of the measured waveform, and has a variance approximated by

$$\overline{i_{phase}^2} \approx \frac{I_0^2\pi^2}{V_\pi^2}\overline{\left(\frac{dV}{dt}\right)^2}\int_0^B \frac{L(f)}{f_0^2}\ df\ ,$$

where $\overline{(dV/dt)^2}$is the mean-squared slope of the voltage waveform $V(t)$ and $L(f)$ is the single-sided laser phase noise spectral density relative to the carrier power at the first harmonic of the laser pulse repetition frequency. The spectral densities of i_{laser}, the laser $1/f$ intensity noise, and $i_{compressor}$, the compressor noise, depend on the design and adjustment of these components.

Of the above noise terms, most can be reduced to a negligible levels compared to the shot noise. Receiver noise is reduced to a level below the shot noise limit by appropriate receiver design; R_L is made large in comparison with $2kT/qI_0$ so that Johnson noise is well below shot noise, and the photodiode amplifier is selected for low input noise. At frequencies below $\simeq$100 kHz the laser intensity noise is

approximately 80 dB above the shot noise level, contributing a 10^4:1 degradation to the minimum detectable voltage. To translate the signal detection to a frequency where this $1/f$ laser noise is below the shot noise limit, the microwave signal to the IC is modulated at 1–10 MHz. The resulting modulation component of the photocurrent, proportional to the sampled voltage, is detected with a 1–10 MHz narrowband vector receiver. For sequential digital circuits which do not operate correctly with chopped excitation, a small-deviation phase modulation is used. In this case the received signal, proportional to the derivative of the sampled waveform, is integrated in software . The pulse compressor is also a source of excess amplitude noise, due to stimulated Raman scattering (SRS) and polarization noise in the non-polarization-preserving fiber [24], which may be the result of stimulated Brillioun scattering.Raman scattering is suppressed by keeping the optical power in the fiber well below the Raman threshold. The effect of polarization noise is suppressed by adjusting the polarization state at the fiber output to maximize the transmission through the grating. Variations in the fiber polarization state then result in only second-order variations in the compressor output intensity. The fiber is place in a temperature stabilized chamber to prevent drift from this polarization state. The pulse compressor typically contributes 10–20 dB of excess amplitude noise. Finally, the laser timing stabilizer reduces $L(f)$ by approximately 25 dB at low frequencies; phase noise is not significant to frequencies of $\sim$ 20 GHz.

Given suppression of these excess noise sources, the system sensitivity is set by the signal to shot noise ratio. Setting the signal current $I_0 \pi V_{min}/V_\pi$ equal to the shot noise current, and normalizing to a 1 Hz acquisition bandwidth, δf=1 Hz, the minimum detectable voltage is

$$V_{min} = \frac{V_\pi}{\pi}\sqrt{\frac{2q}{I_0}} \ . \tag{5.18}$$

For the reflection-mode probing geometries, $V_\pi \simeq 5$ kV, while the average photocurrent I_0 is typically 1 mA. Then, the minimum detectable voltage is

$$V_{min} = 30 \ \frac{\mu V}{\sqrt{\text{Hz}}} \ .$$

Typically, $V_{min} \simeq 70 \ \mu V/\sqrt{\text{Hz}}$ is observed experimentally due to 5- -10 dB of excess noise from the pulse compressor; this sensitivity is sufficient to acquire measurements at scan rates of 10–100 Hz with a noise floor of a few millivolts.

5.7 Measurements on Microwave Distributed Amplifiers

A variety of GaAs circuits have been studied using the electrooptic sampling system, including digital inverter chains, frequency dividers (flip-flops), and multiplexers / demultiplexers, and microwave transmission structures and distributed amplifiers (see Reference 10 for a review of these experiments). Distributed amplifiers were one of the earliest demonstration vehicles for the electrooptic sampling system. Within the various sources of error which then existed in the electrooptic sampling system (strong intensity and pulsewidth fluctuations from the optical pulse compressor) careful circuit measurements were made, and the results were correlated with circuit models and known circuit parameters. In the 18 months since these measurements were made, the stability of the optical pulse compressor has been substantially improved, the system bandwidth increased and noise floor decreased, and a quadrature synchronous detector has been implemented to permit vector network characterization. These time-domain large-signal and scalar frequency-domain small-signal measurements are thus a small subset of the microwave circuit measurements now possible by direct electrooptic sampling.

In a distributed amplifier, a series of small transistors are connected at regular spacings between two high-impedance transmission lines (Fig. 5.7). The high-impedance lines and the FET capacitances together form synthetic transmission lines, generally of 50 Ω characteristic impedance. Series stubs are used in the drain circuit; at low frequencies these act primarily as shunt capacitance to ground, and equalize the phase velocity of the gate and drain synthetic lines. At higher frequencies the series stubs provide partial impedance-matching between the high output impedances of the FET's and the low 50 Ω load impedance. By using small devices at small spacings, the cutoff frequencies of the synthetic lines can be made much larger than the bandwidth limitations associated with the line attenuations arising from FET gate and drain conductances; thus, gain-bandwidth products approaching f_{max} can be attained [25]. The amplifiers studied included a five-FET TWA for 2-18 GHz using microstrip transmission lines (Fig. 5.8), and a similar amplifier using coplanar waveguide transmission lines [26].

The frequency response and distortion characteristics of these circuits depend upon the propagation characteristics of microwave signals along the gate and drain transmission lines. The bandwidth and gain flatness are set by the finite cutoff frequencies of the periodically- loaded lines, the line losses due to FET input and output conductances, and the mismatch between gate and drain propagation velocities. The gain compression characteristics are set by several saturation mechanisms in the transistors, by the power at which each mechanism occurs in each FET, and

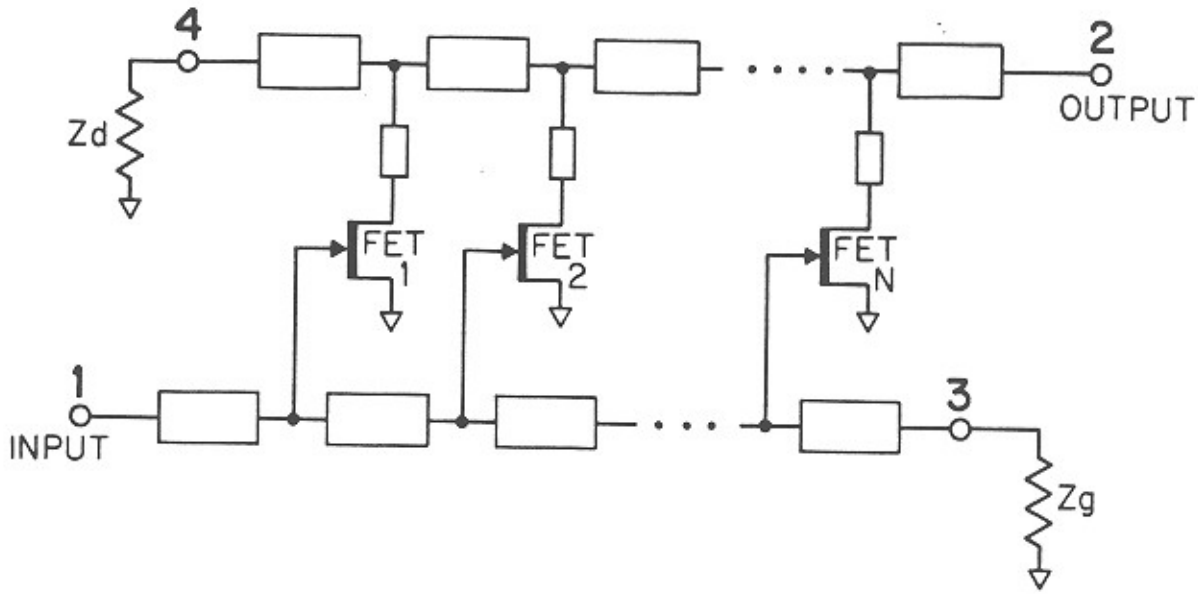


Figure 5.7: Traveling-wave amplifier topology.

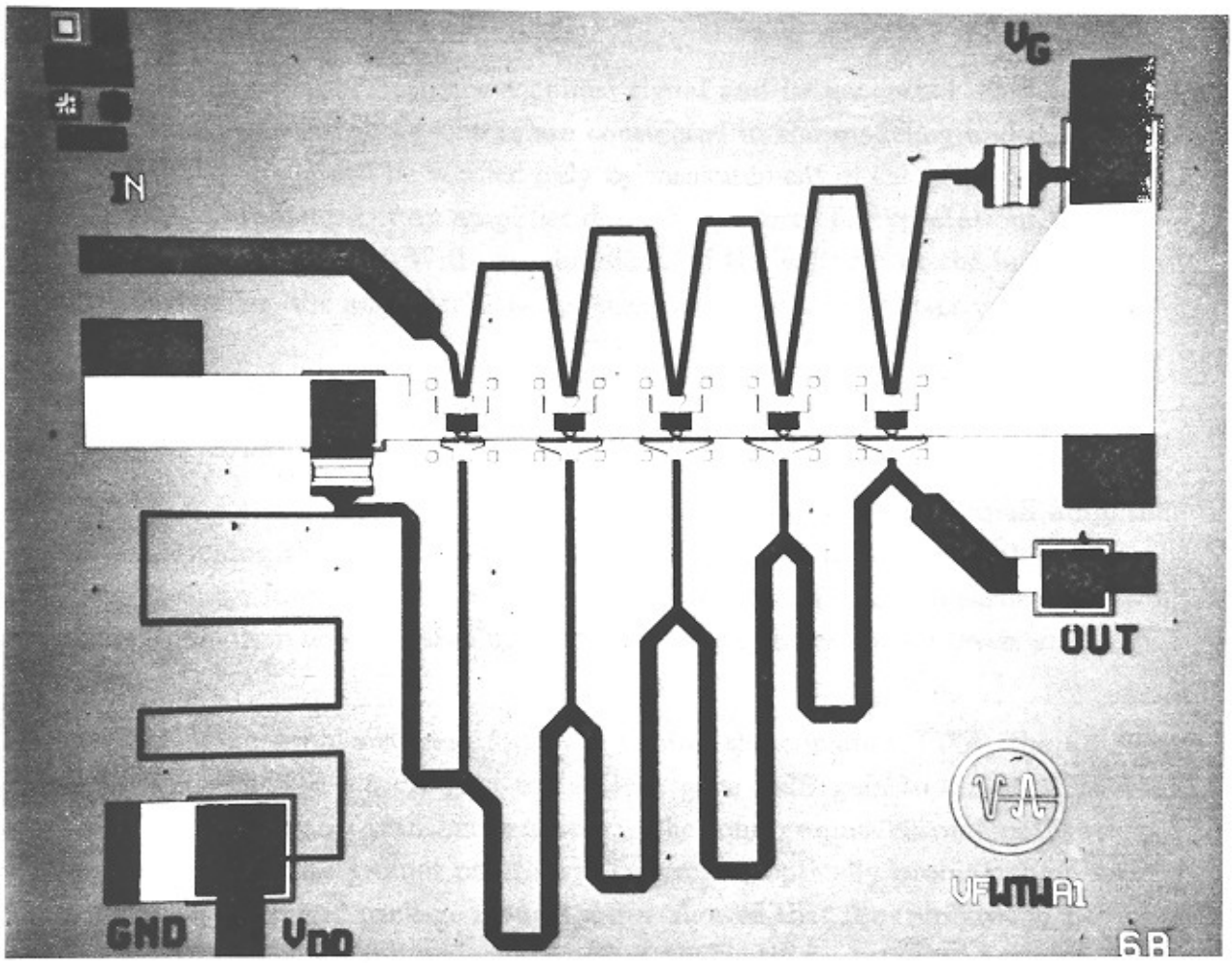


Figure 5.8: 2-18 GHz TWA using microstrip transmission lines. Photo courtesy of Majid Riaziat, Varian Research Center.

by the propagation of both the ampified signal and its generated distortion products. While some of these factors are considered in the modeling and design of of a TWA, the model can be verified only by measurement of the amplifier's external scattering parameters; if an amplifier does not perform to expectations, the cause is not readily identified. With measurements of the voltages at the internal nodes of the amplifier, the amplifier's characteristics can be much better understood.

5.7.1 Small-Signal Measurements

Driving the TWA input with a swept-frequency sinusoid of small amplitude and positioning the laser probe near the FET gate and drain terminals, the small-signal transfer function from the input to each of these nodes is measured, showing the relative drive levels at the FET gates and the relative output levels at the FET drains.

Packaging problems were found in testing the coplanar TWA; the amplifier chip, which showed 5.5 dB gain to 19 GHz, gave 5 dB gain to only 8 GHz when bonded to microstrip transmission lines. The long ground current path between the chip and fixture ground planes was suspected; optically probing the potential between the chip and package ground planes showed that the chip ground potential is only 5 dB below the input signal (Fig. 5.9), indicating substantial package ground inductance. This inductance, in series with the FET sources, provides feedback and thus degrades gain. Coplanar TWA chips directly probed with microwave wafer probes do not exhibit this degradation.

The microstrip amplifier provides 6-dB gain to 18 GHz, while simulation predicts 7 dB gain to 20 GHz. The gate voltage curves for this amplifier (fig 5.10) show several features: the rolloff beyond 18 GHz is the cutoff of the synthetic gate line, the slow rolloff with frequency is the gate line attenuation, and the ripples are standing waves resulting from the gate line being misterminated. These data were compared to simulations using the linear simulation and optimization program SuperCompact; the simulator's optimizer was then used to adjust process- dependent circuit parameters to obtain the best fit to the measured data (Fig. 5.11). Model gate termination resistance increased to 80Ω, C_{gs} increased from 1.0 to 1.14 pF/mm, C_{gd} increased from 0.03 to 0.06 pF/mm, source resistance increased from 0.58 to 0.72 Ω, and the source inductance decreased from 0.14 to 0.10 nH; these parameters fall within normal process variations. Interference between the forward and reverse waves on the drain line results in strong frequency dependence of the drain voltages (figs. 5.12 and 5.13); this can be predicted by simple analysis.

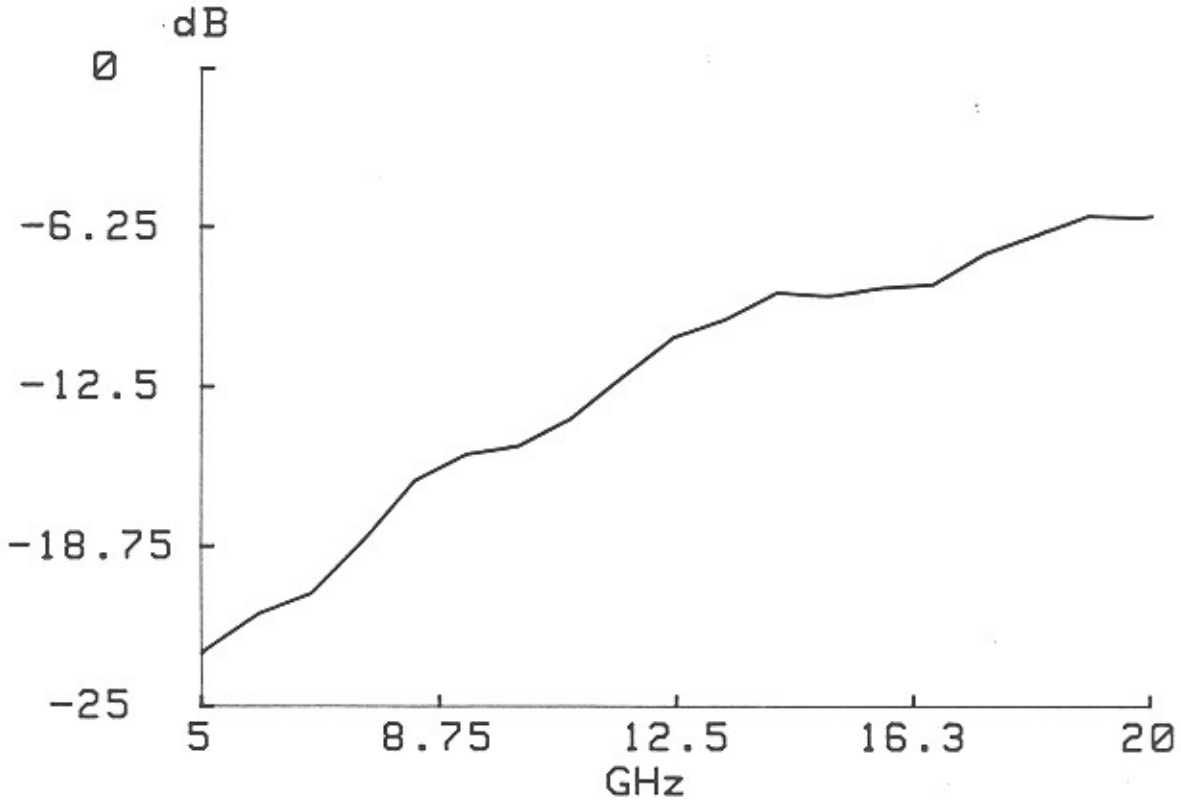


Figure 5.9: Chip ground potential relative to carrier potential, as a fraction of input voltage on the coplanar TWA.

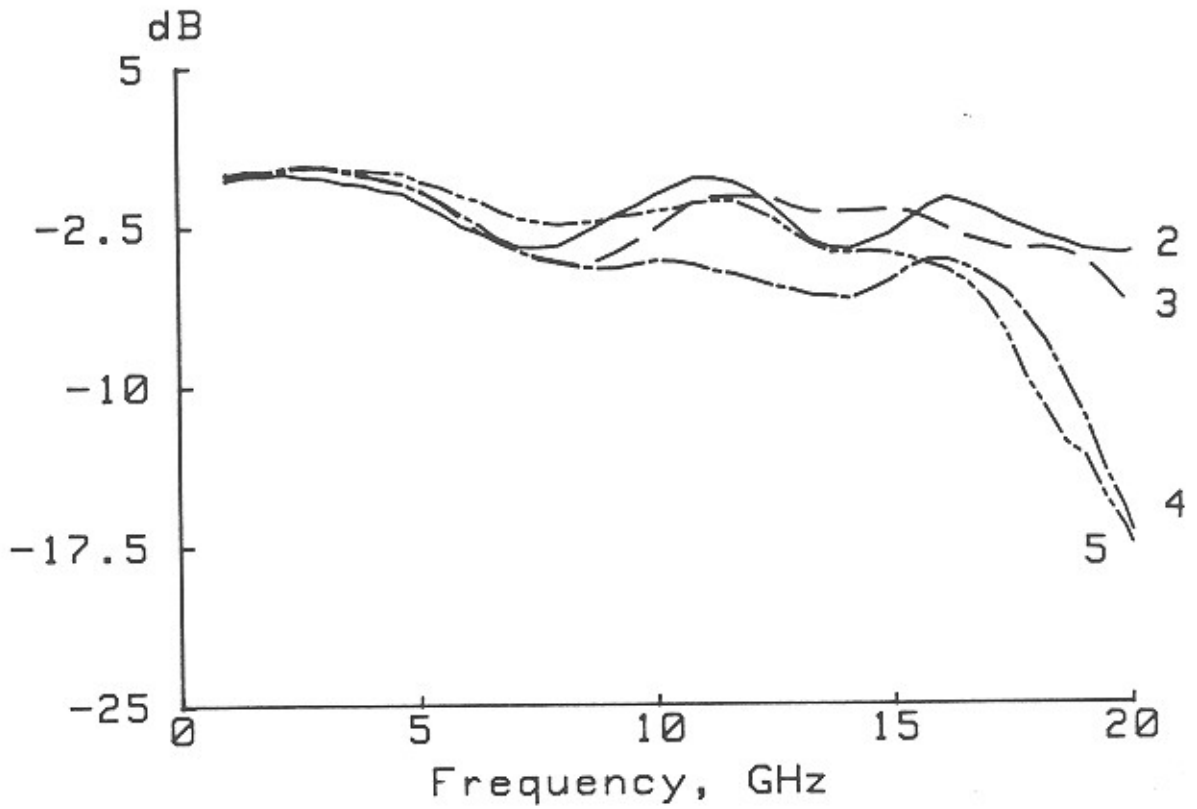


Figure 5.10: Small-signal voltages at the gates of the microstrip distributed amplifier.

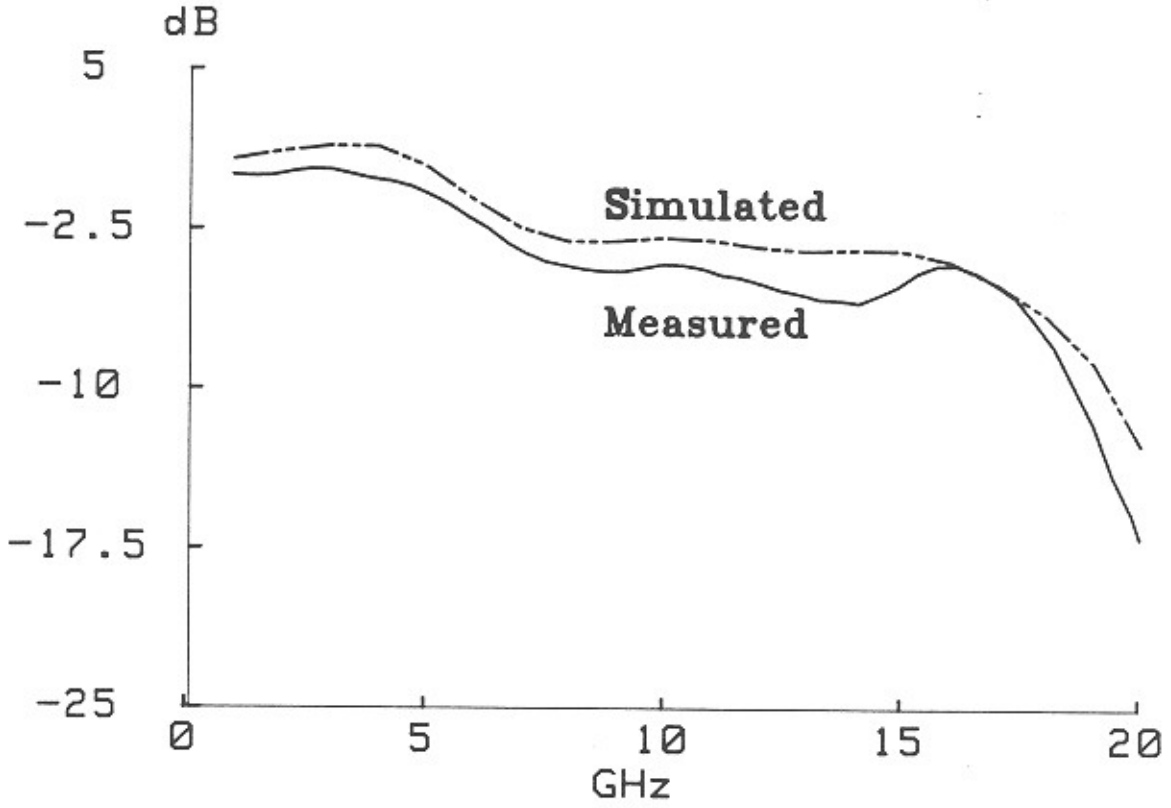


Figure 5.11: Comparison between simulated and measured gate 4 voltage after adjustment of the model to obtain the best fit to measurements.

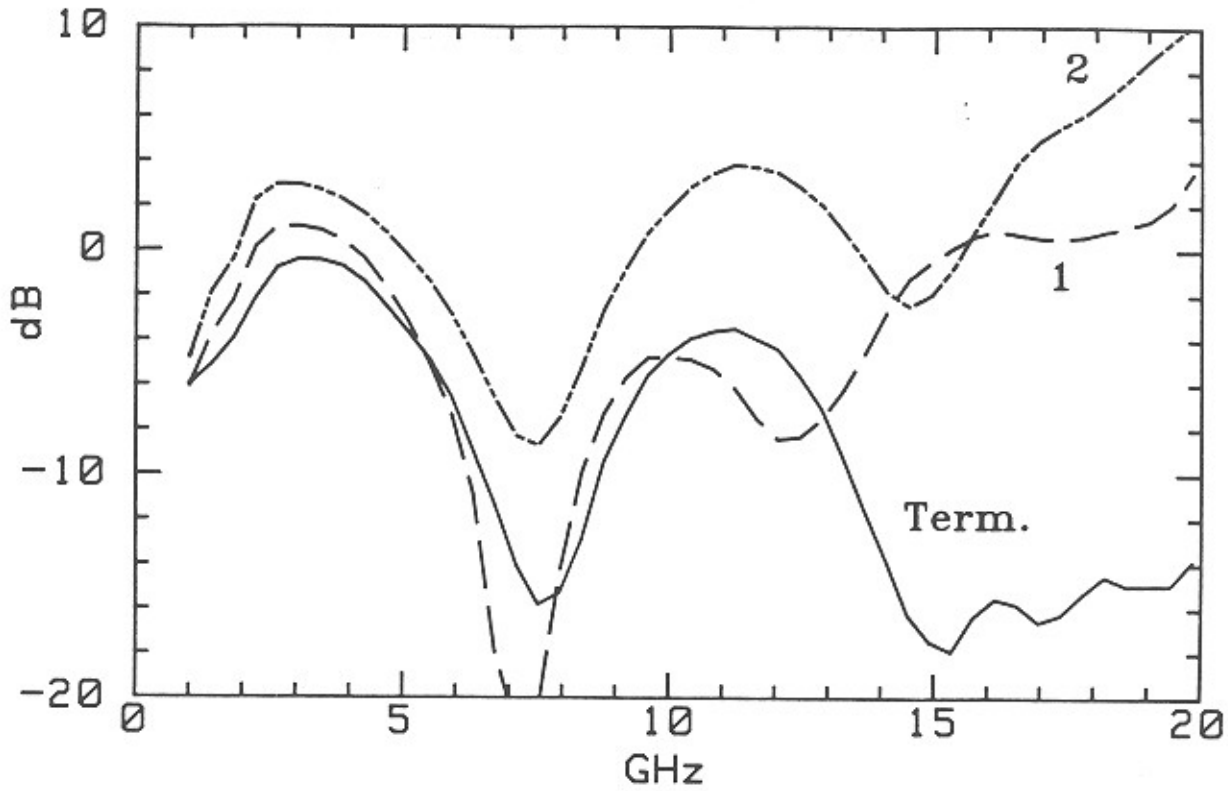


Figure 5.12: Small-signal voltages at the first and second drains, and at the drainline reverse termination resistor, in a 2-18 GHz distributed amplifier using microstrip transmission lines.

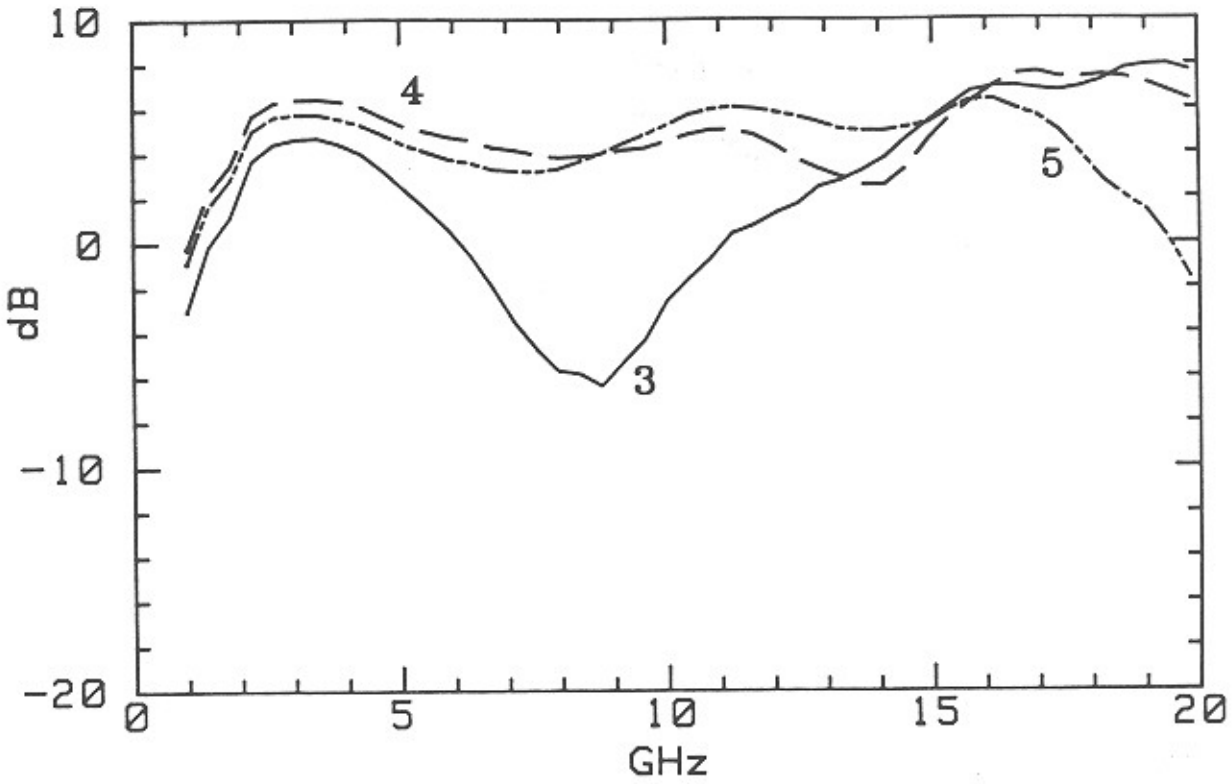


Figure 5.13: Small-signal voltages at drains 3,4, and 5 of the microstrip distributed amplifier.

5.7.2 Drain Voltage Distribution

After Ayasli [27], if the wavelength is much greater than the spacings between the FET's, the synthetic lines can be approximated as continuous structures coupled by a uniformly distributed transconductance. The lines then have characteristic impedances and phase velocities given by the sum of distributed and lumped capacitances and inductances per unit length; the line impedances (Z_{0g},Z_{0d}) and phase velocities (v_{pg}, v_{pd}) are generally made equal. The lines then have propagation constants given by

$$\gamma_g = \alpha_g + \beta_g \simeq \frac{r_g \omega^2 C_{gs}^2 Z_{0g}}{2l} + j\omega/v_{pg} \tag{5.19}$$

$$\gamma_d = \alpha_d + \beta_d \simeq \frac{G_{ds} Z_{0d}}{2l} + j\omega/v_{pd} \ , \tag{5.20}$$

where l is the FET spacing, C_{gs} is the gate-source capacitance, r_g is the gate resistance, G_{ds} is the drain-source conductance, and a forward propagating wave is of the form $e^{-\gamma z}$. The voltage along the drain line is

$$V_d(z) = \frac{-g_m Z_{0d} V_{in}}{2l} \ e^{-\gamma_g z} \ \left(\frac{1 - e^{(\gamma_g - \gamma_d)z}}{\gamma_g - \gamma_d} + \frac{1 - e^{(\gamma_g - \gamma_d)(z-nl)}}{\gamma_g + \gamma_d} \right) \ , \tag{5.21}$$

where n is the number of FET's, g_m is the FET transconductance, V_{in} is the input voltage, and z is the distance along the drain line with the origin located at the drainline reverse termination. Ignoring line attenuation and assuming equal gate and drain phase velocities, converts Eq. (5.21) to

$$\left\| V_d(z) \right\| = \frac{g_m Z_{0d} V_{in}}{2l} \sqrt{z^2 + \frac{z \sin[2\beta(nl - z)]}{\beta} + \frac{\sin^2[\beta(nl - z)]}{\beta^2}} \ , \tag{5.22}$$

which is plotted in Fig. 5.14. The above analysis neglects the matching effect of the series drain stubs. Because these stubs provide partial high-frequency impedance-matching between the FET's and the drain transmission line, at high frequencies the drain voltages are larger than the voltages at the tap points at which the drain series stubs connect to the drain line (Fig. 5.15). Drain-line voltage variation also arises from reflections from the drain-line reverse termination, which, owing to process variations, was 80Ω in the device tested. We see from Fig. 5.14 that at high frequencies the power absorbed by the drain-line reverse termination is negligible. Furthermore, at high frequencies the input power is absorbed primarily in the FET input resistances and not in the gate-line termination. Thus, at high

frequencies, the TWA is a directional coupler with gain. The terminations reduce the low-frequency power gain but do not waste substantial available gain near the cutoff frequency. The predicted frequency-dependent drain voltage distributions also complicate the large-signal operation of the amplifier.

5.7.3 TWA Saturation Mechanisms

The nonlinearities and power limitations in FET amplifiers include gate forward conduction, pinchoff, drain saturation, and drain breakdown. Gate forward conduction, limiting the gate voltage for linear operation to maximum of approximately +0.5 V, thus also limits the drain current to a maximum of I_F, the drain current at V_{gate}=0.5 V. For linear operation, the gate voltage must also be more positive than the pinchoff voltage V_P and the drain voltage must be below the drain breakdown voltage V_{BR}. Finally, nonlinear operation results if the drain voltage is less than $(V_{gate} - V_P)$, the voltage necessary to pinch off the drain end of the FET channel; we refer to this as drain saturation.

In a single FET amplifier, the load-line can be chosen as in Fig. 5.16 so that all these limits are reached simultaneously, maximizing the power output before saturation. From this load-line the sinusoidal gate and drain voltages must be 180 degrees out of phase. Ayasli [28] and Ladbrooke [29] extend this result to traveling-wave amplifiers by assuming that under appropriate design conditions (tapered gate and drain lines) the gate voltages and drain voltages of all FET's can be made equal, thus causing all devices to saturate simultaneously. In the case where the gate and drain losses can be neglected, it can be shown that if the synthetic gate and drain lines have characteristic impedances Z_{0g} and Z_{0d} and phase velocities v_{pg} and v_{pd} of the form

$$Z_{0d}(z) = K/z$$

$$Z_{0g}(Z) = Z_0$$

$$v_{pd}(z) = v_{pg}(z) = v_p \ , \tag{5.23}$$

and if the drain-line reverse termination is omitted and the output load resistance Z_{load} is set at

$$Z_{load} = K/nl \ , \tag{5.24}$$

then the voltage along the drain line will be uniform as follows:

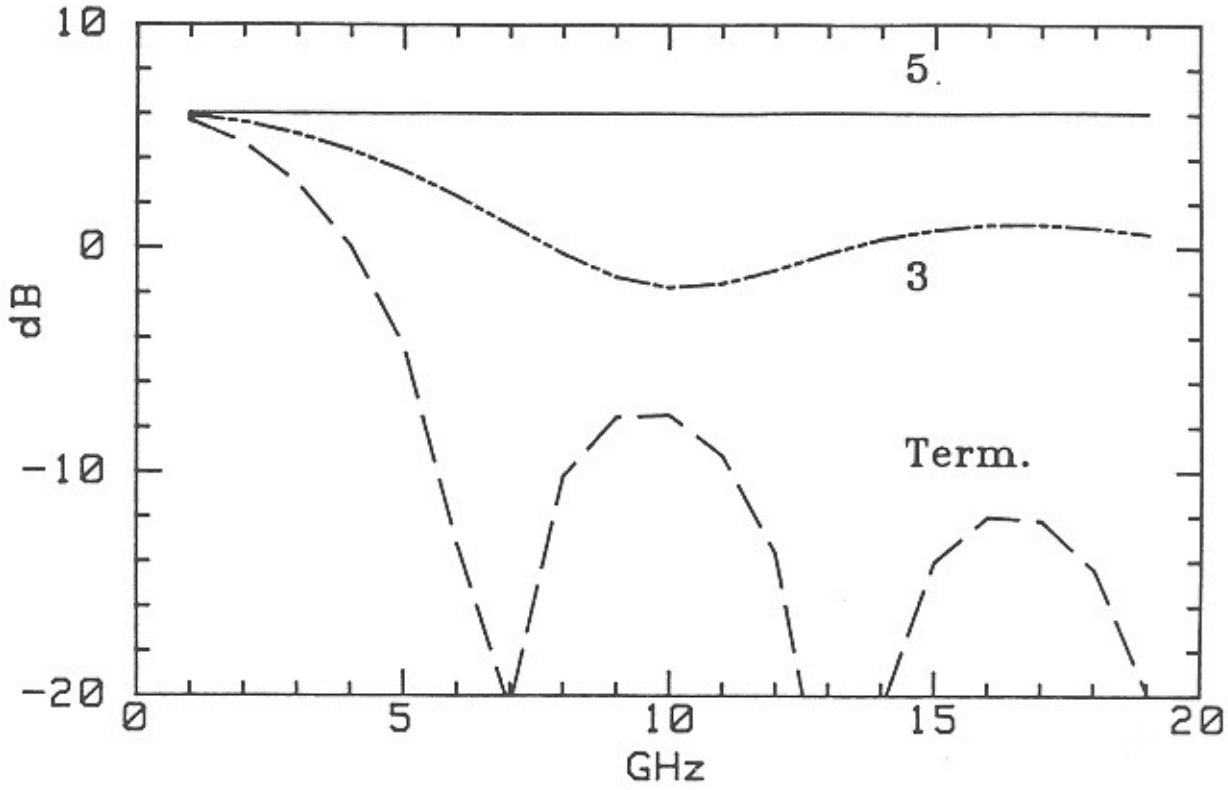


Figure 5.14: Voltage versus frequency, referenced to input voltage, at drain 5, drain 3, and drain reverse termination as calculated by Eq. (5.22).

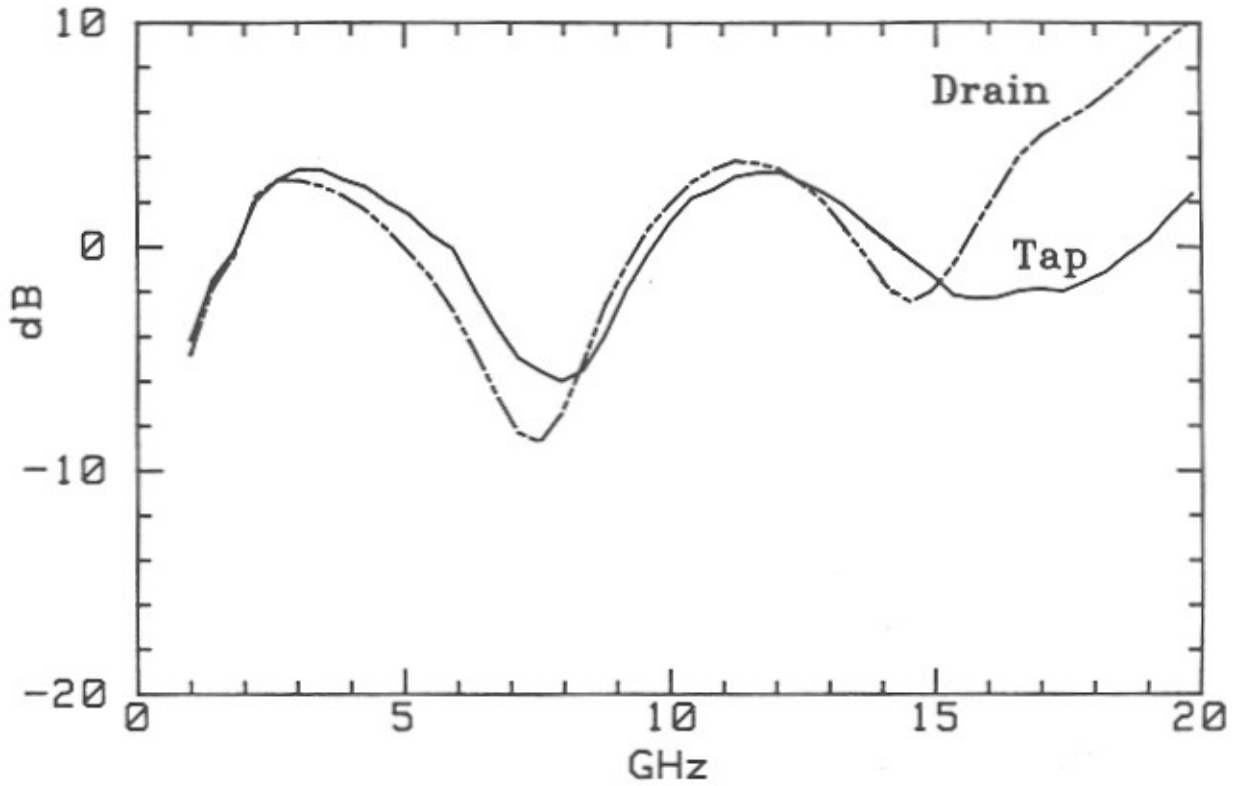


Figure 5.15: Voltage versus frequency at drain 2 and at drain 2 tap point, microstrip amplifier.

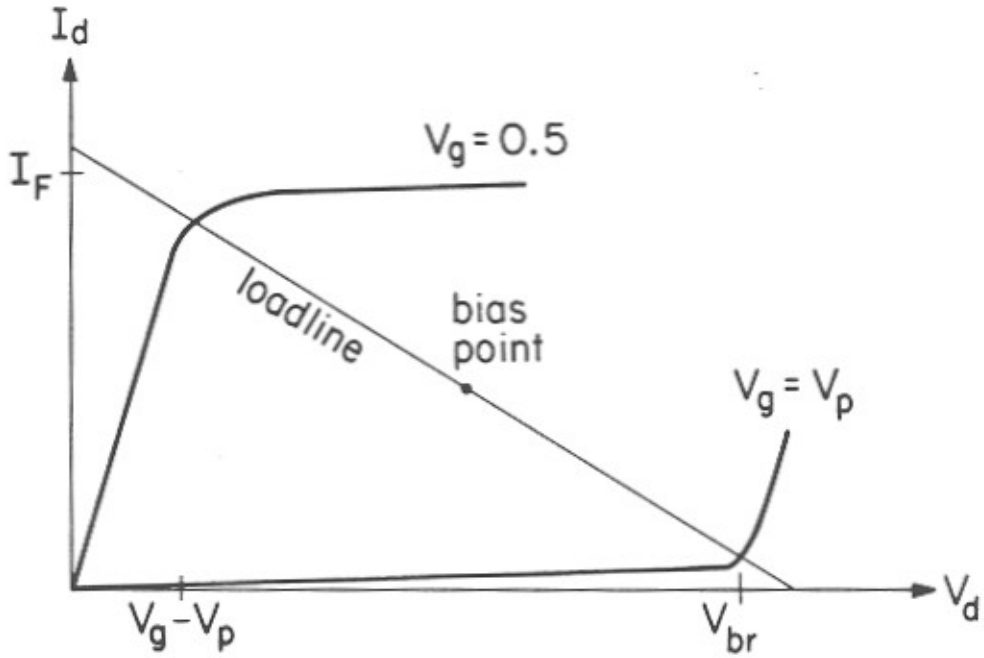


Figure 5.16: FET load-line for maximum output power.

$$V_d(z) = -ng_m Z_{load} V_{in} e^{-j\omega z/v_p} \ . \tag{5.25}$$

The drain-line voltage is uniform and opposite in phase to the gate-line voltage, allowing simultaneous saturation of all FET's and thus maximizing the output power at saturation.

In the uniform drain-line case, as is shown by Eq. 5.21 and by Figs. 5.12 and 5.13, the reverse wave on the drain line complicates the problem; the drain voltages are equal only at low frequencies, and the reverse wave introduces a phase shift between the gate and drain voltages of each FET. Thus, in the uniform drain line case, neither the conditions for simultaneous saturation of all FET's nor the conditions for simultaneously reaching all saturation mechanisms within a given FET can be met.

The 2-18 GHz microstrip amplifier has 1 dB gain compression at 7 dBm input power, and is not optimized for maximum power output; the lines are not tapered in the form of Eq. (5.23) and the bias is not set for maximum uncompressed output power. For a moment the frequency-dependence and position-dependence of the gate and drain voltages are ignored to estimate the gain compression point and identify the predominant saturation mechanism. With a gate bias voltage V_G of -0.3 V, a drain bias voltage V_D of 3.5 V, a pinchoff voltage V_P of approximately -2 V, and an amplifier voltage gain of $A_V = -2$ (6 dB), the gate signal is $\delta V_{gate} = \delta V_{drain}/A_V$, and the maximum negative voltage excursion is limited to $\delta V_{drain} = -(V_D - V_G - \delta V_{gate} + V_P) = -1.2$V corresponding to 11.5 dBm output power. With 6 dB amplifier gain, drain saturation will thus occur at input power levels of approximately 5.5 dBm, while the maximum input voltage before gate forward conduction is $\delta V_{gate} = (0.5 - V_G) = 0.8$ V, corresponding to 8 dBm input power. As pinchoff and drain breakdown occur only at still higher input powers, the maximum power output of the amplifier is limited by drain saturation. Probe measurements of the drain and gate large-signal voltage waveforms with the amplifier driven at its 1 dB gain compression power, show no clipping of the gate voltage waveforms (which would arise from gate forward conduction), but show clipping of the negative excursions of the drain voltage waveforms; the clipping resulting from drain saturation. Adjustment of the gate bias V_G to 0 V results in significant gate forward conduction at 7-dBm input power, and clipping of the positive excursions of the gate voltage waveforms was observed; subsequent test were performed with the amplifier biased normally ($V_G = -0.3$ V), where gain compression is predominantly drain saturation.

At 3 GHz, the small-signal voltages at the drains of the last 2 devices are approximately equal, and are larger than the small-signal voltages at drains 1, 2, and 3 (see Figs. 5.12 and 5.13). Thus, clipping occurs simultaneously at drains 4

and 5 (Fig. 5.17). Because of the smaller voltage swings at drains 1, 2, and 3, these devices show drain saturation only at input power levels several decibels larger than the input power necessary to cause drain saturation in the fourth and fifth FET's.

At 10 GHz, the distortion at the 1 dB gain compression point is complicated by phase shifts between the 10 GHz fundamental and the 20 GHz generated harmonic currents (Fig. 5.18) The 10 GHz small-signal voltage at drain 5 is 1.5 dB larger than that at drain 4; thus, FET 5 saturates more strongly. As with the 3 GHz saturation characteristics, the 10 GHz small-signal voltages at drains 1, 2, and 3 are comparatively small, and thus the first three transistors do not show substantial saturation at the 1 dB gain compression point. The 20 GHz harmonic current generated at drain 5 produces equal forward and reverse drain voltage waves at 20 GHz. With 10 ps loaded line delay between drains 4 and 5, the 20 GHz reverse wave undergoes 20 ps relative phase delay (which is 72^o of a 10-GHz cycle) before combining with the 10 GHz forward wave at drain 4. The resulting voltage waveform then approximates a sawtooth waveform; drain saturation at FET 4 then clips the peak negative excursion. Depending upon the line delay between successive drains, the reverse-propagating harmonic currents can either increase or decrease the peak voltages at other drains, increasing or decreasing the saturation at prior devices.

At 18 GHz, the small-signal drain voltage at drain 4 is larger than that at drain 5. Thus, at the 1 dB compression point, FET 4 will saturate more strongly than FET 5. At this frequency, FET's 2 and 3 have small-signal drain voltages that are 0.5-1 dB smaller than at drain 4, and thus also show significant drain saturation. The 36 GHz harmonic currents generated at drain 4 are beyond the cutoff of the synthetic drain line; the generated 36 GHz forward wave thus experiences significant dispersion and attenuation relative to the 18 GHz fundamental as the two waves propagate towards drain 5 and the output. The voltage waveform at drain 5 is thus visibly distorted, but the phase relationship between the fundamental and the distortion components has been shifted, and the waveform no longer clearly reflects the saturation waveform observed at drain 4 (Fig. 5.19)

5.8 Measurements of Digital Devices

The electrooptic sampling system has also been used to test a large number of digital integrated circuits.One early demonstration of on-wafer measurements of $\simeq$ 10 ps propagation delays was a series of measurements on a chain of 1 μm gate-length buffered-FET logic inverters. The inverters were in a cascaded chain, a standard test structure for measurement of average propagation delays. If the input

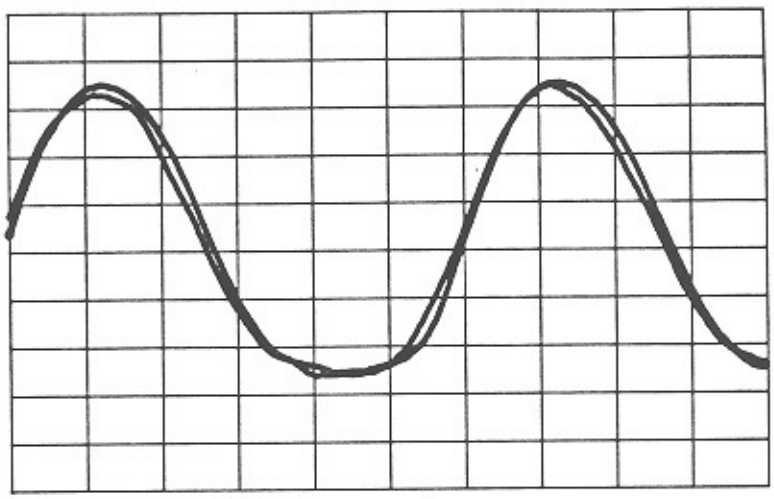


Figure 5.17: Saturation at drains 4 and 5 of microstrip TWA; 3 GHz, 7 dBm input power.

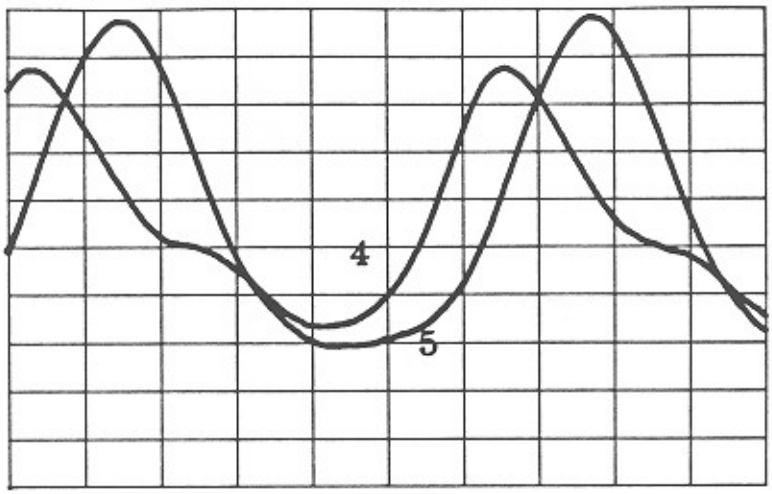


Figure 5.18: Saturation at drains 4 and 5 of the microstrip distributed amplifier at 10 GHz and 7 dBm input power, the 1 dB gain compression point.

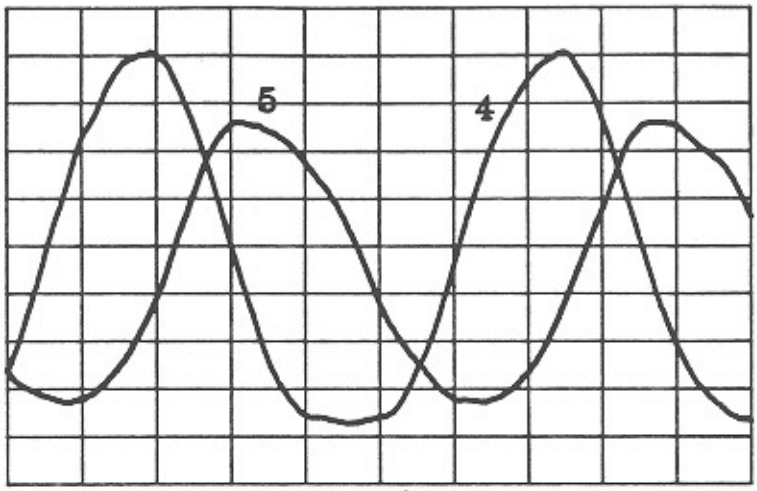


Figure 5.19: Saturation at drains 4 and 5 of the microstrip TWA; 18 GHz, 7 dBm input.

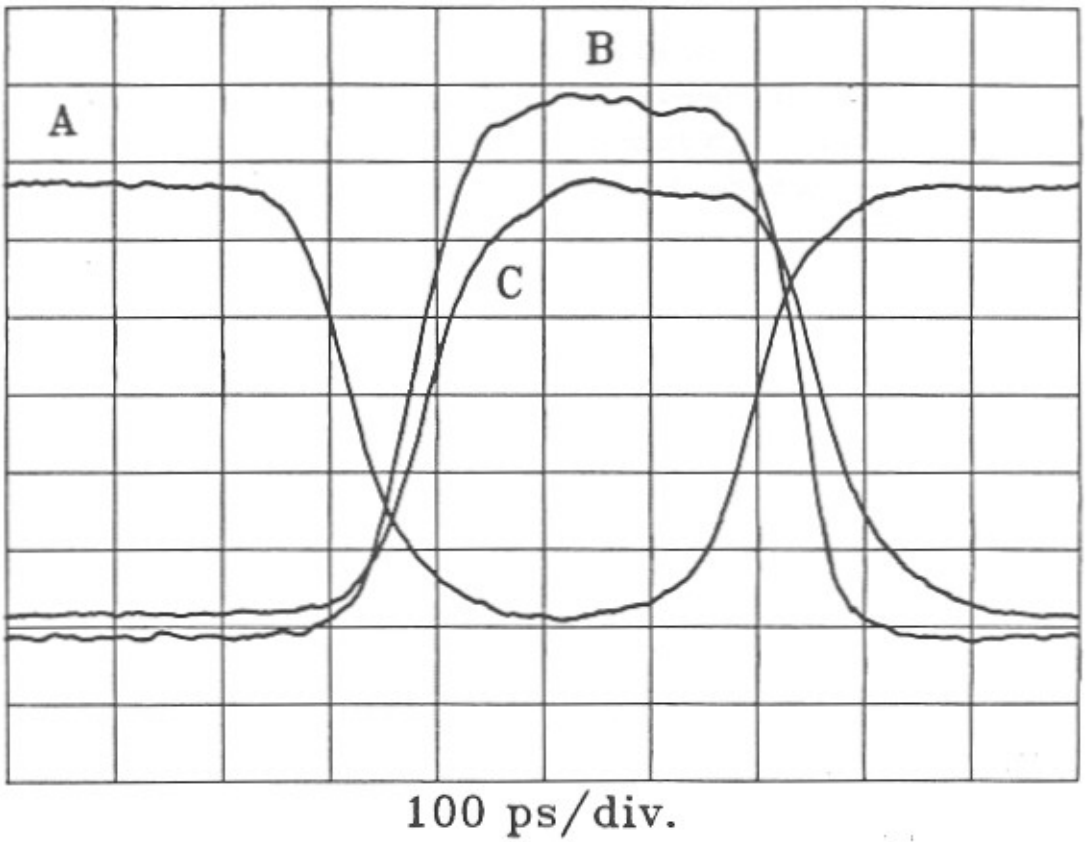


Figure 5.20: Voltage waveforms at the input (a), source-follower gate (b), and output (c) of GaAs buffered-FET-logic inverter gate [43].

to the chain is driven with a microwave signal, an alternating 1010... data sequence propagates through the the inverters. The time waveform, measured directly at individual gate nodes, allowed for direct measurement of the propagation delays and signal risetimes at gate nodes and nodes internal to single gates.

Figure 5.20 shows a gate delay measurement on the inverter chain. The delay between curves A and B of Fig. 15 is the propagation delay of the inverting common-source stage, 60 ps, while the delay between curves B and C is the delay of the source-follower buffer and diode level-shifter, 15 ps. The inverter chain from Lawrence Livermore National Labs [30] consisted of 20 gates each with a fan-in and fan-out of unity.

References.

[1] H.Q. Tserng and B. Kim: "110 GHz GaAs FET Oscillator," Elec. Lett., 21, 178, (1985)

[2] T. Henderson et. al.: "Microwave performance of a Quarter-Micrometer Gate Low-Noise Pseudomorphic InGaAs/AlGaAs Modulation-Doped Field Effect Transistor," IEEE Elect.Dev. Lett., EDL-7, (1986)

[3] T.C.L.G. Sollner, E.R. Brown, W.D. Goodhue, and H.Q. Le: "Observation of milli- meter-wave oscillations from resonant tunneling diodes and some theoretical considerations of ultimate frequency limits", Appl. Phys. Lett., 50 (6), 332, (1987)

[4] J.F. Jensen, L.G. Salmon, D.S. Deakin, and M.J. Delaney: "Ultra-high speed GaAs static frequency dividers," Technical Digest of the 1986 International Electron Device Meeting, p. 476

[5] John B. Gunn: "Sampling System Utilizing Electro-Optic Techniques", United States Patent 3,614,451 filed Aug. 19, 1968, Issued Oct. 19, 1971, Assignee International Business Machines Corporation, Armonk, N.Y.

[6] J.A. Valdmanis, G.A. Mourou, and C.W. Gabel: "Picosecond electro-optic sampling system," Appl. Phys. Lett., 41, 211, (1982)

[7] J.A. Valdmanis, G.A. Mourou, and C.W. Gabel: "Subpicosecond electrical sampling," IEEE J. Quantum Elect., QE-19, 664, (1983)

[8] B.H. Kolner, D.M. Bloom, and P.S. Cross: "Electro-optic sampling with picosecond resolution," Elect. Lett., 19, 574, (1983)

[9] B.H. Kolner and D.M. Bloom: "Electrooptic sampling in GaAs integrated circuits," IEEE J. Quant. Elect., QE-22, 79, (1986)

[10] K.J. Weingarten, M.J.W. Rodwell, and D.M. Bloom: "Picosecond Optical Sampling of GaAs Integrated Circuits", Invited paper to the special issue on ultrafast optoelectronics, IEEE Journal of Quantum Electronics, to be published February 1988.

[11] J.L. Freeman, S.K. Diamond, H. Fong, and D.M. Bloom: "Electrooptic sampling of planar digital integrated circuits," Appl. Phys. Lett., 47, 1083, (1985)

[12] M.J.W. Rodwell, K.J. Weingarten, D.M.Bloom, T. Baer, and B.H. Kolner:

"Reduction of timing fluctuations in a mode-locked Nd:YAG laser by electronic feedback," Optics Letters, 11, 638, (1986)

[13] M.J.W. Rodwell, K.J. Weingarten, J.L. Freeman, and D.M.Bloom: "Gate propagation delay and logic timing of GaAs integrated circuits measured by electro-optic sampling," Elect. Lett., 22, 499 (1986)

[14] M.J.W. Rodwell, M. Riaziat, K.J. Weingarten, B.A. Auld, and D.M. Bloom: "Internal microwave propagation and distortion characteristics of travelling-wave amplifiers studied by electro-optic sampling," IEEE Trans. Microwave Theory Tech., MTT-34, 1356, (1986)

[15] I.P. Kaminow, An Introduction to Electro-Optic Devices. New York: Academic Press, 1974

[16] I.P. Kaminow and E.H. Turner: "Electrooptic light modulators," Proc. IEEE, 54, 1374, (1966)

[17] B.H. Kolner and D.M. Bloom: "Direct electrooptic sampling of transmission-line signals propagating on a GaAs substrate," Elect. Lett., 20, 818, (1984)

[18] A. Yariv, P. Yeh: Optical Waves in Crystals, New York: New York: John Wiley & Sons, Inc, 1984, pp. 286-287

[19] Ibid, p. 230

[20] W.H. Knox, R.L. Fork, M.C. Downer, R.H. Stolen, and C.V. Shank: "Optical pulse compression to 8 fs at a 5-kHz repetition rate," Appl. Phys. Lett., 46, 1120, (1985)

[21] E.B. Treacy,."Optical pulse compression with diffraction gratings," IEEE J. Quant. Elect., QE-5, 454, (1969)

[22] D. Grischkowsky and A.C. Balant: "Optical pulse compression based on enhanced frequency chirping," Appl. Phys. Lett., 41, 1, (1982)

[23] J.D. Kafka, B.H. Kolner, T. Baer, and D.M. Bloom: "Compression of pulses from a continuous-wave mode-locked Nd:YAG laser," Opt. Lett., 9, 505, (1984)

[24] K.J. Weingarten, M.J.W. Rodwell, and D.M. Bloom: "Picosecond sampling of GaAs integrated circuits," To be published in Picosecond Electronics and Optoelectronics, New York: Springer-Verlag, 1987

[25] J.B. Beyer, S.N. Prasad, R.C. Becker, J.E. Nordman, and G.K. Hohenwarter,

"MESFET distributed amplifier design guidelines," IEEE Trans. Microwave Theory Tech., vol. MTT-32, pp. 268-275, Mar. 1984.

[26] M. Riaziat, I. Zubech, S. Bandy and G. Zdasuik, "Coplanar waveguides used in 2-18 GHz distributed amplifier" in Technical Digest, 1986 IEEE MTT-S Int. Microwave Symp., pp. 337- 338, 1986.

[27] Y. Ayasli, R. L. Mozzi, J.L. Vorhaus, L.D. Reynolds, and R.A. Pucel, "A Monolithic GaAs 1-13 GHz traveling-wave amplifier'" IEEE Trans. Microwave Theory Tech., vol. MTT-30, pp.976-981, July 1982.

[28] Y. Ayasli, L.D. Reynolds, R.L. Mozzi, and L.K. Hanes, "2-20 GHz GaAs traveling- wave power amplifier," IEEE Trans. Microwave Theory Tech., vol. MTT-32, pp. 290-295, March 1984.

[29] P.H. Ladbrooke, "Large-signal criteria for the design of GaAs FET distributed amplifiers," IEEE Trans. Electron Devices, vol. ED-32, pp. 1745-1748, Sept 1985.

[30] S. Swierkowski, K. Mayeda, and C. McGonaghy: "A sub-200 picosecond GaAs sample-and-hold circuit for a multi-gigasample/second integrated circuit," Technical Digest of the 1985 International Electron Devices Meeting, pp. 272-275

Chapter 6: Laser Timing Stabilization.

Mode-locked laser systems, in conjunction with optical pulse compressors generate extremely short optical pulses, making these instruments attractive for time-resolved measurements of a variety of fast phenomena, including electrooptic sampling, time-resolved carrier dynamics and electromagnetic transient measurements in semiconductors [1,2], measurement of optical waveforms by optical sampling [3], and time- resolved spectroscopy [4]. In these experiments, the system under test is pumped (excited) by the pulsed laser and subsequently probe by a delayed portion of the same optical pulse; pulse timing fluctuations are therefore unimportant.

In many experiments, laser pulse excitation of the system under test is either inconvenient or is an unrepresentative experimental condition. In electrooptic sampling (Chapter 5), the integrated circuit under test is best excited by microwave sinusoids or digital data sequences (as appropriate) from external electronic oscillators. The timing fluctuations of the laser degrade the time resolution and introduce noise.

For some stimulus-response laser experiments, it may be most desirable to excite the tested system with a pulse from one laser, and probe with a second laser having a different wavelength. Unless the two wavelengths are in an integer ratio (permitting synchronized pulse generation by by optical harmonic generation) the two optical pulse trains must be generated by two separate laser systems. Again, the relative timing fluctuations of the two lasers will both degrade the time resolution and introduce noise.

The performance of a variety of vacuum electron devices, including free electron lasers and laser-excited klystrons, is improved by increasing the cathode electron flux. Increased electron fluxes can be attained by replacement of the thermionic cathode with a photosensitive cathode excited by intense optical pulses from a pulsed laser system. If so excited, the frequency stability of these vacuum devices will be degraded by the timing fluctuations of the laser photoexcitation source. In particular, stringent phase stability constraints [5] must be placed upon the laser-excited klystrons now being developed for particle accelerators. Other potential applications appear to exist—initial publication of preliminary results of laser timing stabilization brought a number of reprint requests from researchers in the area of molecular and cellular biology.

6.1 Spectral Description of Amplitude and Timing Fluctuations

Fluctuations in mode-locked lasers include variations in both pulse intensity and pulse timing. These are random processes whose statistics are described by their power spectral densities. Before embarking on a description of the spectra of these random signals, first consider a single- frequency signal which is both amplitude-modulated and phase-modulated by small deterministic sinusoidal functions.

$$\begin{aligned} v_{am/pm}(t) = &\Big(1 + N_c \cos\omega_m t + N_s \sin\omega_m t\Big) \\ &\exp\Bigg(j\Big(\omega_l t + \theta_c \cos\omega_m t + \theta_s \sin\omega_m t\Big)\Bigg) \ . \end{aligned} \tag{6.1}$$

If the modulations N_c, N_s, θ_c and θ_s are small, the Fourier transform $V_{am/pm}(\omega)$ of $v_{am/pm}(t)$ is then given by

$$\begin{aligned} V_{am/pm}(\omega) = 2\pi\Bigg(&\delta(\omega - \omega_l) \\ &+ \Big(C_a + C_p\Big)\delta(\omega - \omega_l - \omega_m) \\ &+ \Big(C_a^* - C_p^*\Big)\delta(\omega - \omega_l + \omega_m)\Bigg) \ , \end{aligned} \tag{6.2}$$

where

$$C_a = N_c/2 - jN_s/2 \qquad C_p = \theta_s/2 + j\theta_c/2 \ . \tag{6.3}$$

The amplitude-modulation and phase-modulation of a signal can thus be determined from the Fourier spectrum. If the spectrum of a signal is given by

$$V_{am/pm}(\omega) = 2\pi\Bigg(\delta(\omega - \omega_l) + C_u\delta(\omega - \omega_l - \omega_m) + C_l\delta(\omega - \omega_l + \omega_m)\Bigg) \ , \tag{6.4}$$

where C_u,$C_l \ll 1$, then the time-domain signal has small amplitude and phase modulations, and is described by Eq. (6.1). The amplitude and phase modulation indices are then given by

$$\begin{aligned} N_c &= \Re\Big(C_u + C_l^*\Big) \qquad N_s = \Im\Big(C_u + C_l^*\Big) \\ \theta_c &= \Im\Big(C_u - C_l^*\Big) \qquad \theta_s = \Re\Big(C_u - C_l^*\Big) \ . \end{aligned} \tag{6.5}$$

The symbols $\Re(C)$ and $\Im(C)$ denote the real and imaginary parts of a complex number C. Thus, the upper and lower sidebands C_u and C_l can be decomposed into phase and amplitude noise sidebands C_p and C_a by the relationships

$$C_a = \frac{C_u + C_l^*}{2} \qquad C_p = \frac{C_u - C_l^*}{2} \ . \tag{6.6}$$

These relationships will be used later to determine the magnitude of phase measurement errors occurring within a timing stabilization system.

The laser produces a train of pulses, not a single frequency, and the amplitude and phase modulations are random. The laser pulse train is approximated as a series of Gaussian pulses with timing and amplitude fluctuations as given by

$$P(t) = \bar{P}T\Big(1+N(t)\Big) \times \sum_{n=-\infty}^{n=+\infty} \frac{1}{\sqrt{2\pi}\ \sigma_t} \exp\left(-\Big(t-T_0-nT_l-J(t)\Big)^2 \Big/ 2\sigma_t^2\right) , \tag{6.7}$$

where $P(t)$ is the laser intensity, $\bar{P}$ is the beam's average intensity, T_l is the pulse repetition period, σ_t is the rms pulse duration, $N(t)$ is the normalized pulse-intensity variation, $J(t)$ is the timing fluctuation of the pulse train, and T_0 is the static timing offset of the pulse train. The laser intensity then has a power spectral density $S_P(\omega)$ approximated to second order in $n\omega_l\sigma_J$ by

$$\begin{aligned} S_P(\omega) \simeq \bar{P}^2 e^{-\omega^2\sigma_t^2} \sum_{n=-\infty}^{n=+\infty} \Big(& (1 - n^2\omega_l^2\sigma_J^2) 2\pi\delta(\omega - n\omega_l) \\ & + (1 - n^2\omega_l^2\sigma_J^2) S_N(\omega - n\omega_l) \\ & + n^2\omega_l^2 S_J(\omega - n\omega_l) \Big) \ , \end{aligned} \tag{6.8}$$

where σ_J is the standard deviation of $J(t)$ (i.e. the rms timing jitter), $\omega_l = 2\pi/T_l$, $S_N(\omega)$ is the power spectral density of $N(t)$, and $S_J(\omega)$ is the power spectral density of $J(t)$.

Thus, the spectrum of the laser intensity is a series of lines at multiples of ω_l plus spectra associated with timing and amplitude fluctuations, referred to as amplitude noise sidebands $[S_N(\omega - n\omega_l)]$ and phase noise sidebands $[n^2\omega_l^2 S_J(\omega - n\omega_l)]$. The jitter attenuates the pulse train's spectrum at high frequencies, as indicated by the $(1 - n^2\omega_l^2\sigma_J^2)$ term, and degrades the rms time resolution $\sigma_{\text{experiment}}$ of probing experiments:

$$\sigma_{\text{experiment}} = \sqrt{\sigma_t^2 + \sigma_J^2} \ . \tag{6.9}$$

The phase-noise sidebands represent random fluctuations in the phase of the frequency components of the laser pulse train; in probing experiments using a pulsed laser imperfectly synchronized to a second source, the jitter introduces noise proportional to the time derivative of the experimental response. These phase-noise sidebands, having power proportional to n^2, predominate for harmonics of moderate order; at higher harmonics the expansion is inaccurate both because of higher-order terms in $n\omega_l\sigma_J$ and because of spectral terms associated with pulse-width fluctuations [ref]. Monitoring the laser with a photodiode and a spectrum analyzer, and measuring the relative powers of the laser harmonic and its associated sidebands at some harmonic of moderate order, $S_J(\omega)$ can be inferred. The total jitter can then be found:

$$\sigma_J = \sqrt{\frac{1}{2\pi}\int_{-\infty}^{+\infty} S_J(\omega)\ d\omega}\ . \tag{6.10}$$

Spectrum analyzers display the spectra of random signals as the noise power integrated over the analyzer resolution bandwidth. Radio-frequency spectrum analyzers use mixers, filters, and envelope detectors to determine a signal's spectrum, and consequently show small differences in calibration between measurements of deterministic and random signals. Correction factors associated with the difference between spectral densities expressed as a function of Hertz, rather than as a function of radian frequency, resolution bandwidths expressed by half-power bandwidth rather than by equivalent noise bandwidth, and noise measurement by noise envelope detection rather than by noise power detection must be included [6]. Characteristics and methods of measurement of phase noise are discussed in Robbins [6] and in two application notes from Hewlett-Packard [7,8].

The phase noise of several Nd:YAG lasers have been tested, and from this jitter (at rates of 50 Hz and above) in the range of 0.75-10 ps has been determined. Figures 6.1 and 6.2 show the measured phase noise on the 70^{th} harmonic of the 70 MHz pulse repetition rate of a Coherent, co. mode-locked ND:YAG laser. The phase-noise measurements were not digitally stored (the figures are hand tracings of the original data) and the phase noise thus could not be numerically integrated to determine the jitter. A ~ 0.75 ps rms timing jitter in the 50 Hz–5kHz frequency range is estimated.

6.2 Timing Stabilization by Feedback

The pulse-timing fluctuations of a mode-locked Nd:YAG laser can be reduced

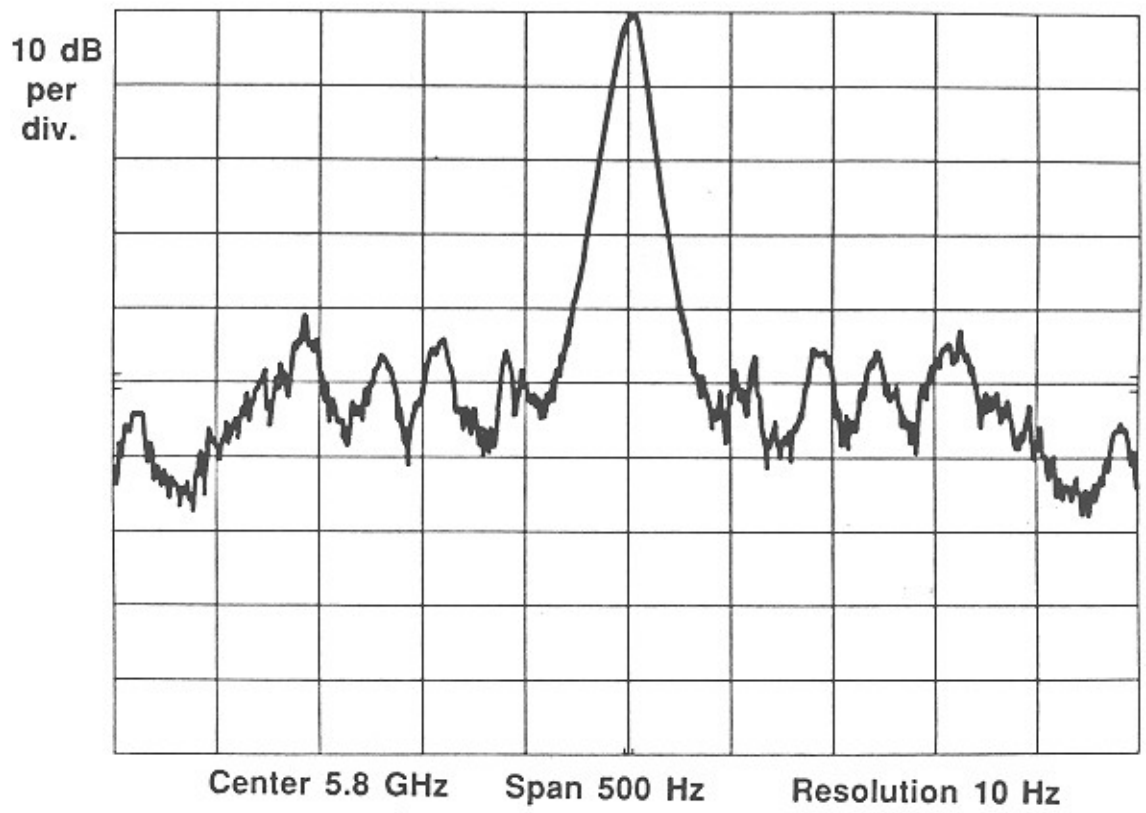


Figure 6.1: Phase noise sidebands, 50–250 Hz, on the 70^{th} harmonic of the 70 MHz laser pulse repetition rate; Coherent Antares mode-locked Nd:YAG laser.

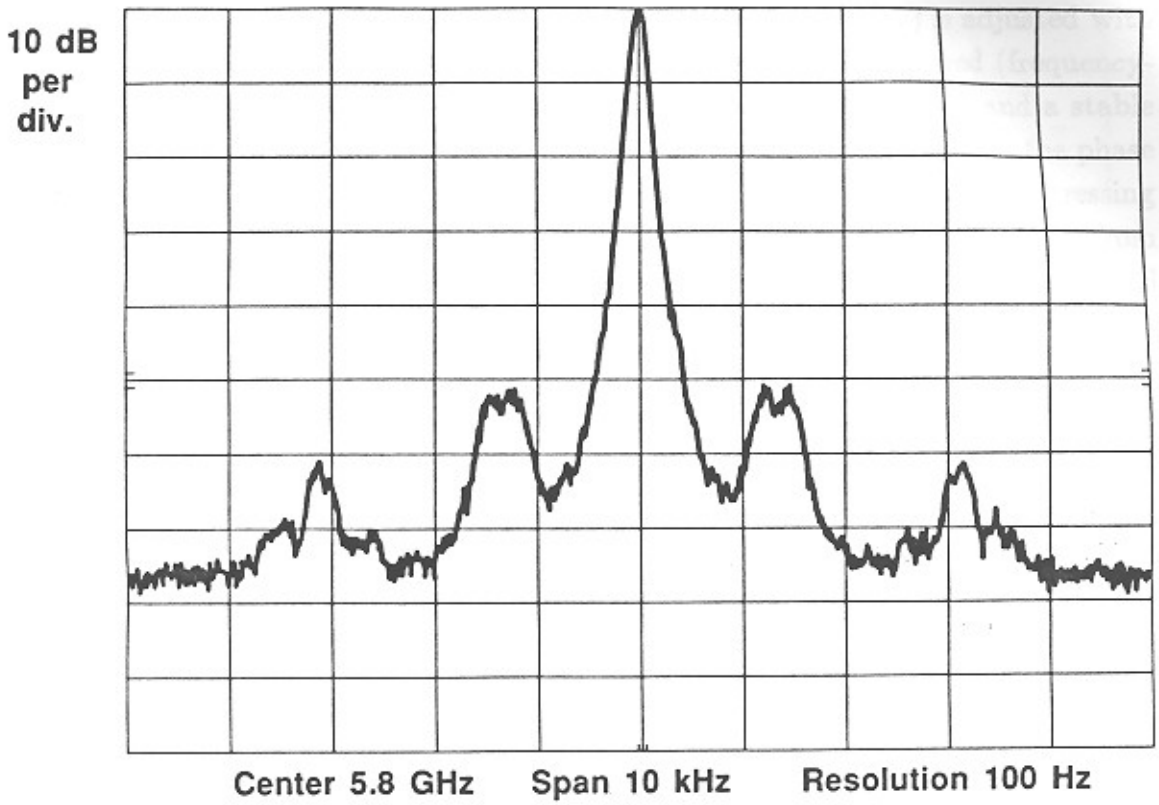


Figure 6.2: Phase noise sidebands, 500 Hz–5 kHz, on the 70^{th} harmonic of the 70 MHz laser pulse repetition rate. Coherent Antares laser.

by a phase-lock-loop feedback system, as first demonstrated by Cotter [9] (Fig. 6.3). A photodiode monitors the laser pulse train, and the phase of its fundamental component is compared to that of a very stable reference oscillator, generating a phase error signal. The signal driving the mode- locking acousto-optic (A-O) cell, whose frequency is half the laser pulse repetition frequency, is generated by frequency division from the reference oscillator; its timing (phase) is adjusted with a voltage-controlled phase shifter driven by the amplified and filtered (frequency-compensated) phase error signal. Given an error-free phase detector and a stable feedback loop with high loop transmission, the loop continuously adjusts the phase of the laser pulse train to equal the phase of the reference oscillator, suppressing the laser timing fluctuations. Jitter suppression is limited by spurious outputs from the phase detector, by the phase noise of the reference oscillator, and by the limited loop gain and limited loop bandwidth attainable in a stable, well-damped feedback control system.

Spurious outputs from the phase detector include both additive noise from the various electronic components in the feedback system, and inadvertent detection of amplitude noise by the phase detector. Becuase of the similarity of the latter problem to well-studied amplitude-noise sensitivity effects in phase-modulation and frequency-modulation radio systems [10], amplitude noise sensitivity can be referred to as AM-PM conversion.

6.2.1 AM-PM Conversion by Phase Detector DC Offset.

Radio-frequency phase detection is usually performed using a double- balanced mixer to form the product of the reference signal and the signal of unknown phase (Fig. 6.4). The multiplier output then contains terms both at DC and at twice the signal frequency; the latter terms are removed by loss- pass filtering. The output voltage of a multiplier (double balanced mixer) is then a function of both laser phase and amplitude,

$$V_{detector} = k_\theta \bar{P}[1 + N(t)]\sin(\theta) + V_{os} \ , \tag{6.11}$$

where $\theta = 2\pi[T_0 + J(t)]/T$ is the phase difference (for convenience, a constant phase factor of $\pi/2$ has been dropped) and V_{os} is the DC offset voltage of the phase detector arising from imbalances in the diodes and transformers [11]. The phase detector slope coefficient can be related to the mixers' RF-IF conversion efficiency A_{M1} (here in units of Volts/Volt) by $k_\theta \bar{P} = A_{M1} V_{in}$. The peak voltage V_{in} of the sinusoidal input voltage is in turn given by $V_{in} = A_{A1}\sqrt{2 P_{detector} Z_0}$, where A_{A1} is the voltage gain of amplifier $A1$ and $P_{detector}$ is the power of the fundamental

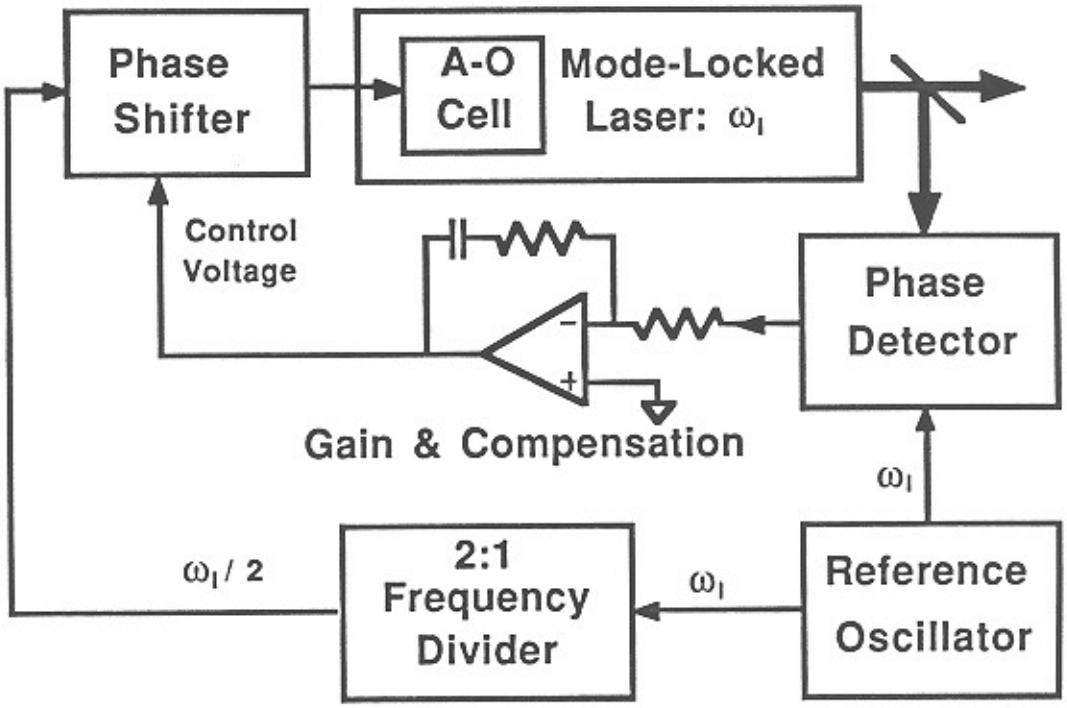


Figure 6.3: Phase-lock-loop laser timing stabilization.

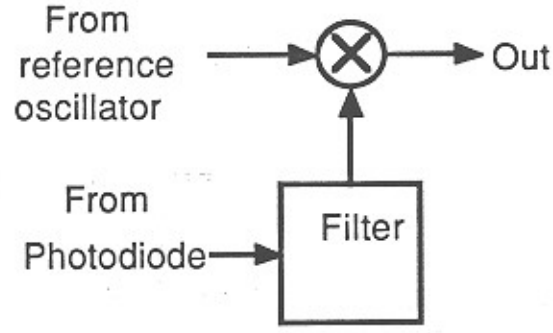


Figure 6.4: Simple phase detector using a double-balanced mixer. The filter passes only the fundamental component at ω_l of the laser pulse train.

component of the photodiode output, as measured in a system with impedance Z_0. The impedance Z_0 is usually 50 Ω. The quantity $\sqrt{2P_{detector}\ Z_0}$ is the peak voltage of the sinusoidal fundamental component of the photodiode signal.

For small θ, the detector output is approximately

$$V_{detector} \simeq k\bar{P}\Big([1 + N(t)]T_0 + J(t) \Big) + V_{os} \ , \tag{6.12}$$

where $k = 2\pi k_\theta / T$. The terms in $N(t)$ in the above equation result in sensitivity of the timing stabilization system to laser amplitude noise. To assess this effect, assume that we have an ideal (unrealizable) feedback loop having unbounded loop gain at all frequencies. The detector output is then forced to zero, and the laser timing becomes

$$\begin{aligned} T_0 &= -V_{os}/k\bar{P} \\ J(t) &= T_0 N(t) \ . \end{aligned} \tag{6.13}$$

Thus, in addition to causing timing drift with changes in mixer offset ($T_0 = -V_{os}/k\bar{P}$), the DC offset results in first-order variation in the phase detector output with laser amplitude fluctuations, thus producing residual closed-loop timing jitter.

An improved phase detector (Fig. 6.5) suppresses the effect of the mixer offset by increasing the slope coefficient k of the detector. The technique is similar to that of chopper-stabilized DC amplifiers and synchronous ("lock-in") signal detectors. The reference oscillator output is multiplied in a double-balanced-mixer by a squarewave chopping signal whose frequency is ω_{IF}. After narrowband filtration to remove spurious mixer outputs, the resulting reference signal is then a sinewave at the laser repetition frequency ω_l whose phase is inverted $+180^0/-0^0$ at the chopping rate ω_{IF}. The fundamental component at ω_l of the photodiode signal is passed to a second mixer, in which the laser phase is compared to that of the phase-modulated reference signal. The output of the second mixer is then a spurious DC offset voltage $V_{os,2}$, plus a squarewave at frequency ω_{IF} whose amplitude varies as the phase difference of the photodiode and reference signals. The squarewave is bandpass filtered, removing both the spurious V_{os} and components at harmonics of ω_l, and amplified by some large gain A_{A2} on the order of 20–40 dB. The phase error signal, now a sinewave, is synchronously detected by a third double-balanced mixer, producing an output voltage proportional to the laser phase. As with the simple phase detector, the chopper-stabilized detector has sensitivity to laser amplitude fluctuations and has a DC offset V'_{os} arising from the third mixer:

$$V'_{detector} \simeq k'\bar{P}\Big([1 + N(t)]T_0 + J(t) \Big) + V'_{os} \ . \tag{6.14}$$

In the chopper-stabilized detector, the phase-detector gain coefficient is given by

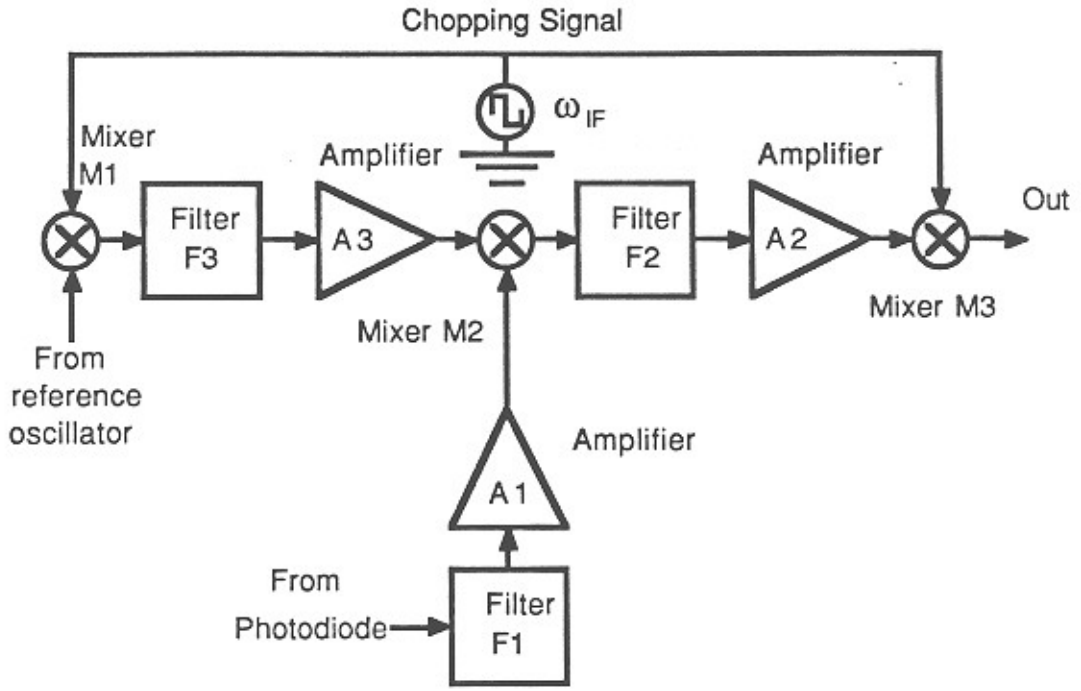


Figure 6.5: Chopper-stabilized phase detector. See text for description of operation. F3 is a filter with passband $\sim \omega_l \pm 5\omega_{IF}$, F1 is a filter with passband $\sim \omega_l \pm \omega_l/2$, and F2 is a filter with passband $\sim \omega_{IF} \pm \omega_{IF}/2$.

$$k' = \sqrt{2P_{detector}\ Z_0}\ A_{A1}\left(\omega_l\right)A_{M2}A_{A2}\left(\frac{8}{\pi^2}\right)A_{M2}\ , \tag{6.15}$$

where A_{A2} is the voltage gain of amplifier $A2$ and A_{M2} is the RF-IF conversion loss (in units of Volts/Volts) of mixer $M2$. The factor of ω_l results from the conversion from phase sensitivity coefficient to timing sensitivity coefficient, and the factor of $8/\pi^2$ is a consequence of chopping with a square-wave signal but subsequently synchronously detecting only its sinusoidal fundamental component. With the chopper-stabilized phase detector, DC offset will result in timing offset and timing jitter given by

$$\begin{aligned} T_0' &= -V_{os,M3}/k'\bar{P} \\ J(t) &= T_0' N(t)\ , \end{aligned} \tag{6.16}$$

where $V_{os,M3}$ is the DC offset voltage of mixer $M3$. Amplitude-noise to phase noise conversion through nonlinear terms in the operation of mixer $M2$ place upper bounds on $\sqrt{2P_{detector}Z_0}\ A_{A1}$, and thus the chopper-stabilized phase detector per-

mits a phase- detector sensitivity k' several orders of magnitude larger than with a simple phase detector; the effect of DC offset is proportionately reduced.

With the continued assumption of infinite loop gain, the spectral density $S_{J,os}(\omega)$ of spurious timing fluctuations resulting from detector DC offset will be

$$S_{J,os}(\omega) = T_0^2 S_N(\omega) \ . \tag{6.17}$$

6.2.2 AM-PM Conversion through Device Saturation

The simplest method of attaining large phase detector sensitivity is to provide large input signal levels to the detector. Problems with DC offsets are reduced, as are the effects of additive noise voltages and currents within the feedback loop. Above some power level, nonlinear effects within both the amplifiers processing the 82 MHz photodiode signal and the mixer used for phase detection will result in AM-PM conversion.

The response of a weakly nonlinear system can be described in the time domain by a Volterra series expansion [10]:

$$\begin{aligned} v_{out}(t) &= \sum_{n=1}^{n=+\infty} \left(\int \int \cdots \int h_n(\tau_1, \tau_2, \ldots, \tau_n) \prod_{i=1}^{i=n} v_{in}(t-\tau_i) d\tau_i \right) \\ &= \sum_{n=1}^{n=+\infty} v_{out,n}(t) \ . \end{aligned} \tag{6.18}$$

Denoting a series of n-dimensional Fourier transforms:

$$V_{out}(\omega) = \int v_{out}(t) e^{-j\omega t} \ dt \ ,$$

$$V_{out,n}(\omega) = \int v_{out,n}(t) e^{-j\omega t} \ dt \ ,$$

$$V_{in}(\omega) = \int v_{in}(t) e^{-j\omega t} \ dt \ ,$$

$$h_n(\tau_1, \tau_2, \ldots, \tau_n) = \frac{1}{(2\pi)^n} \int \int \cdots \int H(\Omega_1, \Omega_2, \ldots, \Omega_n) \left(\prod_{i=1}^{i=n} e^{j\Omega_i \tau_i} d\Omega_i \right) \ ,$$

the Fourier transform $V_{out}(\omega)$ of the output voltage can be written as

$$V_{out}(\omega) = \sum_{n=1}^{n=+\infty} V_{out,n}(\omega) \tag{6.19}$$

where

$$\begin{aligned} V_{out,n}(\omega) = \frac{1}{(2\pi)^{n-1}} \int\int\cdots\int H(\Omega_1, \Omega_2, \ldots, \Omega_{n-1}, \omega - \sum_{i=1}^{i=n-1} \Omega_i) \\ V_{in}(\Omega_1)V_{in}(\Omega_2)\cdots V_{in}(\Omega_{n-1})V_{in}(\omega - \sum_{i=1}^{i=n-1} \Omega_i) \\ d\Omega_1 d\Omega_2 \cdots d\Omega_{n-1} \ . \end{aligned} \tag{6.20}$$

If $v_{in}(t)$ is small (and this is required in order to have adequately small AM-PM conversion), the series can be truncated to third order with sufficient accuracy. Further, the second-order term results in nonlinear responses at DC and twice the input frequency; as these will be outside of the bandwidth of various filters within the system, their effect can be neglected. The nonlinear response of amplifiers and mixers within the timing stabilizer is then modeled in the frequency domain as

$$\begin{aligned} V_{out}(\omega) \simeq & H_1(\omega)V_{in}(\omega) \\ & + \frac{1}{4\pi^2} \int\int_{-\infty}^{+\infty} H_3(\Omega_1, \Omega_2, \omega - \Omega_1 - \Omega_2) \\ & \qquad V_{in}(\Omega_1)V_{in}(\Omega_2)V_{in}(\omega - \Omega_1 - \Omega_2) d\Omega_1 d\Omega_2 \ . \end{aligned} \tag{6.21}$$

For a system with general third-order nonlinearities, the above equation cannot be further reduced. For the case of an input signal at ω_l with narrowband amplitude modulation (i.e. $\omega_m \ll \omega_l$) the terms H_1 and H_3 vary little over the frequency range from $(\omega_l - \omega_m)$ to $(\omega_l + \omega_m)$ and

$$H_1(\pm\omega_l \pm \omega_m) \simeq H_1(\pm\omega_l)$$

$$H_3(\pm\omega_l \pm \omega_m, \pm\omega_l \pm \omega_m, \pm\omega_l \pm \omega_m) \simeq H_3(\pm\omega_l, \pm\omega_l, \pm\omega_l) \ . \tag{6.22}$$

Symmetries in its definition require that $H_3(\alpha, \beta, \gamma)$ be invariant with permutations in the the order of (α, β, γ); we can thus denote:

$$H_3(\omega_l) = H_3(\omega_l, \omega_l, -\omega_l) \text{ and all permutations}$$

$$H_3(-\omega_l) = H_3(-\omega_l, -\omega_l, \omega_l) \text{ and all permutations}$$

Finally, $H_3(-\omega_l) = H_3^*(\omega_l)$ and $H_1(-\omega_l) = H_1^*(\omega_l)$ are required for systems with real response. For compactness, we will write $H_3 = H_3(\omega_l)$ and $H_1 = H_1(\omega_l)$.

Consider the response of a third-order nonlinear system to an amplitude- modulated signal $v_{in}(t)$

$$v_{in}(t) = v_0\Big(1 + N_C \cos\omega_m t\Big)\cos\omega_l t \tag{6.23}$$

The Fourier transform of $v_{in}(t)$ is

$$\begin{aligned} V_{in}(\omega) = \pi v_0 &\bigg(\delta(\omega-\omega_l) + \delta(\omega-\omega_l)\bigg) \\ &+ \frac{\pi v_0 N_c}{2}\bigg(\delta(\omega-\omega_l-\omega_m) + \delta(\omega-\omega_l+\omega_m)\bigg) \\ &+ \frac{\pi v_0 N_c}{2}\bigg(\delta(\omega+\omega_l-\omega_m) + \delta(\omega+\omega_l+\omega_m)\bigg)\ . \end{aligned} \tag{6.24}$$

Substitution of Eq. (6.24) into (6.21) and calculation of the resulting phase modulation from the phase of the spurious third-order products using Eq. (6.5) shows that the output voltage $v_{out}(t)$ is phase-modulated by $\theta_c \cos\omega_m t$, where

$$\frac{\theta_c}{N_c} = \frac{9}{4} v_0^2 \Im\left(\frac{H_3}{H_1}\right)\ . \tag{6.25}$$

The AM-PM conversion factor $\Im(H_3/H_1)$ is rarely specified for an electronic component. An upper bound can be found given knowledge of the commonly-specified (input-referred) two-tone third-order intercept power P_{3oi}. This is

$$P_{3oi} = \left\|\frac{H_1}{H_3}\right\|\frac{2}{3Z_0}\ , \tag{6.26}$$

where Z_0 is the system impedance, usually 50Ω. The spurious phase modulation can thus be bounded by

$$\left\|\frac{\theta_c}{N_c}\right\| \le \frac{9v_0^2}{4}\left\|\frac{H_3}{H_1}\right\| = \frac{3v_0^2}{2P_{3oi}Z_0}\ . \tag{6.27}$$

The spurious phase modulation occurring in any components processing the photodiode signal prior to phase detection will result in an erroneous phase error signal at the output of the phase detector. With the approximation of large loop gain, the output of the phase detector will be again forced to zero, and residual

timing fluctuations will thus result from AM-PM conversion in the phase detector. If the power spectral density of the residual closed-loop timing jitter due to AM-PM conversion through 3^{rd}–order nonlinearities is denoted as $S_{J,3o}(\omega)$, it is found that

$$S_{J,3o}(\omega) \leq \frac{9}{\omega_l^2}\left(\frac{P_{\text{carrier}}}{P_{3oi}}\right)^2 S_N(\omega) \ , \tag{6.28}$$

where $P_{\text{carrier}} = v_0^2/2Z_0$ is the signal power at the frequency ω_l input to the device under consideration.

6.2.3 Additive Noise in the Phase Detection System

For timing jitter below ∼1 ps with a laser whose repetition rate is ∼ 100 MHz, the phase noise sidebands will have very low power. For those components which process the photodiode signal prior to phase detection, the component's additive noise will result in spurious signal sidebands. From the analysis of Section 6.1, these additive sidebands can be decomposed into spurious phase-noise and amplitude-noise sidebands of equal power; the additive noise for these components thus results in a spurious output at the phase detector. The equivalent input noise power of the components processing the pre-phase-detection photodiode signal (i.e. amplifier $A1$ and mixer $M2$) is

$$S_{RF}(\omega) = \frac{kT}{2}\left(F_{A1} + \frac{F_{M2}-1}{A_{A1}^2}\right) \ , \tag{6.29}$$

where F_{A1} and F_{M2} are the noise figures of the amplifier and the mixer. The thermal noise spectral density is $kT/2$ rather than kT because we are using two-sided spectral densities with frequency in the units of radians/second. With large loop transmission, these noise terms will result in a residual closed-loop timing jitter with spectral density

$$S_{J,RF}(\omega) = \frac{S_{RF}(\omega-\omega_l)}{\omega_l^2 P_{photodiode}} \ . \tag{6.30}$$

Components with low noise figures must be used, and the signal levels should be as high as permitted by the limits imposed by third-order nonlinearities.

At the input to the amplifier chain following the phase-detector $M2$, the phase error signal is in the form of the amplitude of a sinewave at the chopping frequency:

$$v(t) = \sqrt{2Z_0}\alpha J(t)\sin\omega_{IF}t \ , \tag{6.31}$$

where

$$\sqrt{2Z_0}\alpha = \sqrt{2P_{photodiode}Z_0}A_{A1}A_{M1}\omega_l\frac{4}{\pi} \ , \tag{6.32}$$

and ω_{IF} is the chopping (synchronous detection) frequency. The equivalent input noise power of the chopping-frequency (intermediate-frequency) components (amplifier $A2$ and mixer $M3$) is

$$S_{IF}(\omega) = \frac{kT}{2}\left(F_{A2} - 1 + \frac{F_{M3}-1}{A_{A2}^2}\right) \ . \tag{6.33}$$

Noise of the intermediate-frequency components will result in residual closed-loop timing jitter with spectral density given by

$$S_{J,IF}(\omega) \simeq \frac{S_{IF}(\omega - \omega_{IF})}{\alpha^2} \ . \tag{6.34}$$

Alternatively, the noise of the intermediate-frequency components can be combined with the noise of the radio-frequency components (i.e. the components processing the photodiode signal prior to phase detection by mixer $M2$). The equivalent input noise power of the RF and IF systems can then be expressed as

$$\begin{aligned} S_{RF/IF}(\omega) =& \frac{kT}{2}\left(F_{A1} + \frac{F_{M2}-1}{A_{A1}^2} + \frac{F_{A2}-1}{A_{A1}^2A_{M1}^2(4/\pi)^2} + \frac{F_{M3}-1}{A_{A1}^2A_{M1}^2(4/\pi)^2A_{A2}^2}\right) \\ =& \frac{kT}{2}F_{equivalent} \ , \end{aligned} \tag{6.35}$$

where $F_{equivalent}$ is an equivalent noise figure for the RF and IF systems. The residual closed-loop timing jitter due to additive noise in both the RF and IF sections is then

$$S_{J,RF}(\omega) + S_{J,IF}(\omega) = \frac{S_{RF/IF}(\omega - \omega_l)}{\omega_l^2 P_{photodiode}} \ . \tag{6.36}$$

Finally, the gain and compensation section has some equivalent input noise voltage with spectral density $S_{amp}(\omega)$, which will result in residual closed-loop jitter having a power spectral density

$$S_{J,amp}(\omega) = \frac{S_{amp}(\omega)}{k'^2} \ . \tag{6.37}$$

To suppress these last two residual noise terms, the signal levels (i.e. the gain factors α and k') should be large, and the components should be selected for low equivalent input noise.

6.2.4 Reference Oscillator Phase Noise

A final error term in the phase detection system is the phase noise (timing jitter) of the reference oscillator itself; in an ideal timing stabilization system, the laser phase will track that of the reference oscillator. The phase noise of a free-running oscillator varies in proportion to the noise of the amplification device coupled to the oscillator's resonator, and the phase noise bandwidth is set by the bandwidth of the resonator and the degree of coupling between resonator and amplifier [12]. The phase noise bandwidth is on the order of the resonator bandwidth. High-stability electronic oscillators use quartz crystal resonators whose quality factor Q is on the order of 10^5. Moderate-cost quartz crystal oscillators at $\omega \simeq 2\pi\times$ 100 MHz are available with phase noise spectral densities ($= \omega_l^2 S_J(\omega)$) on the order of -130 dBc (1 Hz) at Δf =100 Hz, -145 dBc (1 Hz) at Δf =1 kHz [13].

If the oscillation frequency must be tuned to accommodate variations in the laser cavity frequency ω_l , then a simple crystal oscillator cannot be used. Frequency synthesizers, systems in which a broadband tunable oscillator is phase-locked to a precision fixed-frequency quartz oscillator, can then be used for the reference oscillator. Synthesizers having low phase-noise are expensive ($\sim$ \$ 30,000).

6.2.5 Loop Bandwidth and Stability

Given a phase detector with negligible spurious outputs, the timing jitter suppression is set by the feedback loop transmission $G(\omega)$:

$$S_{closed\ loop}(\omega) = \left\| \frac{1}{1+G(\omega)} \right\|^2 S_{open\ loop}(\omega) \ , \tag{6.38}$$

where the loop transmission is given by

$$G(\omega) = k' A_{comp}(\omega) k_s H_L(\omega) \ . \tag{6.39}$$

In the above equation, $A_{comp}(\omega)$ is the transfer function of the amplifier providing loop gain and compensation, k_s is the slope coefficient of the voltage controlled phase shifter (i.e. $d\tau/dV_{control}$, where τ is the delay and $V_{control}$ is the control voltage), and $H_L(\omega)$ is the phase transfer function of the laser.

The design of the loop gain and compensation $A_{comp}(\omega)$ to attain the desired loop characteristics is the subject of classical (frequency- domain) control system theory, and is discussed extensively in D'Azzo and Houpis [14]. Large $G(\omega)$, desirable at all frequencies at which $S_{open\ loop}(\omega)$ has significant energy, is limited by stability constraints arising from $H_L(\omega)$. If the feedback loop is opened (this

is achieved in practice by leaving the loop active but decreasing the loop bandwidth to ~ 1 Hz) and the phase of the laser driving signal is perturbed by applying a step-function drive signal to the voltage-controlled phase shifter, then the output of the phase detector will show the transient behavior of the laser timing in response to a step-function phase shift. The result of this measurement on the Spectra-Physics Nd:YAG laser is shown in Fig. 6.6. This step response conforms well to $h_l(t) = u(t)(1 - e^{-t/\tau_p})$, with $\tau_p = 50\mu$s, and where $u(t)$ denotes the unit step-function. The laser phase transfer function is therefore

$$H_L(\omega) \simeq \frac{1}{1 + j\omega/\omega_p} \ , \tag{6.40}$$

where $\omega_p = 1/\tau_p$; for our Spectra-Physics Nd:YAG laser, $\omega_p \simeq 2\pi \times 3.2$ kHz.

As with any feedback loop, the loop bandwidth (the frequency at which $\|G\| = 1$) must be constrained. The laser phase transient response may have additional poles in its transfer function at frequencies well above 3.2 kHz, whose influences on a time-domain measurement are masked by the response of the dominant pole, and other parasitic poles exist in the transfer functions of the phase shifter, phase detection system, and loop compensation section. Simple feedback systems commonly use an integration in the feedback loop. In this case,

$$A_{comp}(\omega) = \frac{\omega_c}{j\omega} \ , \tag{6.41}$$

and the loop suppression is given by

$$R(\omega) \equiv \frac{1}{1 + G(\omega)} = \frac{j\omega(2\zeta/\omega_n) - \omega^2/\omega_n^2}{1 + j\omega(2\zeta/\omega_n) - \omega^2/\omega_n^2} \ , \tag{6.42}$$

where the loop natural frequency ω_n and the loop damping ζ are given by

$$\omega_n^2 = k' k_s \omega_c \omega_p \qquad \zeta = \frac{1}{2}\sqrt{\frac{\omega_p}{k' k_s \omega_c}} \ . \tag{6.43}$$

If an attempt is made to attain large loop gain and large loop bandwidth by choosing a large gain coefficient $k' k_s \omega_c$, the loop damping factor ζ will be degraded. For $\zeta < 1/\sqrt{2}$, the loop suppression will be *greater than* unity in the vicinity of ω_n, and thus the spectral density $S_J(\omega)$ of the timing jitter will be *increased* for $\omega \approx \omega_n$. For a well-damped control system without peaking of $S_J(\omega)$ near ω_n, the loop damping ζ must be greater than $1/\sqrt{2}$ and the loop natural frequency is then bounded: $\omega_n \leq \omega_p/\sqrt{2}$. Thus, in a simple control system using integrating feedback, the laser pole frequency limits the suppression of timing jitter to those frequency components below $\omega_p/\sqrt{2}$. Even within this bandwidth, the loop rejection is finite,

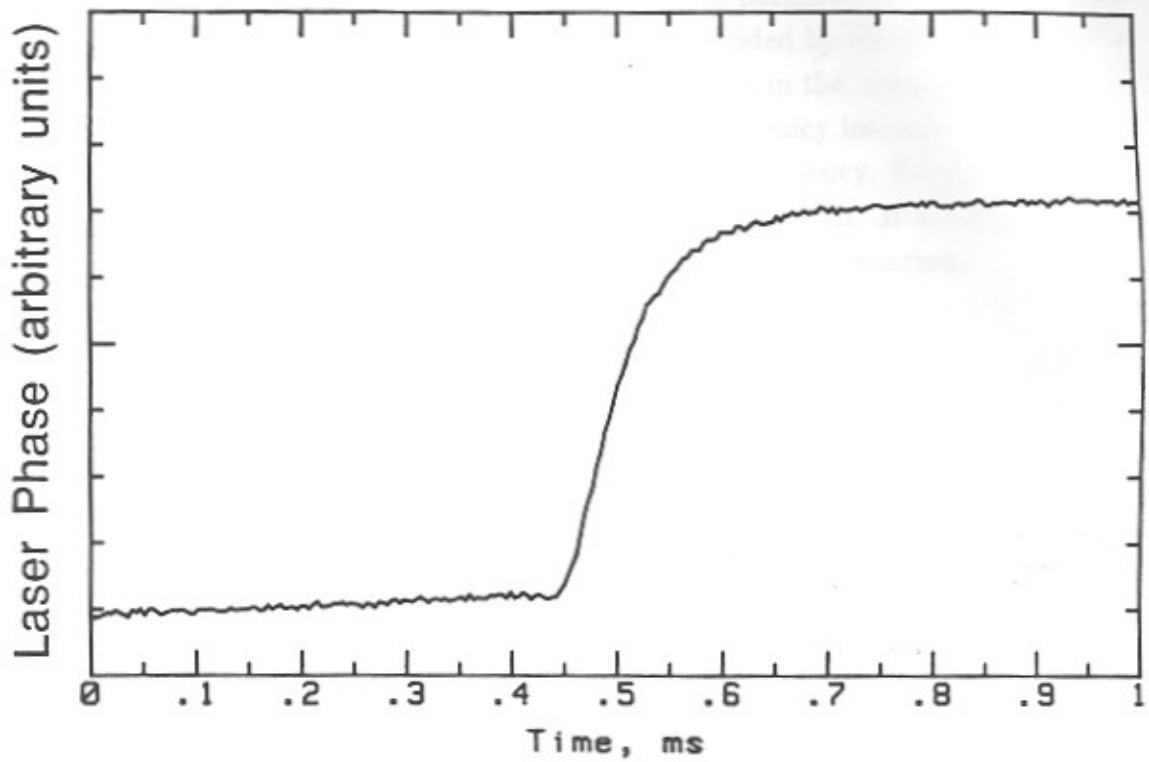


Figure 6.6: Transient response of the phase of a Spectra- Physics laser to a step change in the phase of the mode-locking drive signal. The response is that of a first-order dynamic system, and corresponds to a 3.2 kHz pole.

and is given by the loop rejection Eq. (6.42). Unfortunately, $S_J(\omega)$ has significant energy at frequencies approaching the laser pole frequency ω_p.

Well-damped control loops with larger loop bandwidth can be attained by a variety of compensation techniques, including lead compensators [14] and observer-controller compensators [15]. Given the variations in parameters found in experimental systems, the choice of compensator design is guided by the sensitivity of the system damping and system stability margin to changes in the loop parameters. In particular, it has been observed that the laser pole frequency increases as the laser pulse repetition rate is detuned from the laser cavity frequency. Simple lead compensation will permit a moderate increase in loop bandwidth. If a compensating zero is included in the transfer function of the gain and compensation,

$$A_{comp}(\omega) = \frac{\omega_c}{j\omega}(1 + j\omega/\omega_p) = \frac{\omega_c}{j\omega} + \frac{\omega_c}{\omega_p} \ , \tag{6.44}$$

then the loop suppression becomes

$$R(\omega) = \frac{j\omega/\omega_{loop}}{1 + j\omega/\omega_{loop}} \quad \text{where} \quad \omega_{loop} = k' k_s \omega_c \ . \tag{6.45}$$

The loop suppression is that of a first-order system, and is well- damped even for loop bandwidths ω_{loop} exceeding ω_p.

6.3 Experimental Results

The design guidelines outlined above have been used to implement a feedback system to reduce the timing jitter of a 82 MHz mode-locked Nd:YAG laser. The current design has evolved over the past three years, with early published results [16] being a reduction from 2.9 ps unstabilized to 0.9 ps stabilized jitter; the current system reduces the laser timing fluctuations at rates greater than 50 Hz from ∼1.25 ps unstabilized to ∼0.25 ps stabilized. The decrease with time of the unstabilized jitter deserves some comment: in this period, the laser acousto-optic mode-locker and the end mirrors were replaced. These two changes should have resulted in both a narrower cavity bandwidth and stronger injection-locking, both of which reduce the phase noise; replacement of the two components resulted in the open-loop jitter decreasing from ∼2.9 to ∼1.25 ps (for rates above 50 Hz). An unstabilized jitter of 10 ps was first measured by Brian Kolner [17], while my first measurements on the same laser system were more consistent with 2.9 ps. As unstabilized jitter appears to increase with both laser misalignment and with deviations of the mode-locking

frequency from the natural laser cavity frequency, perhaps the improvement was due to more careful laser operation as we became aware of its effect on jitter.

The laser stabilized is an 82 MHz Spectra-Physics Nd:YAG unit mode-locked by an acousto- optic cell. The amplitude noise sidebands are approximately - 85 dB with respect to the carrier, at ∼100 Hz offset from the carrier, as measured in a 1 Hz resolution bandwidth, i.e. $S_N \simeq -85$ dBc (1 Hz) at $\Delta f = 100$ Hz. The unstabilized phase noise sidebands have significant spectral density up to a frequency of ∼2 kHz, and a loop bandwidth of ∼5 kHz is thus required for significant suppression of the strong phase-noise components at 1–2 kHz (Eq. 6.45). Given this estimate of the loop bandwidth, if timing fluctuations of ∼ 100–200 fs are to be attained, phase noise spectral density must be on the order of - 125 to -130 dBc (1 Hz) at frequencies within the ∼5 kHz loop bandwidth. The spurious phase-noise sidebands due to DC offset, 3^{rd}- order nonlinearities, and additive noise must be well below the objective stabilized phase noise sidebands; the system is designed so that each spurious effect results in sidebands with power less than -135 dBc (1 Hz).

The timing stabilization system has been implemented following the general design guidelines outlined in the previous sections. The phase detection system is chopper-stabilized as in Section 6.2.1, with chopping at 1 MHz; the phase detector slope k' is 1 Volt/ns, and the mixer offset is less than 0.5 mV, resulting in a static timing offset T_o of less than 5 ps. The resulting AM-PM conversion due to DC offset is less than -50 dB; thus the laser's -85 dBc (1 Hz) amplitude noise results in less than -135 dBc (1 Hz) spurious closed-loop phase noise through AM-PM conversion by detector DC offset.

With amplitude noise sidebands some 50 dB more powerful than the objective phase noise sidebands, AM-PM conversion through third-order nonlinearities must be suppressed by at least 50 dB; using Eq. (6.25), the input power levels to the amplifier $A1$ and to the phase detector (mixer) $M2$ must be at least 30 dB below the third-order intercept points of each component. In the current implementation, the input power levels are greater than 35 dB below the 3^{rd}-order intercept power, and the suppression of AM-PM conversion is at least 60 dB, resulting in less than - 145 dBc (1 Hz) spurious closed-loop phase noise through AM-PM conversion by 3^{rd}-order products.

The photodiode power $P_{photodiode}$ at the laser repetition frequency ω_l is -24 dBm, while the equivalent noise figure $F_{equivalent}$ of the RF and IF systems is 9 dB, and kT is -173 dBm (1 Hz). By Eq. 6.36, the additive noise of these subsystems thus will result in a spurious phase noise spectral density of -140 dBc (1 Hz).

Finally, the reference oscillator for the timing stabilization system is a Hewlett-Packard model 8662A low phase-noise signal generator. No phase- noise specifica-

tionis available for this synthesizer operating at 82 MHz, but over the 320–640 MHz range, the phase noise specification is -104 dBc (1 Hz) at 100 Hz offset, -121 dBc (1 Hz) at 1 kHz offset, and -131 dBc (1 Hz) at 10 kHz offset. Extrapolation to 82 MHz is uncertain, but it is likely that the oscillator's timing deviations are relatively independent of oscillation frequency, in which case phase noise will scale as the square of oscillation frequency. The phase noise of the HP8662A at 82 MHz is then estimated as - 116 dBc (1 Hz) at 100 Hz offset, -133 dBc (1 Hz) at 1 kHz offset, and -143 dBc (1 Hz) at 10 kHz offset. Given both the small difference between the estimated reference phase noise and the objective stabilized laser phase noise, and the uncertainty of our estimate, this data will have to be verified if the laser phase noise is to be further reduced.

The control loop bandwidth is set at approximately 6 kHz; the control loop is lead compensated, with a compensating zero at 3.2 kHz cancelling the response of the 3.2 kHz laser pole. Due to variability in the laser pole frequency, with our laser the compensation does not consistently provide good loop damping if the loop bandwidth is increased much beyond 6 kHz. The bandwidth and damping of the control loop can be determined by measuring the suppression of a step-function error signal injected into the loop (Fig. 6.7). The response of the system to an injected error signal is a simple exponential decay, with no evidence of an underdamped second-order system response. In contrast, Fig. (6.8) shows a similar measurement of an earlier laser timing stabilizer in which the lead compensation was incorrectly set; the response contains both an exponentially decaying response and a exponentially decaying oscillatory response, and is characteristic of an underdamped 3^{rd}-order control system.

The suppression of phase noise by the feedback loop is shown as a series of three phase noise measurements (Figs. 6.9–6.11). In the narrowband 500 Hz span (Fig. 6.9) 10–20 dB suppression is seen. Some sidebands at harmonics of the 60 Hz power line frequency are observed. The phase noise on the 82 MHz fundamental can be calculated from this; -113 dBc (1 Hz) at 100 Hz offset. In this bandwidth, the phase noise is very close to the estimate of the reference oscillator phase noise. Indeed, the measured phase noise may be that of the local oscillator in the spectrum analyzer used for these measurements, as the spectrum analyzer likely has a poorer local oscillator than the low-phase-noise oscillator from the same manufacturer. These points need to be investigated.

In the broader 10 kHz frequency span (Fig. 6.10), the phase noise is most likely that of the laser. Broad spectral peaks at 1.2 kHz are suppressed by approximately 16 dB, while some shoulders on the spectral density at 2 kHz are suppressed by 10 dB. These two spectral features might be ascribed to mechanical resonances

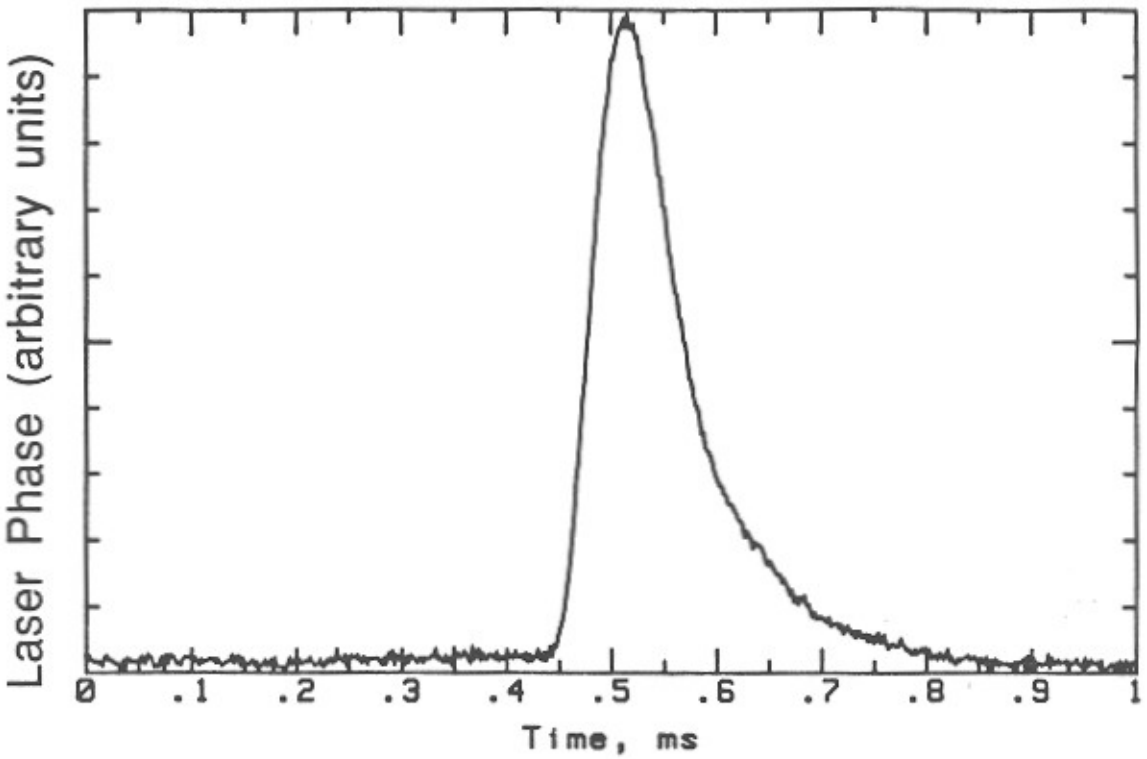


Figure 6.7: Suppression of a step-function error signal injected into the phase-lock loop. The exponentially decaying response is indicative of a stable control loop with large damping.

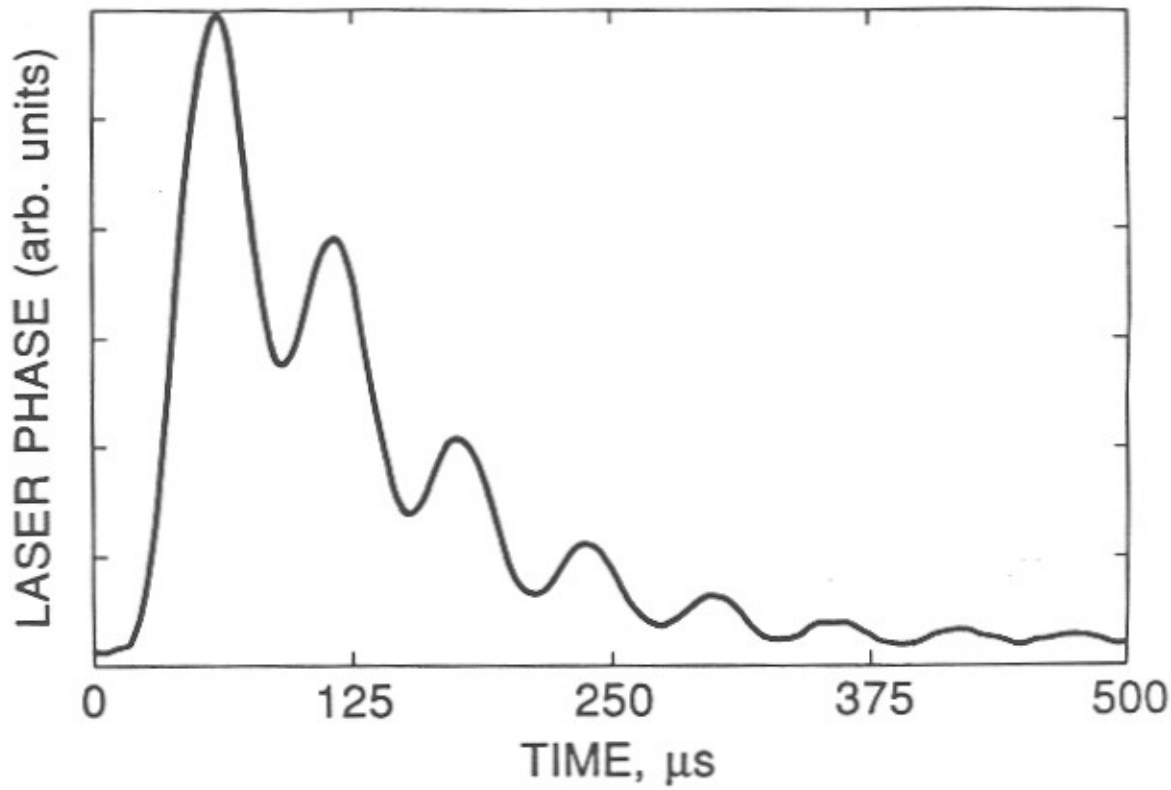


Figure 6.8: Suppression of an injected error signal by a previous implementation of the laser timing stabilizer, in which the lead compensation was incorrectly set. The system has 3^{rd}–order dynamics, and the response contains both an exponentially decaying component and a damped oscillatory component. The stability margin of this system was small.

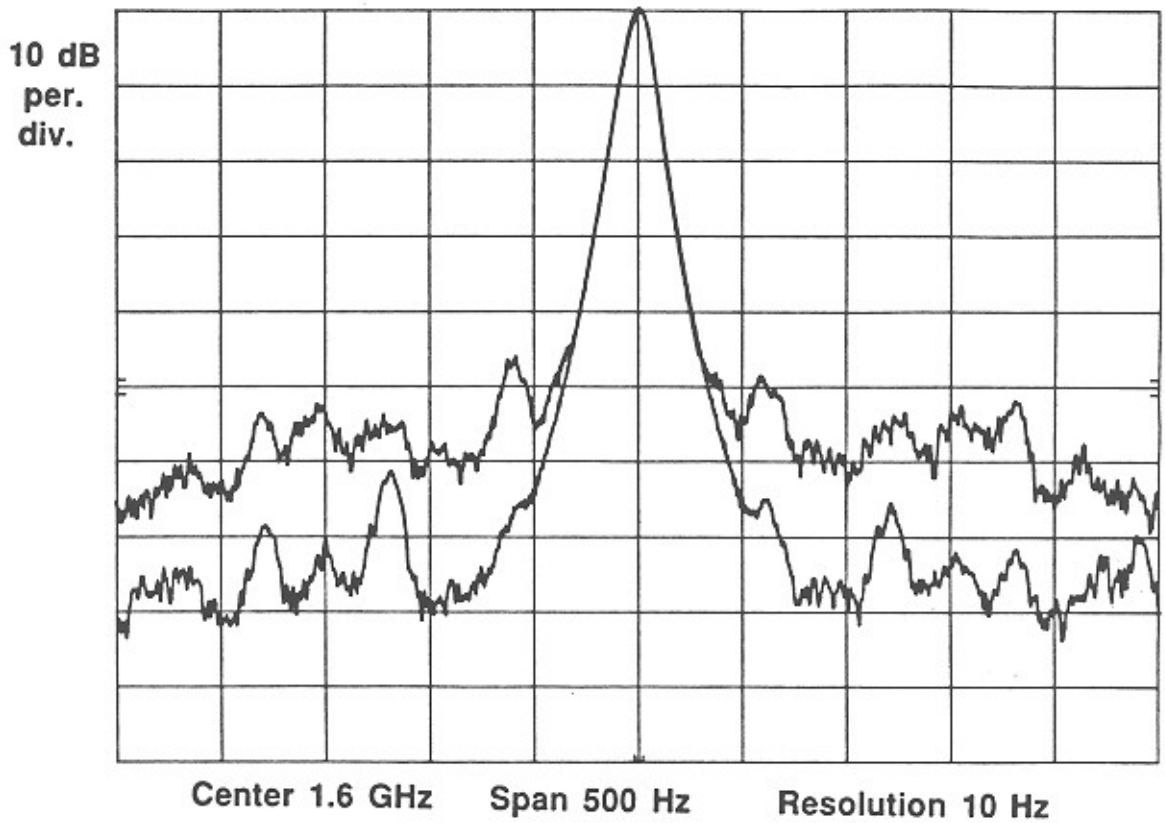


Figure 6.9: Suppression of the 50–250 Hz phase-noise sidebands of an 82 MHz Spectra-Physics mode-locked Nd:YAG laser, as measured by spectum analysis of the 20^{th} harmonic of the laser pulse repetition rate.

within the laser mirror supports. The low-energy point of the sidebands is at $\sim$ 2.5 kHz; the phase noise spectral density at this point corresponds to -131 dBc (1 Hz) phase noise on the 82 MHz fundamental. It is likely that this corresponds to the combined spurious effects of reference oscillator phase noise, AM-PM conversion through DC offset, additive noise in the RF and IF systems, and additive noise ($\sim$ 20 dB noise figure) of the photodiode, amplifier, and spectrum analyzer used for these measurements. Of these effects, spectrum analyzer phase noise and additive photodiode, amplifier, and spectrum analyzer noise are not components of the laser phase noise, but are measurement error. For this reason, at offsets less than 2 kHz, we see strong suppression of the laser's open-loop phase noise sidebands consistent with a $\sim$ 6–7 kHz loop bandwidth ω_{loop}, but see little phase-noise suppression at offsets from 2–5 kHz, consistent with a background phase noise floor of $\sim$-131 dBc (1 Hz).

For completeness, the phase noise spectrum in a very broad 50 kHz span is also shown (Fig. 6.11). The stabilization system causes a slight increase in the spectrum at frequencies from $\sim$6–10 kHz, with little difference thereafter. The phase noise at 10 kHz offset corresponds to -131 dBc (1 Hz) on the 82 MHz fundamental. At offsets greater than 15 kHz the sidebands are below the additive noise of the instrumentation system, and the phase noise cannot be determined. Improved measurements will require an instrumentation system with improved noise figure in the $\sim$ 2–6 GHz microwave frequency range.

After integrating these phase noise spectral densities, the timing jitter of the laser is determined to be:

$$\sigma_J = \begin{cases} 1.2 \text{ ps}, & \text{open loop, 50 Hz–25 kHz;} \\ 0.24 \text{ ps}, & \text{closed loop, 50 Hz–25 kHz.} \end{cases}$$

The reduction in phase noise is 5:1 over the frequency range 50 Hz–25 kHz. Frequency components above 25 kHz should be measured, but the trend of Figs. (6.9–6.11) shows that the noise is rapidly decreasing above 5 kHz. The unstabilized phase noise components below 50 Hz have large energy, and should be measured; unfortunately, the resolution of the spectrum analyzer prohibits measurement of these components. Phase noise test sets can measure these very low-frequency components, but the instrument cost ($\sim$ \$300,000) is well beyond the laboratory resources. Further, the AM rejection of these instruments (the Hewlett- Packard 11729B carrier noise test set has a specified AM rejection of 20 dB) may be inadequate unless measurements are made at high harmonics of the laser pulse repetition rate. Observing the timing fluctuations of the laser pulses on a sampling oscilloscope, the unstabilized jitter appears to be on the order of 20 ps at low frequencies, while the

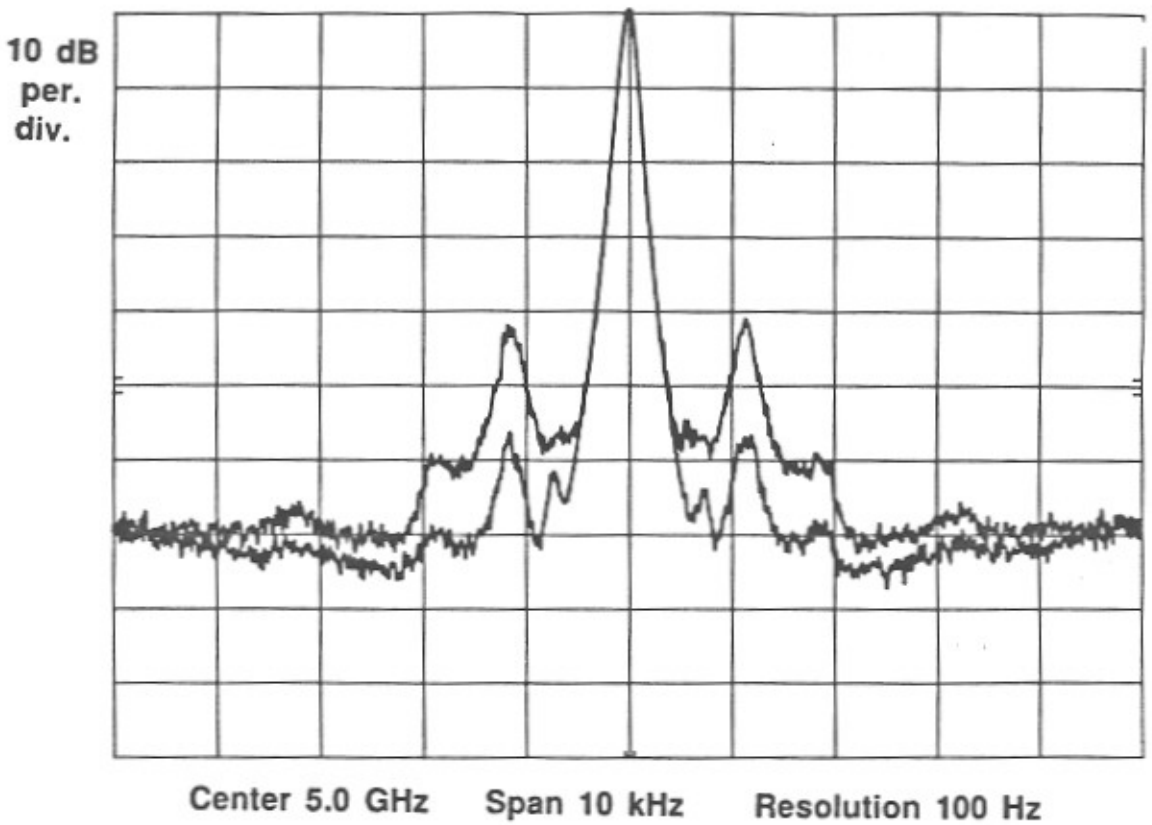


Figure 6.10: Suppression of the 500–5 kHz phase-noise sidebands of the Spectra-Physics laser, measured on the 60^{th} harmonic of the laser pulse repetition rate.

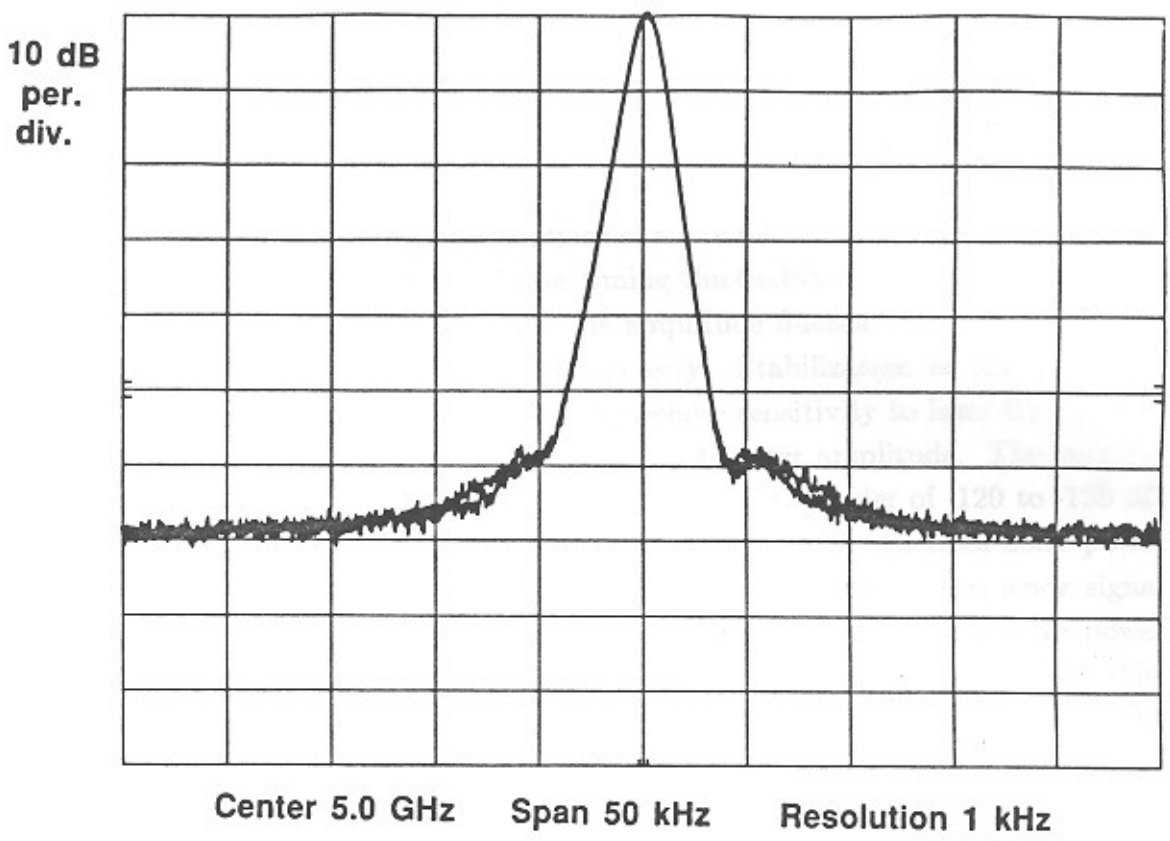


Figure 6.11: Suppression of the 5 kHz–25 kHz phase-noise sidebands of the Spectra-Physics laser, measured on the 60^{th} harmonic of the laser pulse repetition rate.

the stabilized jitter is below the timing fluctuations ($\sim$ 5 ps) of the sampling oscilloscope. Phase noise components below 50 Hz clearly need further investigation.

6.4 Conclusions

The (50 Hz-25 kHz) timing fluctuations of a mode-locked laser has been reduced by 5:1, attaining 0.24 ps jitter. These timing fluctuations are 2 parts in 10^5 of the laser pulse repetition period, while the amplitude fluctuations of the laser are approximately 5% of the average laser intensity. Stabilization of the timing to these levels requires a precise phase detector whose sensitivity to laser timing is at least a factor of 10^4 greater than its sensitivity to laser amplitude. The spectral density of the stabilized phase noise sidebands are on the order of -120 to -135 dB (1 Hz) with respect to the 82 MHz laser fundamental; as the thermal noise power power has spectral density $kT = -173$ dBm (1 Hz), the radio- frequency signal levels must be selected so that the associated phase-noise sidebands have power well above kTF, where F is the noise figure of the phase detection system. The bandwidth of the unstabilized phase noise is significant in comparison with the bandwidth of the laser's dynamic phase response, and the feedback loop must be appropriately compensated if the loop is not to show instabilities or weakly damped resonances with associated peaking of the phase noise. With these results, the laser phase noise is now within an order of magnitude of the phase noise of high- stability radio-frequency sources. Further reduction of the jitter will require decreased offset, decreased noise figure, and decreased 3^{rd}- order products in the phase detection system, increased loop bandwidth or more complex (lead-lag) compensation, and perhaps a reference source with further improved stability (it is likely that only a fixed-frequency crystal oscillator will provide this).

Instrumentation is also a difficulty; the noise figure and the residual phase noise of the photodetector, amplifier, and spectrum analyzer measurement system are becoming a significant fraction of the measured sideband power. Precision carrier noise test sets may permit measurements with lower residual phase noise, and will permit measurements of phase noise at offsets below 50 Hz. Unfortunately, the cost of such systems is beyond the resources of most scientific laser laboratories, and their AM rejection is unestablished.

Despite these difficulties, no fundamental limits have been reached, and with continued refinement jitter below 0.1 ps is expected.

References.

[1] J. Bokor, R. Haight, J. Stark, R.H. Storz, R.R. Freeman, and P.H. Bucksbaum: "Picosecond Time-Resolved Photoemission Study of the InP (100) Surface", in *Picosecond Electronics and Optoelectronics*, G.A. Mourou, D.M. Bloom, and C.-H. Lee, eds., Springer-Verlag, New York, 1985.

[2] D.H. Auston, K.P. Cheung, J.A. Valdmanis, and D.A. Kleinman, Phys. Rev. Lett., 53, 1555-1558 (1984).

[3] J.M. Weisenfeld, R.S. Tucker, P.M. Downey, and J.E. Bowers, Electron Lett, 22, 396-397, 1986.

[4] J.A. Kash, J.C. Tsang, and J.M. Hvam, "Subpicosecond Raman Spectroscopy of Electron-LO Phonon Dynamics in GaAs", in *Picosecond Electronics and Optoelectronics*, G.A. Mourou, D.M. Bloom, and C.-H. Lee, eds., Springer-Verlag, New York, 1985.

[5] J.D. Fox and H. Schwarz, in *Digest of the 1982 IEEE Symposium on Microwave Theory and Techniques*, pp. 334–336.

[6] W.P. Robbins, *Phase Noise in Signal Sources*, IEEE Telecommunications Series, No. 9, Peter Peregrinus, London, 1982.

[7] *Phase Noise Characterization of Microwave Oscillators, Phase Detector Method*, Product note 11729B-1, Hewlett-Packard Co., P.O. Box 10301, Palo Alto, CA. 94303-0890

[8] *Phase Noise Characterization of Microwave Oscillators, Frequency Discriminator Method*, Product note 11729C-2, Hewlett-Packard Co., P.O. Box 10301, Palo Alto, CA. 94303-0890

[9] D. Cotter: "Technique for Highly Stable Active Mode-Locking", in *Ultrafast Phenomena IV*, D. A. Auston and K. B. Eisenthal, eds., Springer- Verlag, New York, 1984.

[10] D. R. Jordan and P.L. Penny, eds.: *Transmission Systems for Communications, Fifth Edition*, Bell Laboratories, 1982. See especially Chapters 17 and 24.

[11] S.R. Kurtz: "Mixers as Phase Detectors" Watkins-Johnson Company Technical Notes, vol.5, no. 1, January/February 1978. Watkins-Johnson Company, Palo Alto, CA.

[12] A. van der Ziel: *Noise in Solid State Devices and Circuits*, Wiley-Interscience,

New York, 1986. See especially Chapter 10.

[13] Data on high-stability crystal oscillators are available from a number of manufacturers. One large supplier is Vectron Laboratories, Inc. 166 Glover Avenue, Norwalk, Connecticut 06850.

[14] J. J. D'Azzo and C. H. Houpis: *Linear control system analysis and design: conventional and modern*, McGraw-Hill, Inc. 1975.

[15] T.M. Kailath: *Linear Systems*, Prentice-Hall, 1980.

[16] M. J. W. Rodwell, K. J. Weingarten, D. M. Bloom, T. Baer, and B. H. Kolner: "Reduction of timing fluctuations in a mode-locked Nd:YAG laser by electronic feedback", Optics Letters, vol. 11, page 638, October 1986.

[17] B. H. Kolner and D. M. Bloom: "Electrooptic Sampling in GaAs Integrated Circuits", IEEE J. Quant. Elect., vol. QE-22, 79-93, 1986.

Chapter 7: Future Directions

We have reviewed both the development of picosecond electrical pulse generators and the picosecond optoelectronic instrumentation used to characterize them. The results with the first-generation nonlinear transmission lines (Chapter 4) demonstrate wavefront compression, with generated transition times which are a factor of 2–3 better than those reported for tunnel diodes and step-recovery diodes. But, recalling the predictions of Chapter 1, falltimes below 1 picosecond should be feasible; in this context, 8 ps is a poor result. The 8 ps minimum compressed falltime of the first-generation devices is primarily due to the high (10 Ω) diode series resistance. Further, in contrast to compression of step-functions, where signal attenuation occurs only from transmission-line series resistance, diode series resistance introduces strong attenuation during compression of sinusoidal input signals. As a consequence of the smaller wavefront magnitude at the line output, longer minimum compressed falltimes should be observed with sinusoidal excitation.

The devices described within the thesis were fabricated with a 10 μm minimum feature size. As the fabrication process has been refined, smaller feature sizes and hence higher diode cutoff frequencies and higher periodic- line cutoff frequencies have become possible. With diode series resistance identified as the predominant limitation in the first devices, improved devices fabricated to 3.5–5 μm design rules were recently designed and fabricated by Chris Madden. In these improved devices, the diode cutoff frequency has been increased from 300 GHz to approximately 1 THz, while the periodic-line cutoff frequency is either 140 GHz (as in the first design) or 280 GHz. The line lengths in these designs are increased sufficiently for compression to the asymptotic falltime given a 6 V, $\sim$35 input transition. Given a step-function input, SPICE simulations predict compressed falltimes of 3–4 picoseconds. In initial evaluation of these devices, 5–7 ps falltimes have been attained.

Shorter measured transition times will require improvements in the excitation signal used in device testing, in the electrooptic sampling system, and in the device itself. To date, the nonlinear transmission lines have been evaluated only by observing compression of sinusoidal input signals. With step-function excitation from a step-recovery diode, signal attenuation due to diode resistance is eliminated, and shorter compressed falltimes should be attained over a larger range of device operating conditions. The fastest (35 ps) step-recovery diodes (SRD's) are available only as device die and not as packaged step-function generators. In designing an SRD drive circuit and fixture for 35 ps step-function generation, there are considerable

difficulties with undesired structure on the generated waveform (pre-edge pedestal, post-edge "dribble-up"). An additional difficulty is AM-PM conversion; the 1–10 MHz pulse modulation used for $1/f$ noise suppression in the electrooptic sampling system results in significant phase modulation of the SRD output. Possibly, the pre-transition and post-transition artifacts on the SRD output can be suppressed by diode clamping, while the difficulties with AM-PM conversion can be avoided by employing alternate signal-processing methods.

The 5 ps wavefronts generated by the improved devices are the shortest electrical transients yet measured on the electrooptic sampling system. While a $\sim$ 100 GHz system bandwidth has been estimated from measurements of optical pulse autocorrelation and phase noise, the time resolution of the sampling system may be introducing substantial error in the measurement of 5 ps transients. While the optical pulse duration is $\sim$ 1.5 ps, as measured at the 50% points, the optical pulse autocorrelation exhibits measurable wings. If a significant fraction, on the order of 10%, of the optical pulse intensity lies within these wings, then the 10%-90% risetime of the electrooptic sampling system will be significantly larger than 2 picoseconds. While laser timing fluctuations are less than 1/4 picosecond at the laser output, several effects could potentially lead to larger jitter at the compressor output, and hence poorer system time resolution. Finally, the lateral dimensions of the nonlinear transmission line are $\sim$ 60 μm; if the interconnecting transmission lines are probed, the optical transit time will be approximately 3 picoseconds. While the probe beam was passed directly through the diode depletion region to avoid this effect, fields are still present in the substrate below each diode, and transit-time effects may still occur to some extent.

Finally, the devices themselves can be further improved. Higher diode cutoff frequencies will both reduce line attenuation and decrease the minimum compressed falltime. With the minimum feature sizes now possible in processing, the optimum N- layer doping is greater than $3 \times 10^{16}/cm^3$. To maximize this cutoff frequency, both the ohmic-Schottky spacing and the Schottky contact width should be at the process minimum dimensions. With $\sim$ 1 THz diode cutoff frequencies, the quasi-static diode analysis of Chapter 3 is unreliable, and skin losses and electron plasma effects (see references [3] and [4] of Chapter 3) must be assessed. From microwave measurements of the improved devices, it appears that the interconnecting transmission lines contribute significant skin loss, and the effect of skin impedance on compressed falltime may need to be reevaluated. Line losses can be reduced by using a tapered structure in which the greatest fraction of the length of the nonlinear transmission line uses diodes of moderate capacitance, connected by line sections of large cross-section and hence low loss. Because the wavefront falltime decreases

during propagation, and would thus approach the minimum compressed falltime of a large-dimension structure, the final compression sections near the line output uses scaled sections, i.e. smaller diodes interconnected with lines of smaller lateral dimensions. The scaled line sections near the line output permits compression to the smaller minimum compressed falltime of the scaled structure while incurring the smaller loss of the unscaled line.

Both the initial and the improved monolithic devices use a uniform doping profile in the N- layer. As a result, the capacitance varies as the inverse square root of the voltage, and (neglecting line capacitance) the per-section delay then varies as the inverse fourth root of the voltage. As a result, the per-diode compression is only a small fraction of the per-diode delay, which is the inverse of the periodic-line cutoff frequency ω_{per}. Large numbers of diodes are thus required for large compression ratios. If a hyperabrupt N- doping profile were adopted, the change in capacitance with voltage could be increased, increasing the compression per section. Fewer diodes would then be required for a given compression ratio, and the line losses would be proportionally reduced. Further, the approximate analysis of the stretch-compress model suggests that the minimum compressed falltime will also be decreased given the increased per-section compression.

Nonlinear transmission lines require the same circuit components and fabrication processes as Schottky diode sampling bridges. Given small improvements in the nonlinear transmission line, monolithic integration with a diode sampling bridge would produce a 100 GHz monolithic GaAs sampling head for sampling oscilloscopes and network analyzers.