

THE MAY-JUNE 1983
VOL. 62, NO. 5, PART 1



BELL SYSTEM
TECHNICAL JOURNAL

A Long Wavelength Optical Receiver Using a Short Channel Si-MOSFET K. Ogawa, B. Owen, and H. J. Boll	1181
Modal Analysis of Loss and Mode Mixing in Multimode Parabolic Index Splices I. A. White and S. C. Mettler	1189
Experimental Results of 20-Mb/s FSK Digital Transmission on 4-GHz (TD) Radio Y. Y. Wang	1209
An Experimental Broadband Imaging Feed T. S. Chu and R. W. England	1233
An M/G/1 Queue With a Hybrid Discipline B. T. Doshi	1251
Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design M. S. Phadke, R. N. Kackar, D. V. Speeney, and M. J. Grieco	1273
The Effects of Selected Signal Processing Techniques on the Performance of a Filter-Bank-Based Isolated Word Recognizer B. A. Dautrich, L. R. Rabiner, and T. B. Martin	1311
CONTRIBUTORS TO THIS ISSUE	1337
PAPERS BY BELL LABORATORIES AUTHORS	1341
CONTENTS, JULY-AUGUST ISSUE	1349

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

D. E. PROCKNOW, *President*

I. M. ROSS, *President*

W. M. ELLINGHAUS, *President*

Western Electric Company

Bell Telephone Laboratories, Incorporated

American Telephone and Telegraph Company

EDITORIAL COMMITTEE

A. A. PENZIAS, *Committee Chairman, Bell Laboratories*

M. M. BUCHNER, JR., *Bell Laboratories*

R. P. CLAGETT, *Western Electric*

T. H. CROWLEY, *Bell Laboratories*

B. R. DARNALL, *Bell Laboratories*

B. P. DONOHUE, III, *American Bell*

I. DORROS, *AT&T*

R. A. KELLEY, *Bell Laboratories*

R. W. LUCKY, *Bell Laboratories*

R. L. MARTIN, *Bell Laboratories*

J. S. NOWAK, *Bell Laboratories*

L. SCHENKER, *Bell Laboratories*

G. SPIRO, *Western Electric*

J. W. TIMKO, *American Bell*

EDITORIAL STAFF

B. G. KING, *Editor*

PIERCE WHEELER, *Managing Editor*

LOUISE S. GOFFER, *Assistant Editor*

H. M. PURVIANCE, *Art Editor*

B. G. GRUBER, *Circulation*

THE BELL SYSTEM TECHNICAL JOURNAL (ISSN0005-8580) is published by the American Telephone and Telegraph Company, 195 Broadway, N. Y., N. Y. 10007; C. L. Brown, Chairman and Chief Executive Officer; W. M. Ellinghaus, President; V. A. Dwyer, Vice President and Treasurer; T. O. Davis, Secretary.

The Journal is published in three parts. Part 1, general subjects, is published ten times each year. Part 2, Computing Science and Systems, and Part 3, single-subject issues, are published with Part 1 as the papers become available.

The subscription price includes all three parts. Subscriptions: United States—1 year \$35; 2 years \$63; 3 years \$84; foreign—1 year \$45; 2 years \$73; 3 years \$94. Subscriptions to Part 2 only are \$10 (\$12 foreign). Single copies of the Journal are available at \$5 (\$6 foreign). Payment for foreign subscriptions or single copies must be made in United States funds, or by check drawn on a United States bank and made payable to The Bell System Technical Journal and sent to Bell Laboratories, Circulation Dept., Room 1E-335, 101 J. F. Kennedy Parkway, Short Hills, N. J. 07078.

Single copies of material from this issue of The Bell System Technical Journal may be reproduced for personal, noncommercial use. Permission to make multiple copies must be obtained from the editor.

Comments on the technical content of any article or brief are welcome. These and other editorial inquiries should be addressed to the Editor, The Bell System Technical Journal, Bell Laboratories, Room 1J-319, 101 J. F. Kennedy Parkway, Short Hills, N. J. 07078. Comments and inquiries, whether or not published, shall not be regarded as confidential or otherwise restricted in use and will become the property of the American Telephone and Telegraph Company. Comments selected for publication may be edited for brevity, subject to author approval.

Printed in U.S.A. Second-class postage paid at Short Hills, N. J. 07078 and additional mailing offices. Postmaster: Send address changes to The Bell System Technical Journal, Room 1E-335, 101 J. F. Kennedy Parkway, Short Hills, N. J. 07078.

© 1983 American Telephone and Telegraph Company.

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 62

May-June 1983

Number 5, Part 1

Copyright © 1983 American Telephone and Telegraph Company. Printed in U.S.A.

A Long-Wavelength Optical Receiver Using a Short-Channel Si-MOSFET

By K. OGAWA, B. OWEN, and H. J. BOLL

(Manuscript received April 21, 1982)

Recent improvements in fine-line technology have resulted in silicon metal oxide semiconductor field-effect transistors (MOSFETs) with channel lengths between 0.2 and 0.8 μm . We have measured the low-frequency noise in these transistors and find it to be smaller than that in comparable GaAs-metal Schottky valve field-effect transistors (MESFETs). Theoretical considerations on the FET noise and experimental results at 45 Mb/s indicate that Si-MOSFETs can compete with GaAs-MESFETs in hybrid photoamplifier circuits. As a natural extension, Si-MOSFETs can also be used for the complete monolithic integration of the receiver circuit with the benefits of reliability and improved performance.

I. INTRODUCTION

In the absence of high-performance avalanche photodiodes for long-wavelength optical receivers, p-i-n photodiodes with low-noise amplifiers have been used.¹ The amplifiers are designed with ultra-low-noise components to realize high receiver sensitivities. Up until now, GaAs-metal Schottky valve field-effect transistors (MESFETs) were used exclusively as low-noise components at bit rates less than 300 Mb/s.^{2,3} We have fabricated a short-channel Si-metal oxide semiconductor field-effect transistor (MOSFET)⁴ and used this MOSFET in a hybrid integrated receiver circuit at 45 Mb/s. The receiver's performance is similar to that of a receiver employing a GaAs-MESFET. The use of

Si-MOSFETs creates the opportunity for monolithic integration of the entire front-end amplifier, with the ensuing benefits of circuit reliability and stability.

In the past, Si-FETs were ignored for this application because they were believed to have a lower transconductance than GaAs-FETs of comparable dimensions. This was due to the difference in mobility between the two materials. Recent improvements in fine-line technology have resulted in silicon MOSFETs with short channels from 0.2 to 0.8 μm . Also, we find that a short-channel Si-FET operating in the saturation region of the electron drift velocity has a transconductance comparable to that of the best GaAs-FET.^{5,6} GaAs-FETs exhibit additional noise because of electron scattering in the high electric field of the channel. Since this effect is generally absent in Si-FETs,^{7,8} the receiver sensitivity obtained using a Si-MOSFET is now expected to be comparable to or slightly better than that obtained using a GaAs-MESFET.

II. SILICON SHORT-GATE MOSFET

Table I lists the characteristics of a Si-NMOSFET with a channel length between 0.5 and 0.8 μm , and for comparison shows the typical characteristics of a GaAs-MESFET with a channel length between 0.5 and 1.0 μm . The figure of merit, g_m/C of a Si-N-channel metal oxide semiconductor field-effect transistor (NMOSFET) is smaller than that of a GaAs-MESFET when structures with the same dimensions are compared.⁶ However, the noise factor, Γ , of a Si-NMOSFET is smaller than that of the GaAs FET if induced gate noise and its correlation with channel noise are considered.^{5,7-9} The mean square of the equivalent input noise current of an FET is given by

$$i_n^2 = 4k\Gamma \frac{\omega^2(C_{gs} + C_{in})^2}{g_m} \Delta f$$

$$\Gamma = P + \left| \frac{C_{gs}}{C_{in} + C_{gs}} \right|^2 R - 2Q \left| \frac{C_{gs}}{C_{in} + C_{gs}} \right|,$$

where Γ is the noise factor, g_m the transconductance, C_{gs} the gate-source capacitance, and C_{in} the input capacitance consisting of the

Table I—Typical Si-NMOSFET characteristics

	Gate-Source Capacitance (pF)	Transcon- ductance (mS)	Figure of Merit, g_m/C_{gs} (mS/pF)
Si-NMOS	0.5 to 0.8	40 to 50	60 to 70
GaAs-MESFET (0.5 to 1.0 μm)	0.2 to 0.5	25 to 50	60 to 140

Table II—Noise factor Γ , P , Q , R
for Si-NMOSFET

P	0.763
Q	-0.206
R	0.245
Typical Γ	1.03
$C_{gs} = 0.5_{pF} C_{in} = 0.5_{pF}$	

photodiode capacitance and any parasitic capacitance. P is the noise factor for the channel noise, R is the noise factor for the induced gate noise, and Q is the correlation factor. Table II indicates the values of Γ , P , Q , and R for a Si-FET with a $0.5\text{-}\mu\text{m}$ channel length. The noise factor, Γ , for the Si-FET (1.03) is much smaller than the value for a typical GaAs-FET (1.78).

Low-frequency $1/f$ noise in FETs has an important effect on the performance of an optical receiver at low bit rates. We have measured the low-frequency noise of both a Si-NMOSFET and a GaAs FET. The results are shown in Fig. 1. Whereas the low-frequency noise for the GaAs FET does indeed have a $1/f$ dependence, the results show a $f^{-1/2}$ dependence for the Si-NMOSFET. This result has not yet been explained.

The noise measured in Fig. 1 was normalized to the expected channel noise $4kT\Gamma\Delta f/g_m$. The FET transconductance, g_m , was 48 mS; the filter bandwidth, Δf , was 3.1 kHz; and the noise factor, Γ , was 1.03 for

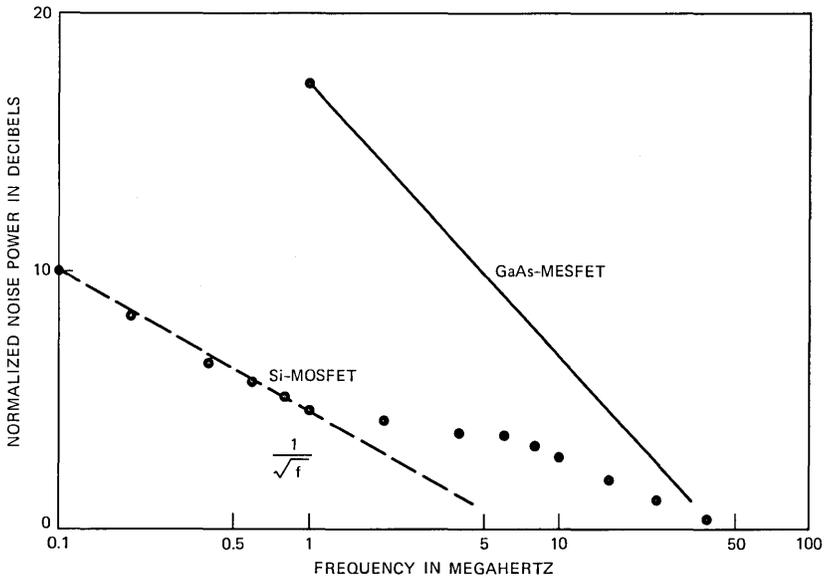


Fig. 1—Low-frequency characteristics of Si-NMOSFET. The dotted points show results measured with a 3.1-kHz filter. The best fit showed by the dashed line has a $f^{-1/2}$ slope. The $1/f$ noise of the GaAs-MESFET is shown (solid line) for comparison.

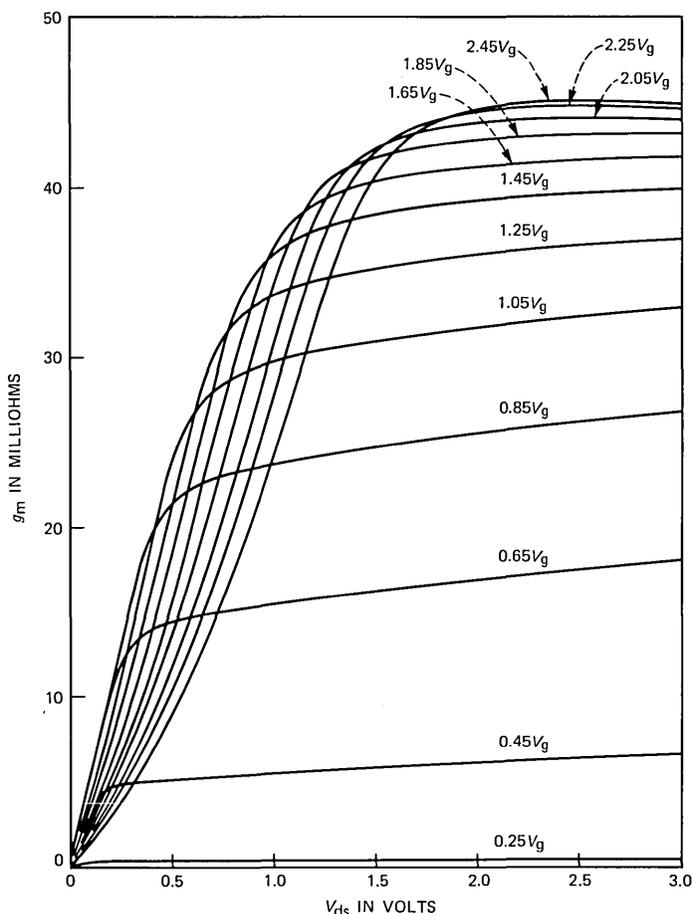


Fig. 2—Transconductance of Si-NMOSFET. Parameters are the gate-source voltage.

the Si-NMOSFET and 1.78 for the GaAs FET. In the Si-NMOSFET, the channel noise exceeded the $4kT\Gamma\Delta f/g_m$ value. The excess noise is believed to be thermal noise associated with the large series resistance of the polysilicon gate. This series gate resistance can be reduced by improved fabrication techniques, such as metallizing the gate. From Fig. 1, the noise corner frequency, f_c , for the Si-NMOSFET is ≈ 5 MHz. The noise corner frequency for the GaAs FET is ≈ 30 MHz. Therefore, even with the excess noise from the gate resistance, the low-frequency noise contribution of the Si-NMOSFET is clearly smaller than that of the GaAs FET.

Another FET parameter that affects its performance in an optical receiver is gate leakage current. The gate leakage current contributes shot noise at the receiver front end. Again, the Si-NMOSFET is

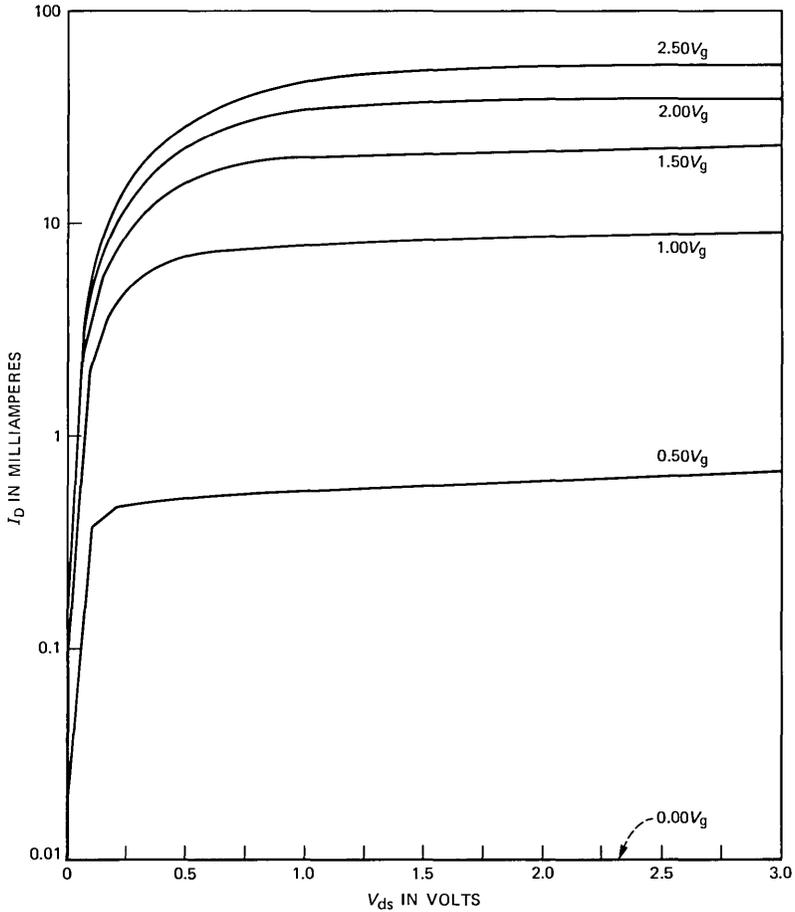


Fig. 3—I-V characteristics of Si-NMOSFET. The parameters are the gate-source voltage.

superior to the GaAs FET. The measured gate leakage current of the GaAs FET is ≈ 5 nA. The measured gate leakage current of the Si-NMOSFET is ≈ 10 pA.

Based on these measurements, we have calculated the sensitivity of a 45-Mb/s optical receiver using both a Si-NMOSFET and a GaAs FET. The receiver sensitivity is given by

$$\eta \bar{P} = \left| \frac{h\nu Q}{q} \right| \langle i^2 \rangle_T^{1/2},$$

where the prefactor $(h\nu Q/q)$ is 4.950 W/A at $1.3 \mu\text{m}$ and at 10^{-7} bit error rate; and where the equivalent input noise current, $\langle i^2 \rangle_T$, is given by

$$\langle i^2 \rangle_T = 2qI_L I_2 B$$

$$+ 2qI_D I_2 B$$

$$+ \frac{4kT}{R_F} I_2 B$$

$$+ \frac{16\pi^2 k T \Gamma (C_T^2) I_3 B^3}{g_m}$$

$$+ \frac{16\pi^2 k T \Gamma (C_T)^2 f_c I_f B^2}{g_m}$$

$$+ \langle i^2 \rangle_c$$

Shot noise from the gate leakage current, I_L .

Shot noise from the photodiode dark current, I_D .

Thermal noise from the bias resistor, R_F .

Channel noise in the FET.

1/f noise in the FET.

Postamplifier noise,

where B is the bit rate, and I_2, I_3 , etc. are Personick integrals associated with the circuit noise. Assuming R_F to be 500 k Ω , and the p-i-n photodiode dark current, I_D , at 30 nA, the sensitivity of a 45 Mb/s receiver at 10^{-7} bit error rate is -51.3 dBm for a GaAs FET front-end amplifier and -51.8 dBm for a Si-NMOSFET front-end amplifier.

III. EXPERIMENTS

The Si-NMOSFET used for our experiments was fabricated on a p -type substrate (carrier concentration $\approx 2 \times 10^{15}/\text{cm}^3$) with an implanted

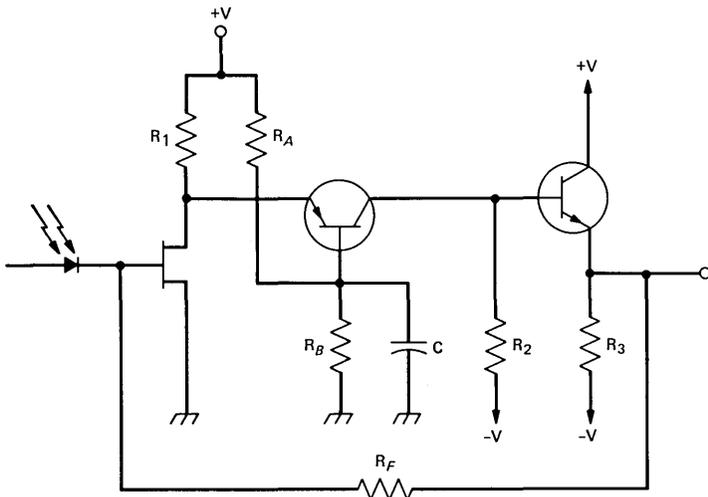


Fig. 4—Front-end amplifier circuit diagram with three active components involved. The first stage with the Si-NMOSFET provides a high input impedance. The second stage p-n-p transistor cascode circuit provides a gain stage and reduces the Miller effect. The last stage is an emitter follower for low output impedance.

n-layer (carrier concentration $\approx 5 \times 10^{16}/\text{cm}^3$), which had a thickness between 0.5 and 0.6 μm .⁴ The channel width was 500 μm , and the effective channel length was 0.45 μm .⁵ The gate-source capacitance was 0.5 pF and the transconductance was 45 mS. The gate leakage current was ≈ 10 pA. Figures 2 and 3 show the drain current and transconductance of the Si-NMOSFET versus drain source voltage with different gate voltages.

We have fabricated a transimpedance front-end circuit³ using an InGaAs p-i-n photodiode with the Si-NMOSFET as the first amplifier stage. The circuit is shown in Fig. 4. The primary gain was achieved in the second stage, which used a p-n-p transistor. The third stage was an emitter-follower circuit using an n-p-n transistor. The feedback

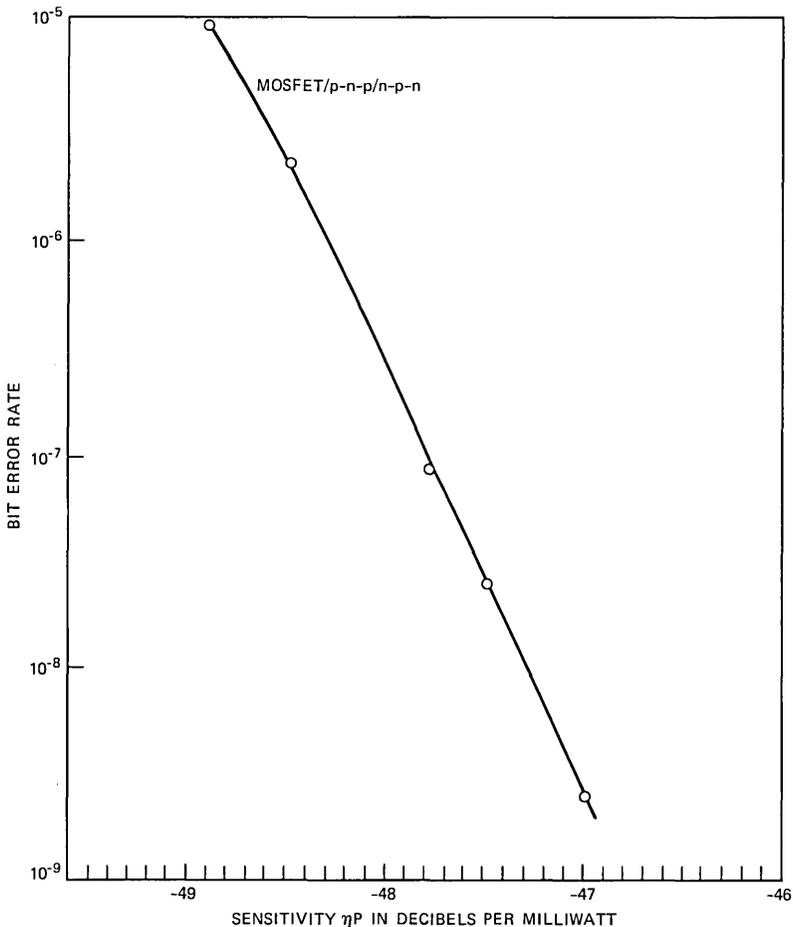


Fig. 5—Error-rate measurement of an InGaAs p-i-n-Si-NMOSFET receiver at 45 Mb/s.

circuit was 500 k Ω . The bit error rate was measured using a 44.7 Mb/s pseudorandom nonreturn to zero (NRZ) optical signal from an InGaAsP LED emitting at 1.3 μm . The receiver circuit was combined with a regenerator circuit and a retiming circuit designed for optical receivers. As shown in Fig. 5, the measured sensitivity at 10^{-7} bit error rate was -47.8 dBm (-51.8 dBm theoretical). With a GaAs FET first-amplifier stage, the same receiver circuit had a measured sensitivity of -49.5 dBm (-51.3 dBm theoretical).

IV. CONCLUSION

Further work is in progress to improve the Si-NMOSFET receiver performance and to integrate the front-end amplifier. The circuit, especially the second stage, is not presently optimized. Also, the Si-NMOSFET has a high series gate resistance because the gate was fabricated with polysilicon. The noise penalty associated with this resistance can be eliminated by metallizing the gate.

In conclusion, theoretical considerations and our first experiments indicate that Si-NMOSFETs can have sensitivity performance comparable to that of GaAs FETs in a 45-Mb/s optical receiver. The natural extension of this result is the complete monolithic integration of the entire receiver circuit using silicon fine-line MOS technology.

REFERENCES

1. Tingye Li, "Optical Fiber Communication—The State of the Art," IEEE Trans. Commun., *COM-26* (July 1978), pp. 946–55.
2. D. R. Smith, R. C. Hooper, R. P. Webb, and M. F. Saunders, "PIN Photodiode Hybrid Optical Receivers," Proc. Optical Commun. Conf., Amsterdam, September 17–19, 1979.
3. K. Ogawa and E. L. Chinnock, "GaAs-FET Transimpedance Front-End Design for a Wide Band Optical Receiver," Elec. Lett., *15*, No. 20 (September 1979), pp. 650–3.
4. P. I. Suoiu, E. N. Fuls, and H. J. Boll, "High-Speed NMOS Circuits Made with X-Ray Lithography and Reactive Sputter Etching," IEEE Elec. Device Lett., *EDL-1*, No. 1 (January 1980), pp. 10–11.
5. W. Baechtold, "Si and GaAs 0.5 μm Gate Schottky Barrier Field Effect Transistors," Elec. Lett., *9*, No. 10 (May 1973), p. 232.
6. T. Wada and J. Frey, "Physical Basis of Short Channel MESFET Operation," IEEE Trans. Elec. Dev., *ED-20*, No. 4 (April 1979), pp. 476–89.
7. W. Baechtold, "Noise Behavior of Schottky Barrier Gate Field-Effect Transistors at Microwave Frequencies," IEEE Trans. Elec. Dev., *ED-18*, No. 2 (February 1971), pp. 99–104.
8. W. Baechtold, "Noise Behavior of GaAs Field Effect Transistors with Short Gate Lengths," IEEE Trans. Elec. Dev., *ED-19*, No. 5 (May 1972), pp. 676–80.
9. K. Ogawa, "Noise Caused by GaAs-MESFETs in Optical Receivers," B.S.T.J., *60*, No. 6 (July–August 1981), pp. 923–8.

Modal Analysis of Loss and Mode Mixing in Multimode Parabolic Index Splices

By I. A. WHITE and S. C. METTLER

(Manuscript received November 18, 1982)

In this paper we present an electromagnetic modal theory for characterizing parabolic-index multimode fiber splices with either intrinsic or extrinsic mismatches. The theory agrees with previously published theoretical results for transverse offset using a uniform power distribution. It also agrees with new experimental measurements made with a long, spliced input fiber using a published, theoretical, steady-state modal power distribution. This modal theory predicts, and experiment confirms, a previously unreported periodic fluctuation in splice loss as a function of wavelength for intrinsic parameter mismatch. The analysis also predicts a large degree of mode mixing for transverse offset but negligible mode mixing for parameter mismatch in typical multimode fiber splices.

I. INTRODUCTION

Theoretical predictions of splice loss in multimode fibers have been attempted by many researchers (see Ref. 1 for a list). However, these analyses¹⁻³ do not adequately predict measured splice loss results, and most do not address mode mixing effects, which are important for the prediction of the bandwidth of concatenated lengths of fiber. The best agreement with splice loss measurement data is an empirical model based on geometric optics.² However, such an analysis of splice loss provides only a limited description of the effect of splices. Electromagnetic theory, using the coupling coefficients of individual modes, provides a complete treatment of splice loss and mode mixing. The only published electromagnetic theory³ appears to be in error, since predictions do not approach the well-known correct geometric optics limit for splice loss of fibers with a large normalized frequency, V , and a uniform modal power distribution. This paper presents an electromagnetic analysis for splices, which gives single-term expressions for mode

coupling coefficients for the case of either transverse offset or profile parameter mismatch for parabolic-index multimode fibers. These coupling coefficients are used to calculate loss and degree of mode mixing in splices. The results approach the proper geometric optics limit for a uniform modal power distribution. Furthermore, if we compare theoretical and experimental loss to transverse offset results, we can verify a published theoretical "steady-state" power distribution⁴ in multimode parabolic-index fibers. Comparing the results of this theory with a measurement of splice loss as a function of transverse offset gives the modal power distribution in a fiber. A wavelength dependence of splice loss for intrinsic parameter mismatch is predicted by this theory and has been experimentally verified. Because splices change the modal power distribution, the power redistribution can cause additional loss as the system evolves towards the steady state in the receiving fiber. In the past² this loss has been considered part of the total splice loss, but the improvement of fiber quality has reduced this effect significantly for typical lengths between splices. Because we have ignored the power redistribution effects (after the splice), these theoretical results are only strictly valid for short lengths of fiber after the splice; however, they should remain valid for typical distances between splices.

II. THEORY

The modal amplitude coupling coefficient, $C[m_1\alpha_1; m_2\alpha_2]$, for mode $m_1\alpha_1$ of the transmitting fiber and mode $m_2\alpha_2$ of the receiving fiber is obtained directly from the theory of excitation of weakly guiding fibers.⁵ (The variable m is the radial mode number and α is the azimuthal mode number.)

$$C[m_1\alpha_1; m_2\alpha_2] = \frac{1}{2} \left(\frac{\epsilon_{co}}{\mu} \right)^{1/2} \int \bar{E}_{m_1\alpha_1} \cdot \bar{E}_{m_2\alpha_2}^* dA, \quad (1)$$

where for parabolic-index fibers (using the infinite profile approximation) the electromagnetic modal field, $E_{m_1\alpha_1}$, propagating in the z direction in the fiber core is:

$$\bar{E}_{m_1\alpha_1} = \begin{pmatrix} \hat{x} \\ \hat{y} \end{pmatrix} A_{m_1\alpha_1} t^{\frac{|\alpha_1|}{2}} L_{m_1}^{|\alpha_1|}(t) e^{-\frac{t}{2}} e^{-i(\alpha_1\phi + \beta z)}, \quad (2)$$

where \hat{x} and \hat{y} are the linearly polarized unit field vectors of the mode. ϵ_{co} and μ are the permittivity and permeability of the core, respectively, and $\epsilon_{co} = n_{co}^2 \epsilon$. Then

$$A_{m_1\alpha_1} = \left[\left(\frac{\mu}{\epsilon_{co}} \right)^{1/2} \frac{V}{\pi a^2} \frac{m_1!}{(m_1 + |\alpha_1|)!} \right]^{1/2}, \quad (3)$$

and

$$\left. \begin{aligned} t &= V \left(\frac{r}{a} \right)^2 \\ \beta^2 &= \left(\frac{2\pi}{\lambda} n_{co} \right)^2 - 2N \frac{V}{a^2} \\ V &= \frac{2\pi}{\lambda} n_{co} a \sqrt{2\Delta}, \end{aligned} \right\} \quad (4)$$

where the principal mode number is $N (= 2m + \alpha + 1)$, and $L_m^\alpha(t)$ is an associated Laguerre polynomial. The variable a is the fiber core radius and $\Delta = (n_{co} - n_{cl})/n_{co}$ is the normalized maximum refractive index difference. For weakly guiding fibers the polarization property of the incoming mode is preserved by the splice, as can be seen from eq. (1).

In general we can write:

$$C(m_1\alpha_1; m_2\alpha_2) = K \cdot I(m_1\alpha_1; m_2\alpha_2), \quad (5)$$

where

$$K = \left[\frac{m_1! m_2!}{(m_1 + |\alpha_1|)! (m_2 + |\alpha_2|)!} \right]^{1/2}. \quad (6)$$

$I(m_1\alpha_1; m_2\alpha_2)$ is a function that depends on the type of splice mismatch. Table I shows the expressions for $I(m_1\alpha_1; m_2\alpha_2)$ for the two cases of intrinsic parameter (V) mismatch and transverse offset between identical fibers (r_o). The derivation of these equations is given in the appendix.

For the parameter mismatch case, the coupling coefficients are a function of the ratio of the normalized frequencies of the two fibers. The hypergeometric function, ${}_2F_1$, is, in this case, a power series in y^2 of order m_1 . Note that because azimuthal symmetry is preserved in the splice, only modes with the same azimuthal mode number couple. In the identical fiber transverse offset case, the coupling coefficients are simple products of Laguerre-Gaussian polynomials with argument proportional to the normalized offset (r_o/a). In this case all azimuthal modes have finite coupling coefficients, but, for small offsets, coupling is much stronger for nearest neighbor azimuthal modes.

Assuming a random phase relationship between the modes of the transmitting fiber, the total power coupled into mode $m_2\alpha_2$, $P_{m_2\alpha_2}$, is:

$$P_{m_2\alpha_2} = \sum_{m_1, \alpha_1}^T |C(m_1\alpha_1; m_2\alpha_2)|^2 P_{m_1\alpha_1}, \quad (7)$$

and the total splice loss, δ_s , is then:

$$\delta_s = -10 \log \left(\frac{\sum_{m_2, \alpha_2}^R P_{m_2 \alpha_2}}{\sum_{m_1, \alpha_1}^T P_{m_1 \alpha_1}} \right) \quad (8)$$

The mode mixing factor, Γ (i.e., the fraction of the power that is redistributed into different modes), is:

$$\Gamma = 1 - \frac{\sum_{m_1=m_2, \alpha_1=\alpha_2}^T |C(m_1 \alpha_1; m_2 \alpha_2)|^2 P_{m_1 \alpha_1}}{\sum_{m_2, \alpha_2}^R P_{m_2 \alpha_2}} \quad (9)$$

The superscripts R and T refer to sums over the bound mode spectra of the receiving and transmitting fibers, respectively.

The range of allowed radial and azimuthal mode numbers, m and α , for a fiber with normalized frequency V is:

$$\begin{aligned} m &\leq m_{\max} = \text{INT}(V/4) \\ \alpha &\leq \alpha_{\max} = \text{INT}(V/2) - 1 \\ N_{\max} &= \text{INT}(V/2), \end{aligned} \quad (10)$$

Table I—Mode coupling functions

Splice	Parameter of Fiber		$I(m_1 \alpha_1; m_2 \alpha_2)$
	Transmitting	Receiving	
Parameter Mismatch			$= 0$ for $\alpha_1 \neq \alpha_2$
	a_T	a_R	$= \frac{(m_1 + m_2 + \alpha)!}{m_1! m_2!} \cdot \frac{(1 + \epsilon)^{(a+1)/2}}{\left(1 + \frac{\epsilon}{2}\right)^{a+1}} \cdot (-1)^{m_1}$
	Δ_T	Δ_R	$\cdot y^{-(m_1+m_2)} {}_2F_1[-m_2; -m_1; -(m_1+m_2+\alpha); y^2]$
	V_T	V_R	where $y = (2 + \epsilon)/\epsilon$ and $\epsilon = \left(\frac{V_R}{a_R} \cdot \frac{a_T^2}{V_T}\right) - 1$
Transverse Offset for Identical Fibers	a	a	$= (R_o)^{p/2} e^{-R_o} \cdot \phi(m_1 \alpha_1; m_2 \alpha_2)$ for $\alpha_1 > 0, \alpha_2 < 0$;
	Δ	Δ	$\phi = L_{m_1}^{ \alpha_1 - \Delta m} (R_o) L_{m_2}^{ \alpha_1 + \Delta m} (R_o)$ for $\alpha_1 > 0, \alpha_2 > 0$;
Transverse displacement of the fiber axes - r_o			$\phi = \frac{(m_2 + \alpha_2)!}{m_1!} \cdot (-R_o)^{\Delta m} L_{m_2 + \alpha_2}^{p + \Delta m} (R_o) \cdot L_{m_2}^{\Delta m} (R_o)$,
			where $R_o = \frac{V}{4} \left(\frac{r_o}{a}\right)^2$; $\Delta m = m_1 - m_2$; $p = \alpha_1 - \alpha_2 $

where INT indicates the integer part. It is essential to realize the discrete nature of the bound mode spectra to understand the splice loss predictions for intrinsic parameter (V) mismatch. The explicit wavelength dependence of splice loss and mode mixing is demonstrated by the presence of V in the coupling coefficients. However, an implicit wavelength dependence due to the allowed spectrum of bound modes occurs from eq. (10).

III. RESULTS

Equations (8) and (9) can be used to calculate splice losses and degree of mode mixing for the two cases of parameter mismatch and transverse offset. To validate this analysis, they must be compared with existing theories wherever possible. Most existing results are geometric optics evaluations of splice loss, and the modal power distribution equivalent to any assumed ray distribution is difficult to evaluate in general. However, the ray distribution equivalent to the uniform modal power distribution has been well documented,⁶ and comparison of splice loss predictions for this case demonstrates that this analysis is correct. Calculations and experimental results for the effects of different modal power distributions on the mode mixing and loss at a splice are shown in the sections that follow.

3.1 Identical fiber transverse offset

This section presents the results for mode mixing and splice loss associated with transverse offset of the fiber axes for identical fibers.

3.1.1 Splice loss

Splice loss predictions for transverse offset with a uniform power distribution for several values of the normalized frequency, V , are shown in Fig. 1. It is well known that, as V increases, the electromagnetic analysis should asymptotically approach the geometric optics predictions. This is confirmed in Fig. 1. The only other published electromagnetic splice loss theory³ does not appear to converge to the correct geometric optics limit, e.g., from Fig. 3 of Ref. 3, the splice loss for $0.2a$ offset (for $2a = 50 \mu\text{m}$) is ~ 1.0 dB, whereas geometric optics predict a value of 0.8 dB. The small deviation of our results for small offset ($< 0.1a$) is believed to be caused by coupling of power to the bound modes of the receiving fiber through the evanescent cladding fields. This effect cannot be predicted using ray optics.

Uniform power distribution results do not agree with realistic splice loss measurements.² Figure 2 demonstrates loss versus transverse offset for several power distributions, the Gaussian splice loss model,² and experimental data.⁸ The experimental data shown were obtained using $0.82\text{-}\mu\text{m}$ laser excitation of a 7-km input fiber containing about

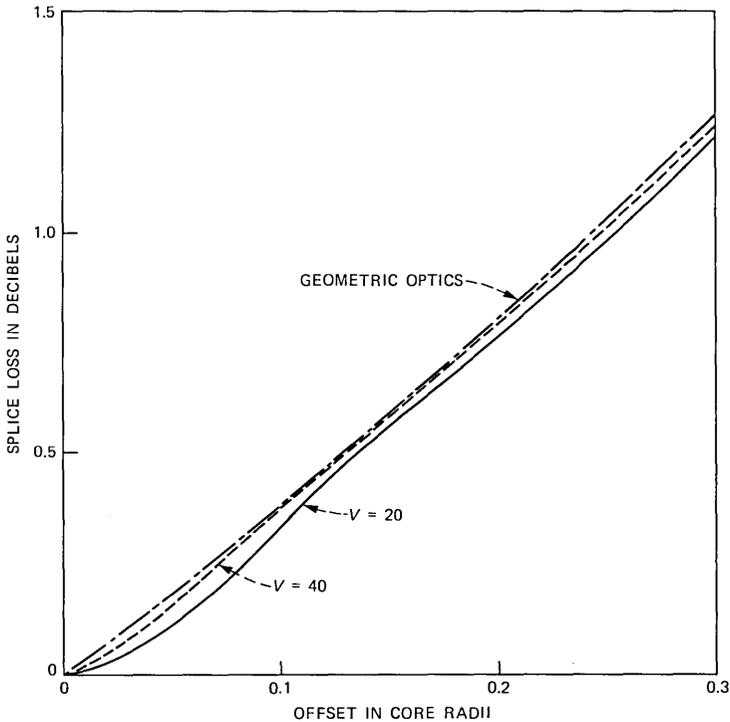


Fig. 1.—Splice loss vs. transverse offset with uniform input power.

14 splices into a final splice that was offset in both x and y orthogonal axes. The offset data are in good agreement with previous data and should represent a realistic “steady-state” condition. The theoretical power distribution for steady state due to microbending⁴ gives excellent agreement with this data. (A power distribution that has been used to approximate this steady-state distribution is also plotted to show the sensitivity of the loss to the choice of power distribution.) The excellent agreement between the data and this analysis using the steady-state power distribution confirms both the distribution and the splice loss theory. Although this power distribution gives excellent agreement with the data for very long lengths of fiber before the splice, it does not agree as well as the Gaussian model² with shorter-length input fiber measurements because the power distribution has not achieved “steady state.”

Figure 3 compares experimental data for shorter lengths of input fiber (~ 1 km) and different excitation conditions⁸ with theoretical predictions using both the Gaussian model² and various power distributions in this modal analysis. Note that the different excitation conditions significantly affect the splice loss for larger offsets. The

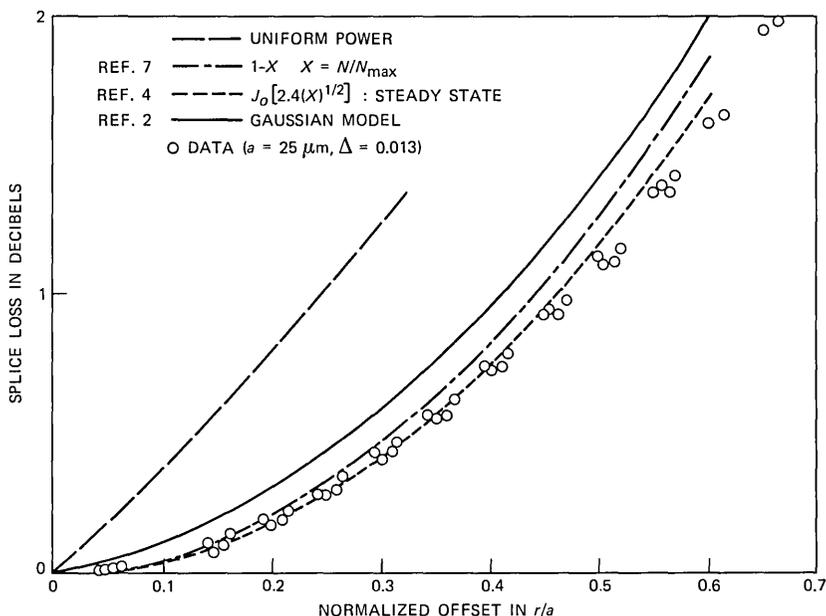


Fig. 2—Loss vs. transverse offset for different modal power distributions.

empirical model of Ref. 2 agrees well with loss measurements for typical offsets expected in practice of 0.1 to 0.2 (r_o/a) for both sources and agrees very well over a wider range for the laser source. Modal power distributions for the electromagnetic analysis to describe this situation were empirically chosen so that, for very long fiber lengths, they would degenerate to the steady-state distribution. The simplest choice for a power distribution satisfying this criteria is:

$$P(m_1\alpha_1) = J_0 \left[2.405 \sqrt{\frac{N}{N_{\max}}} (1 - e^{-\eta L}) \right]. \quad (11)$$

The choices of η shown in Fig. 3 (for $L = 1$ km) demonstrate that excellent agreement between the modal analysis and the splice loss versus transverse offset for the shorter input fiber lengths is possible for both sources. Furthermore, for $\eta = 3$, the results agree with the predictions of the empirical model² within 0.03 dB. The theoretically predicted steady-state distribution gives optimistic results for this short length. The existence of a modal power distribution that is in good agreement with the results of Ref. 2 for transverse offset supports the basic assumptions of the empirical model.

3.1.2 Mode mixing

Equation (9) can be used to calculate the amount of power in the

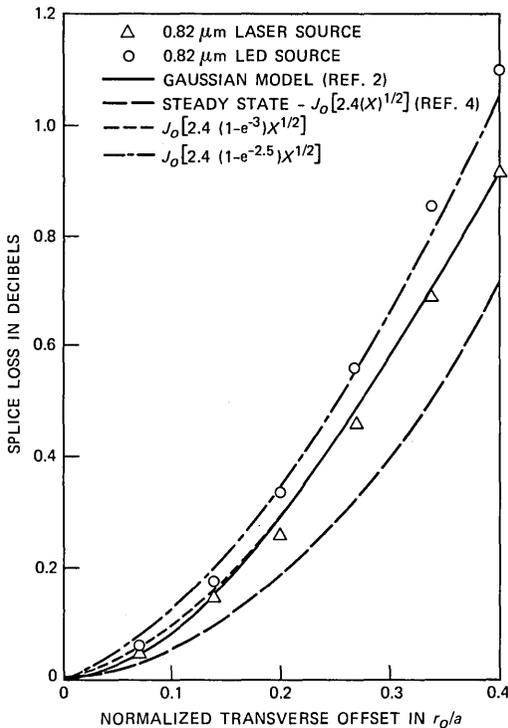


Fig. 3—Splice loss vs. transverse offset for 1-km input fiber.

receiving fiber that has changed propagation mode because of the splice. By changing the summation criteria one can calculate any desired mode mixing factor, such as power coupled within the same principal mode group, nearest neighbor mode groups, etc., where the summation is over all m_1 , α_1 and m_2 , α_2 that differ in principal mode number by ΔN , which is written as:

$$\sum_{2m_2 + \alpha_2 - 2m_1 - \alpha_1 = \Delta N} \quad \text{for } \Delta N = 0, 1, 2, \dots \quad (12)$$

Mode mixing as a function of normalized transverse offset for two different input power distributions is shown in Fig. 4 as the percentage of receiving fiber power that has been redistributed into neighboring mode groups. These differ in principal mode number, N , by ΔN of 0, 1, 2, 3, and 4. ($\Delta N = 0$ represents power coupled within the same degenerate mode group in both transmitting and receiving fibers.) The power distributions used were the uniform and the theoretically predicted power distribution for the steady state for microbending.⁴ It is important to note that the percentage of the power-changing propagating mode groups is very high even for small offsets with low splice

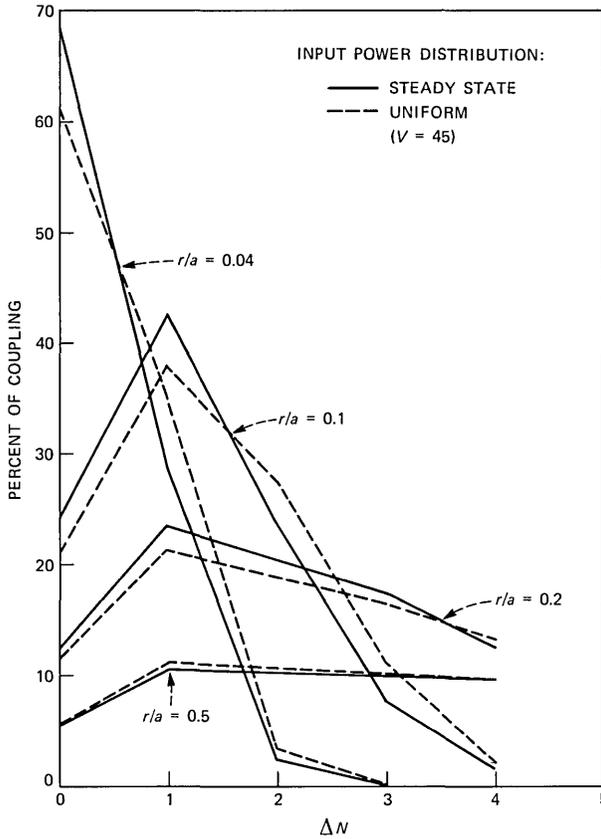


Fig. 4—Power coupling into neighboring mode groups for transverse offset.

loss, i.e., 31 percent for $0.04a$ offset and 0.07 dB splice loss for a $V = 45$ fiber with the steady-state power distribution. Mode mixing is a fairly weak function of the modal power distribution, which indicates that the mode mixing is uniform over all the mode groups as can be seen from the local numerical aperture arguments.² As offset increases, the mode mixing changes from being primarily to the nearest neighbor mode groups ($\Delta N = \pm 1$) to being redistributed over a wider range of principal mode groups. The strength of the mode coupling for even small offsets with low splice loss, and the relatively small offsets at which it spreads over many mode groups, are both initially surprising.

3.2 Parameter mismatch

For fibers with no axial offset, the individual mode fields are insensitive to small changes in the V parameter, so that for splices with only slight parameter mismatch, as can be seen in eq. (1), self-coupling

dominates (i.e., mode $m_1\alpha_1$ of the transmitting fiber couples primarily with mode $m_1\alpha_1$ of the receiving fiber). For example, there is less than 3-percent mode mixing for a 5-percent parameter mismatch with uniform modal excitation over the wavelength range of 0.8 to 1.4 μm . Therefore, splice loss due to parameter mismatch is caused only by this slight field mismatch unless the ranges of the bound mode spectra of the two fibers [eq. (10)] differ, i.e., if the normalized frequencies of the fibers are such that $\Delta N_{\text{max}} \geq 1$. In that case:

$$\Delta N_{\text{max}} = N_{\text{max}}^T - N_{\text{max}}^R \quad (13)$$

or

$$\Delta N_{\text{max}} = \text{INT} \left[\frac{\pi}{\lambda} n_{co}(a_T \sqrt{2\Delta_T}) \right] - \text{INT} \left[\frac{\pi}{\lambda} n_{co}(a_R \sqrt{2\Delta_R}) \right]. \quad (14)$$

When ΔN_{max} is zero, the splice loss is small, as discussed above. However, when $\Delta N_{\text{max}} > 0$, the modes in the highest-order mode group of the transmitting fiber are not bound modes in the receiving fiber. Because self-coupling dominates, most of the power in the highest-order modes couples to leaky/radiation modes in the receiving fiber and is therefore lost. The large degeneracy of the higher-order mode groups [eq. (4)] accentuates this loss. From this discussion we see that parameter mismatch splice loss is caused primarily by the difference between the bound mode volumes of the two fibers, which is a function of the normalized frequencies, V . Therefore, both Δ and radius mismatches can be expressed simply as V mismatch. (An x -percent radius mismatch is equivalent to a $2x$ -percent Δ mismatch for small degrees of mismatch.)

Theoretical predictions of splice loss versus normalized parameter mismatch ($\Delta V/V$) for uniform and steady-state input power distributions at 0.82 and 1.3 μm are shown in Fig. 5. The discontinuities of the splice loss curves are caused by the discreteness of the bound mode spectra, as can be seen from the definition of ΔN_{max} in eq. (14). In Fig. 5a this discreteness is most obvious due to the choice of a uniform input power distribution, which accentuates the power in the highest-order mode group. Note that at the longer wavelength the loss is initially higher for a given parameter mismatch because the percentage of the total power that is in the highest-order mode groups is greater than at shorter wavelengths. Each discontinuous increase of splice loss is caused by the stripping of an additional high-order mode group of the input fiber, which does not propagate in the receiving fiber as $\Delta V/V$ increases. The splice loss predicted by geometric optics (which ignores the discreteness of the mode spectra) is also shown.¹ The results for the steady-state power distribution used previously are shown in Fig. 5b displaying similar discontinuous features. However,

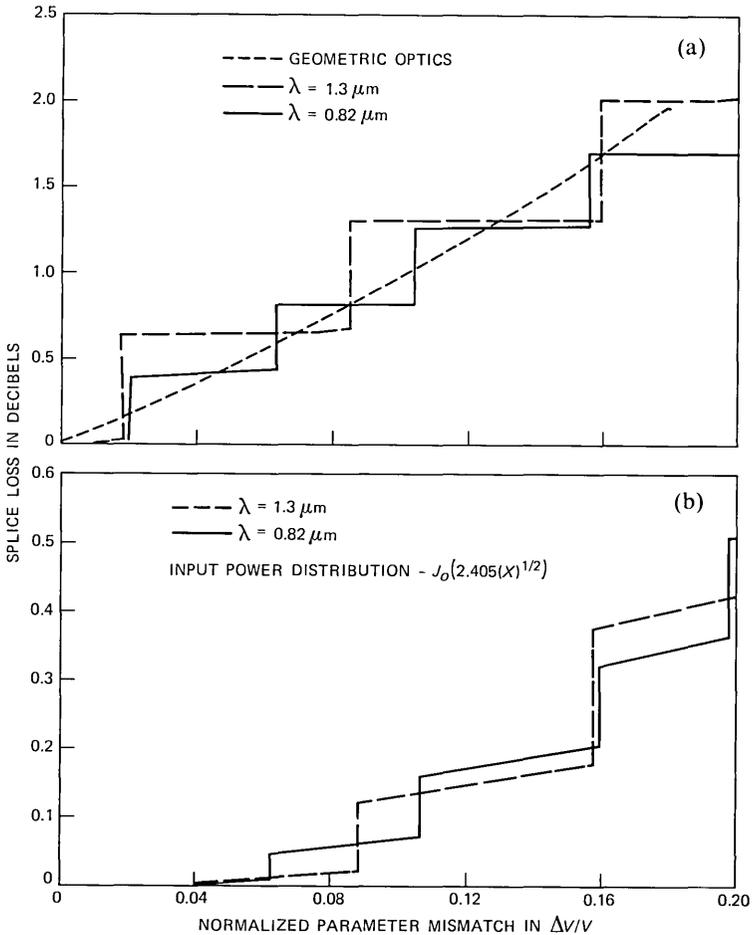


Fig. 5—Splice loss vs. normalized parameter mismatch in $\Delta V/V$ for (a) uniform power distribution and (b) steady-state power distribution.

the decrease in the order of magnitude of the loss allows the observation of the small loss component caused by the slight field mismatch between similar modes of the two fibers. This is reflected in the small slope of the levels. Because the steady-state distribution assumes that there is no power in the highest-order mode group of the transmitting fiber, the first discontinuous change of splice loss of Fig. 5a is negligible in Fig. 5b.

Figure 6 shows the percentage of mode mixing at a splice caused by parameter mismatch for 0.82- and 1.3- μm wavelengths, and the uniform and steady-state power distributions. The substantially higher mode mixing for the uniform power case compared to the steady-state indicates that mode mixing is dominated by the higher-order modes.

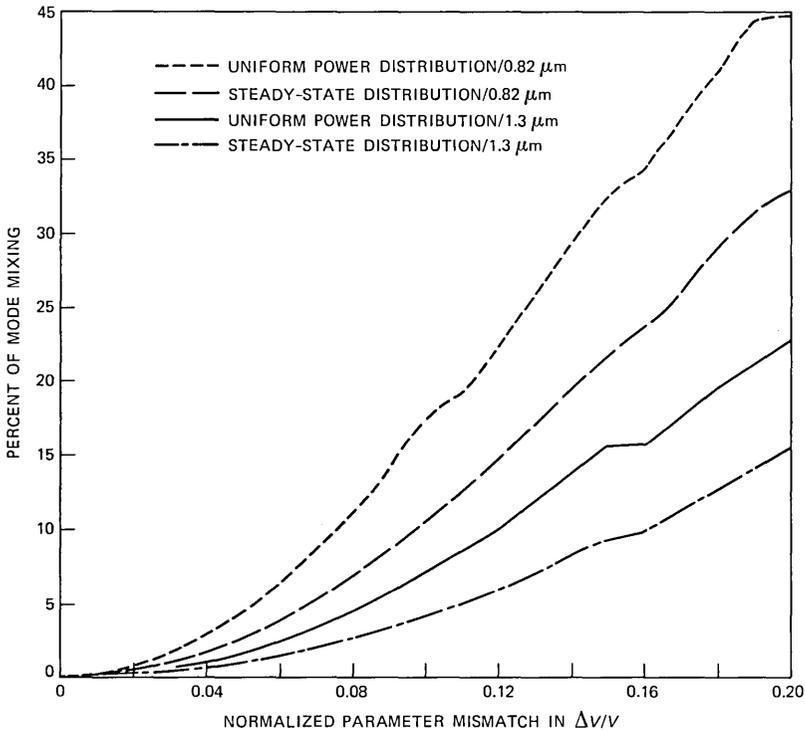


Fig. 6—Percent of mode mixing vs. normalized parameter mismatch.

This contrasts with the transverse offset case shown in Fig. 4, where mode mixing is essentially the same for uniform and steady-state distributions. Reduced mode mixing at longer wavelengths is due to the more diffuse bound mode fields, which result in stronger self coupling between modes. In comparison to the transverse offset case, the degree of mode mixing for the case of parameter mismatch is substantially reduced. Furthermore, a detailed analysis of the mode mixing shows that the coupling is almost completely nearest neighbor coupling ($\Delta N = 1$), e.g., for $\Delta V/V \sim 0.2$, 38 percent of the total 44-percent mode mixed power is coupled to nearest neighbor modes. The deviations from a smoothly increasing function in these curves occur at the transition points of the splice loss curves (see Fig. 5). This is caused by the transition of the highest-order mode group in the receiving fiber from propagating to lossy, removing its contribution to the overall mode mixing.

3.3 Wavelength dependence of splice loss for parameter mismatch

As we can see from eq. (13), for a given degree of parameter mismatch, ΔN_{max} is also a discontinuous function of wavelength. For

example, if, at some particular wavelength, V_T is 46.1 and V_R is 45.9, ΔN_{\max} is 1 and splice loss is ~ 0.5 dB. A slight change of wavelength (~ 1 percent) will cause both V 's to simultaneously be greater or less than 46 with $\Delta N_{\max} = 0$, and splice loss is ~ 0.02 dB. Theoretical considerations [eq. (14)] also indicate that the period of the splice loss fluctuations should increase with increasing wavelength because the mode volume is proportional to $1/\lambda$. The magnitude of the fluctuation should also increase with wavelength because the relative amount of power in the highest-order mode group increases as the mode volume decreases. The splice loss wavelength dependence for a parameter mismatch ($\Delta V/V$) of 0.04 and uniform input power distribution is shown in Fig. 7. The magnitude of the splice loss changes considerably over the entire wavelength range. Splice loss calculated from geometric optics for this parameter mismatch case is also shown in Fig. 7 and is approximately the wavelength averaged splice loss as calculated from modal theory. Although the splice loss displays this pathological behavior, the mode mixing remains relatively small and continuous.

Figure 8 gives two further examples of the theoretical splice loss versus wavelength for uniform power excitation and core radius mismatches of 1.0 and 1.5 μm . For Δa of 1.5 μm , there are ranges of wavelength over which $\Delta N_{\max} = 2$, resulting in even larger fluctuations. The choice of a uniform power distribution unrealistically enhances the magnitude of this effect. For example, with the choice of the steady-state distribution, there would be very little power in the highest-order mode group, and therefore this change in loss would be very small, as we can see in Fig. 5b. Splice loss is not expected to vary as abruptly as shown in Figs. 7 and 8, even for a uniform power distribution, because the cladding splits the degeneracy of the highest-order mode group and some modes of this group remain bound even when the group itself, as predicted by the infinite parabolic approxi-

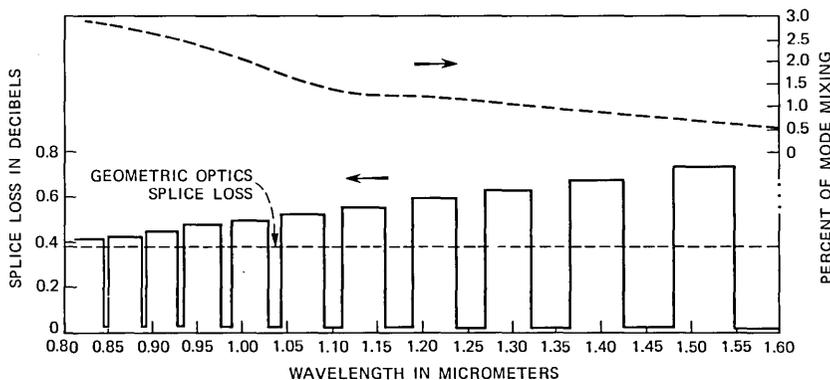


Fig. 7—Splice loss and mode mixing vs. wavelength for $\Delta V/V = 0.04$.

mation, is not bound. This results in a smoothing of the loss versus λ curve, but the periodicity should not be affected. The observation of this effect would further verify this analysis.

3.4 Experiment

The experimental setup shown in Fig. 9 was designed to enhance the power in higher-order mode groups in an attempt to observe the predicted wavelength dependence of splice loss shown in Figs. 7 and 8. A 5 m, 63- μm core, step-index fiber was over-filled and used to excite

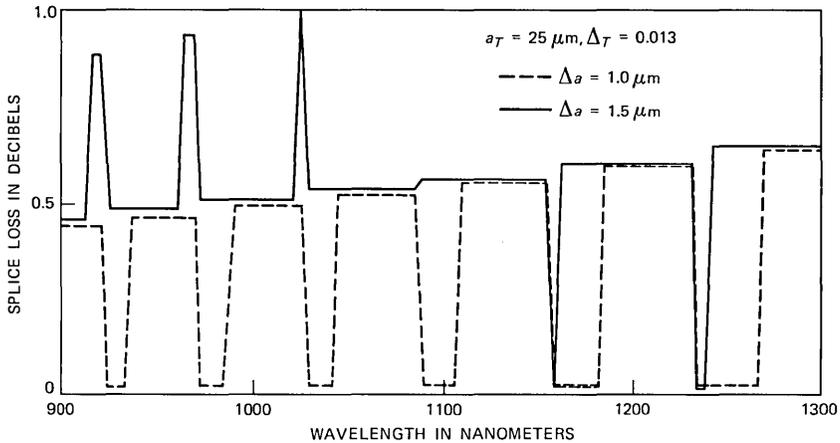


Fig. 8—Theoretical loss vs. wavelength for parameter mismatch.

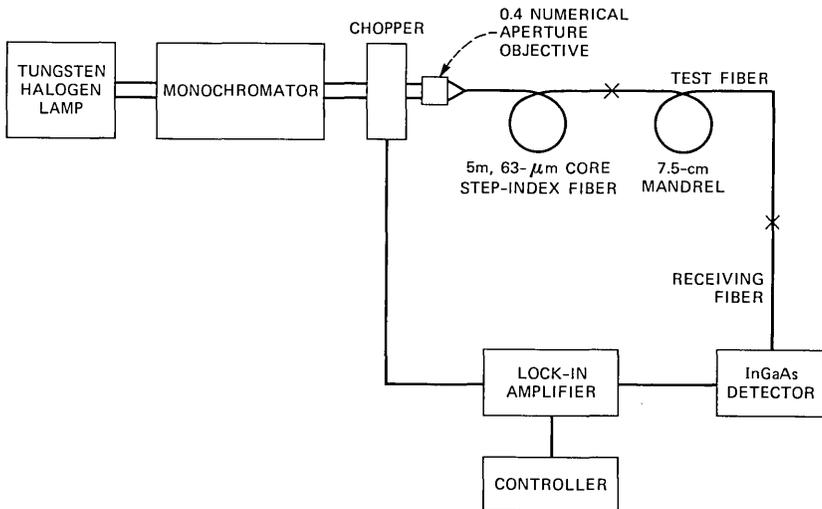


Fig. 9—Splice loss measurement system.

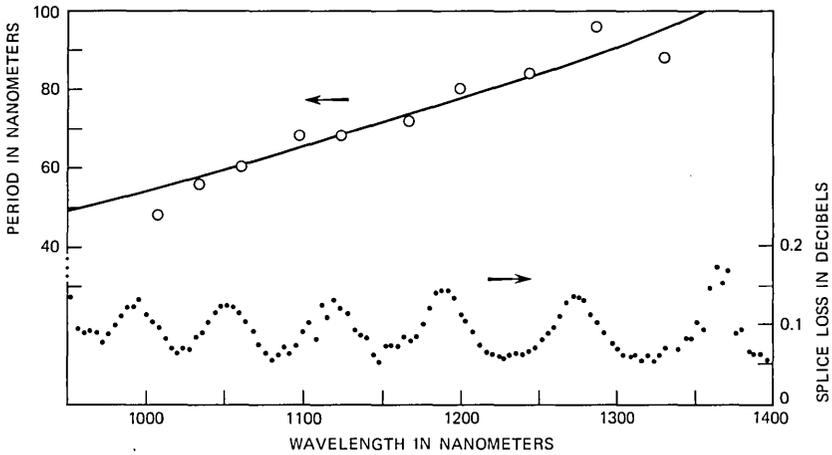


Fig. 10—Measured period and splice loss for parameter mismatch.

the test fiber, which was wound on a 7.5-cm mandrel. A cut-back test determined that approximately 15 m of fiber on this mandrel successfully eliminated leaky modes in the transmitting fiber. This fiber was spliced with micropositioners to a 1 to 2 m length of fiber, which differed in core radius by $1.0 \pm 0.2 \mu\text{m}$ (but with the same Δ), as determined by refracted near-field measurements. Monochromator resolution was 2 nm and the repeatability of the splice loss measurements was 0.02 dB. As a final check the test fiber was spliced to a length of identical fiber. Maximum splice loss change seen in this case was 0.02 dB, with no discernible wavelength dependence.

The measured wavelength dependence of splice loss and its period are plotted in Fig. 10 for the mismatched fibers. The data show excellent agreement with the period versus wavelength curve obtained from Fig. 8. The increase in maximum amplitude with wavelength seen in Fig. 10 is in qualitative agreement with the theory. Cladding effects (ignored in the infinite parabolic approximation) and only partial success in achieving a uniform power distribution are the probable reasons that the maximum amplitude is less than expected from Fig. 8.

IV. CONCLUSIONS

In this paper we have presented a modal theory that allows the calculation of mode mixing and loss resulting from parabolic-index multimode fiber splices. The predictions of this model have been confirmed by experiment, including a previously unreported wavelength dependence of splice loss for intrinsic parameter mismatch.

This phenomenon is not expected to be significant for practical "steady-state" system losses but is important for providing insight into the actual loss mechanisms at a splice as well as verifying the validity of the theory. Another effect of this discreteness of the bound mode spectrum has previously been reported for the wavelength dependence of bending loss.⁹ The theory also successfully describes splice loss versus transverse offset phenomena and has been used to verify the predicted steady-state power distribution. The analysis provides a basis for understanding mode-mixing effects at splices, which is important for bandwidth studies on concatenated fiber lengths. Splice losses caused by offset or parameter mismatch produce greatly different degrees of intermodal coupling. This implies that the effect of the splice on the bandwidth of a concatenated length is insignificant for the case of small parameter mismatch, but may be significant for transverse offset (due to the strong mode mixing). The sensitivity of the splice loss predictions to the choice of the input power distribution may allow this analysis to provide an evaluation of modal power distributions in fibers from experimental measurements of splice loss versus transverse offset.

The particular case of longitudinal offset has been obtained from a generalization of the parameter mismatch theory but has not been treated here due to its relative unimportance in practical systems. Gloge used geometric optics to show that fiber axis tilt at the splice is equivalent to a transverse offset via the relationship $(r_o/a)^2$ is equal to $\sin^2\theta/2\Delta$, where θ is the angle between the fiber axes.¹⁰ This equivalence is also valid from the electromagnetic analysis for pure tilt or offset. The modal analysis of combinations of offset (both transverse and longitudinal) with tilt and parameter mismatch (as discussed in Ref. 10) are subjects of continuing work.

V. ACKNOWLEDGMENTS

We would like to thank D. W. Peckham for providing loss versus transverse offset data; M. J. Saunders for index profile data; and C. M. Miller, P. J. Rich, and R. B. Kummer for helpful discussions.

REFERENCES

1. S. E. Miller and A. G. Chynoweth, Eds., *Optical Fiber Telecommunications*, New York, NY: Academic Press, 1979, p. 74.
2. C. M. Miller and S. C. Mettler, "A Loss Model for Parabolic Profile Fiber Splices," *B.S.T.J.*, 57, No. 9 (November 1978), pp. 3167-80.
3. N. Kashima, "Transmission Characteristics of Splices in Graded-Index Multimode Fibers," *Appl. Optics*, 20, No. 22 (November 15, 1981), pp. 3859-66.
4. D. Marcuse, "Losses and Impulse Response of a Parabolic Index Fiber With Random Bends," *B.S.T.J.*, 52, No. 8 (October 1973), pp. 1423-37.
5. A. W. Snyder, "Excitation and Scattering of Modes on a Dielectric or Optical Fiber," *IEEE Trans.*, *MTT-17* (December 1969), pp. 1138-44.
6. D. Gloge and E. A. J. Marcatili, "Multimode Theory of Graded-Core Fibers," *B.S.T.J.*, 52, No. 9 (November 1973), pp. 1563-78.

7. R. Olshansky, "Mode Coupling Effects in Graded-Index Optical Fibers," *Appl. Optics*, 14, No. 4 (April 1975), pp. 935-45.
8. D. W. Peckham, unpublished work.
9. T. Ohshima et al., "Cutoff Wavelength Measurements of a Multimode Fiber," *Proc. of 7th E.C.O.C., Copenhagen, Denmark, 1981*, pp. 5-1-5-4.
10. D. Gloge, "Offset and Tilt Loss in Optical Fiber Splices," *B.S.T.J.*, 55, No. 7 (September 1976), pp. 905-16.
11. S. C. Mettler, "A General Characterization of Splice Loss for Multimode Optical Fibers," *B.S.T.J.*, 58, No. 10 (December 1979), pp. 2163-82.
11. I. S. Gradshteyn et al., *Tables of Integrals, Series, and Products*, New York, NY: Academic Press, 1980, p. 844, Eqn. 7.414.4.
12. N. N. Lebedev, *Special Functions and Their Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1965, p. 83, Eq. 4.20.3.
13. J. A. Stratton, *Electromagnetic Theory*, New York, NY: McGraw-Hill, 1941, p. 373.
14. A. Erdélyi, Ed., *Higher Transcendental Functions*, The Bateman Manuscript Project, New York, NY: McGraw-Hill, Vol. 2, p. 188.
15. Ref. 11, p. 848, Eqn. 7.422.2.
16. A. Erdélyi, Ed., *Tables of Integral Transforms*, The Bateman Manuscript Project, New York, NY: McGraw-Hill, 1954, Vol. 2, p. 63.

APPENDIX

Derivation of Modal Coupling Coefficients

The principal steps in the derivation of the modal coupling coefficients (eq. 1) presented in Table I are shown in this appendix.

A.1 Intrinsic parameter (V) mismatch

In this case both fibers have fields as defined in eq. (2), where the argument of the field, Z , is

$$Z_T = V_T(r/a_T)^2 \quad (15)$$

and

$$Z_R = Z_T(1 + \epsilon), \quad (16)$$

where

$$V_T = (2\pi/\lambda) \cdot n_{co} \cdot \sqrt{2\Delta_T} \cdot a_T \quad (17)$$

and

$$\epsilon = \left(\frac{V_R}{a_R^2} \right) \left(\frac{a_T^2}{V_T} \right) - 1, \quad (18)$$

where the normalization expressions $A_{m\alpha}$ also contain V_T and V_R .

Examination of the azimuthal integral of eq. (1) shows that, because there is no transverse offset of the fiber axes, azimuthal symmetry is preserved and only modes with the same azimuthal symmetry couple.

Changing the radial integral from an integration over r to one over Z_T , we find that it is in the form of a well-documented integral of Laguerre Gaussian functions,¹¹ giving the expression shown in Table I. The hypergeometric function, ${}_2F_1(-m_2, -m_1, -(m_1 + m_2 + \alpha), y^2)$ is a finite power series in y^2 whose order is the lesser of m_1 or m_2 , since both m_1 and m_2 are integers.

A.2 Transverse offset—identical fibers

The modal fields to be used in eq. (1) can still be expressed by eq. (2) for both fibers. However, the radial and azimuthal variations of the receiving fiber (r_R, ϕ_R) need to be expressed in terms of the transmitting fiber (r_T, ϕ) and the transverse offset, r_o , of the axes. Using

$$r_R^2 = r_T^2 + r_o^2 - 2r_T r_o \cos \phi \quad (19)$$

and the integral relation for Laguerre-Gaussian functions¹²

$$e^{-Z/2} Z^{|\alpha|/2} L_m^{|\alpha|}(Z) = \frac{(-1)^m}{2} \int_0^\infty J_\alpha(\sqrt{xZ}) e^{-x/2} x^{|\alpha|/2} L_m^{|\alpha|}(x) dx \quad (20)$$

we can then use the addition theorem for circular cylindrical waves¹³ to write the receiving fiber field amplitude $E_R(m, \alpha)$ as

$$E_R(m, \alpha) = \frac{(-1)^m}{2} \sum_{p=-\infty}^{\infty} e^{-i(p+\alpha)\phi} \cdot \int_0^\infty J_p(\sqrt{xZ_o}) J_{p+\alpha}(\sqrt{xZ_T}) e^{-x/2} x^{|\alpha|/2} L_m^{|\alpha|}(x) dx, \quad (21)$$

where

$$\begin{aligned} Z_o &= V(r_o/a)^2 \\ Z_T &= V(r_T/a)^2. \end{aligned} \quad (22)$$

The azimuthal integral of eq. (1), for coupling between mode ($m_1\alpha_1$) of the transmitting fiber and mode ($m_2\alpha_2$) of the receiving fiber, reduces the infinite summation over p , introduced by eq. (21), to a single term for

$$p + \alpha_2 = \alpha_1. \quad (23)$$

After a change of variable from r_T to Z_T defined above, the radial integral of eq. (1) can be evaluated by again using eq. (20). (In evaluating this integral it is imperative to realize that the radial field variation of the modes is a function only of the magnitude of the azimuthal mode number, $|\alpha|$.) This evaluation produces two cases, which must be treated separately:

- (i) α_1 and α_2 —both the same sign
- (ii) α_1 and α_2 —different sign.

For both cases the final integral over the dummy variable x , introduced by the Laguerre-Gaussian integral identity [eq. (20)], can be evaluated by using the relationship¹⁴

$$L_n^{-\delta}(x) = (-1)^\delta x^\delta \frac{(n-\delta)!}{n!} L_{n-\delta}^\delta(x) \quad (24)$$

to write the integral in the form

$$I = \int_0^{\infty} y^{p+1} e^{-\alpha y^2} L_m^{p-\sigma}(\alpha y^2) L_n^{\sigma}(\alpha y^2) J_p(x \cdot y) dy, \quad (25)$$

which is a well-documented integral giving

$$I = (-1)^{m+n} (2\alpha)^{-(p+1)} x^p e^{-x^2/4\alpha} L_m^{\sigma-m+n}(x^2/4\alpha) \cdot L_n^{p-\sigma+m-n}(x^2/4\alpha). \quad (26)$$

In both Refs. 15 and 16, the formula is slightly incorrect due to a reversal of the subscripts of the Laguerre polynomials. The formula of Refs. 15 and 16 is seen to be incorrect by noting that eq. (25) above is a Hankel transform of order p , so that, by applying the inverse transform to the result, one should expect to obtain the original function, i.e., the argument of the original Hankel transform. The error persists in Ref. 15 because it used Ref. 16 as its source for this result. The use of this relationship then provides the results in Table I.

Experimental Results of 20-Mb/s FSK Digital Transmission on 4-GHz (TD) Radio

By Y. Y. WANG

(Manuscript received June 9, 1982)

This paper presents the results of tests of 20-Mb/s frequency-shift-keying digital transmission on the 4-GHz TD radio system. On the basis of the test results, the performance of the 20-Mb/s system is projected to be satisfactory to support long-haul digital services. The system employs a 20-Mb/s terminal that multiplexes 12 signals at the first level of digital signal hierarchy (DS-1) (1.5 Mb/s) into a 10-Mbaud, 4-level signal to be transmitted by the Bell System standard 4-GHz (TD) FM microwave radio system. The maximum distance of a digital regeneration span in normal operation is limited by the intermodulation noise to approximately 10 typical hops of TD radio. The 20-Mb/s TD radio system uses the standard frequency-diversity protection switching system, which was designed for analog message service. A fundamental system trade-off is, therefore, the choice of switch threshold: long periods of error-free transmission interrupted by infrequent error bursts due to switch transients versus an occasional low background error rate with less frequent switch transients. We concluded that the protection switch threshold of a 1500-message-circuit channel is suitable for 20-Mb/s TD radio channels.

I. INTRODUCTION

The growth of *Dataphone** Digital Service¹ (DDS)[†] and the needs of new services such as *Picturephone*‡ Meeting Service² (PMS) require a substantial increase in digital long-haul transmission capacity. The existing long-haul digital facilities, mainly Data-Under-Voice³ (DUV), are near exhaustion in many areas. Furthermore, some existing frequency modulation (FM) radios will be replaced by the single-sideband

* Service mark of AT&T.

† Acronyms and abbreviations are defined in the Glossary at the back of this paper.

‡ Service mark of AT&T.

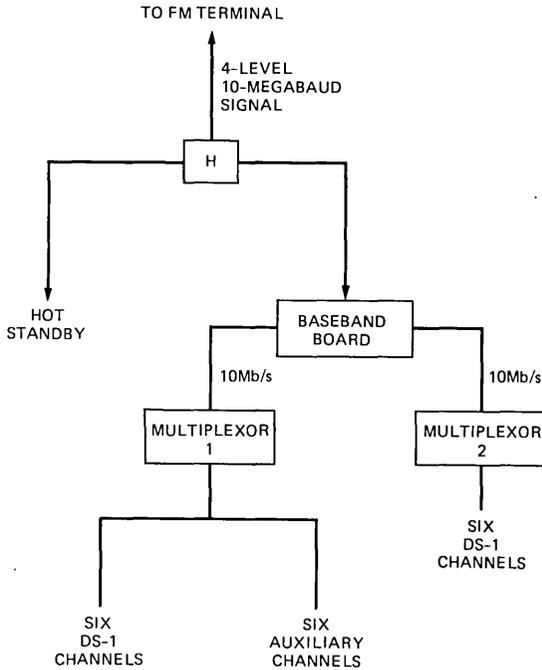


Fig. 1—The 20-Mb/s terminal.

AR-6A⁴ radio to increase analog transmission capability. Such replacement reduces the number of DUV channels. New technologies, such as 45-Mb/s transmission on TD radio^{5,6} (TD-45A^{7,8}) and long-haul digital radio at 4 GHz, will not be available until 1983 and beyond. During the interim, the 20-Mb/s TD system* will be used to provide long-haul digital connectivity.

AT&T Long Lines is deploying an approximately 3000-mile, 20-Mb/s TD network in 1981 and 1982 for DDS application. By the end of 1982, the 20-Mb/s network will have 70 terminals providing 159 digroups. This number of DDS digroups will double the capacity of the 1980 DDS network. The 20-Mb/s terminal has other applications in the Bell System. For example, a portable microwave radio facility can be set up quickly to work with the 20-Mb/s terminals to carry digital services on a temporary basis.

The 20-Mb/s terminal is capable of multiplexing up to 12 asynchronous signals at the first level of digital signal hierarchy (DS-1) into a 9.856-Mbaud, 4-level signal (see Fig. 1). This 4-level signal can be transmitted on the long-haul microwave network using standard FM

* Using the VIDAR DM-12A 20-Mb/s terminal.

terminals and a dedicated 20-MHz bandwidth radio channel suitable for 1500-message-circuit loading. Recommended maximum digital regeneration span is ten TD radio hops (about 250 miles). The existing 100A⁹ or 400A¹⁰ protection switch equipment will protect against propagation fading and radio equipment failures.

Following a summary of the test results in Section 2, the test system configuration and the deployment of test equipment are discussed in Section 3. Section 4 records six tests in the absence of multipath fading, e.g., the jitter performance test, fade margin measurements, and the protection switch compatibility test. Section 5 discusses the performance under the multipath fading condition. The projected performance of the 20-Mb/s TD system for DDS and PMS is discussed in Section 6.

Subsequent to these tests, we found that the preferred power level at the input to the Frequency Modulation Transmitter (FMT) could be reduced by 2 dB, from -14.9 dBm to -16.9 dBm. Combined with the use of a narrower band receiver filter, this power reduction improves the adjacent channel interference and the intermodulation noise performance significantly without degrading the performance of the 20-Mb/s channel. The results presented in this report have been adjusted to reflect the reduced drive level and the tighter receiver filters that are being implemented.

II. SUMMARY

Tests of a 20-Mb/s terminal on a 12-hop, 188-mile TD radio loop were conducted in New Jersey from July to November, 1980. The main results are as follows:

(i) It satisfactorily transmitted the digital signal over 12 hops of TD radio without baseband digital regeneration.

(ii) The 20-Mb/s terminal showed satisfactory jitter performance.

(iii) The system is essentially error-free during normal propagation and operating conditions. System performance as tested in New Jersey is satisfactory for DDS and PMS.

(iv) The required cochannel Carrier-to-Interference Ratio (CIR) into a desired 20-Mb/s TD channel is 25 dB at the protection switching point.

(v) A switch transient of the 100A frequency-diversity protection switch causes a 6 to 30 millisecond burst of errors. The switch threshold of a 1500-message-circuit channel is suitable for the 20-Mb/s TD channel.

Despite the fact that the protection switch transients will cause transmission errors, the tests confirmed that the use of the 20-Mb/s terminal with the TD radio system can meet long-haul transmission requirements.

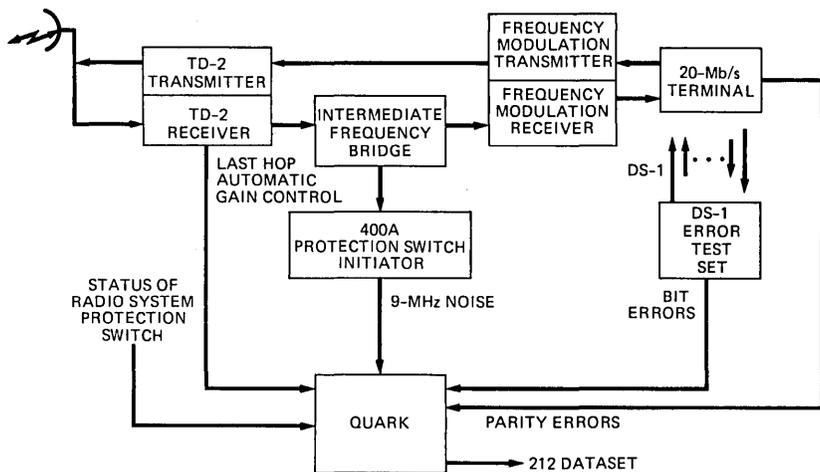


Fig. 2—20-Mb/s TD radio system: test arrangement in Freehold, NJ.

III. TEST DESCRIPTION

The 20-Mb/s TD radio system under test consists of a 20-Mb/s terminal, a 4A FM terminal, a 12-hop loop of TD-2 radio, and a 100A frequency-diversity protection switch (Section 3.1). Two bit-error-rate test sets (Bowmar), an errored-bit accumulator, a 400A protection switch initiator, and a Quantizer, Analyzer, and Record Keeper (QUARK)¹¹ were used to monitor and to record the performance statistics (Fig. 2).

3.1 System configuration

The 20-Mb/s terminal multiplexes six even-numbered and six odd-numbered DS-1 (1.544 Mb/s) channels into two 10-Mb/s rails and then encodes the two rails into a 4-level, 10-Mbaud baseband signal. The 4-level, 10-Mbaud baseband signal is connected to an FM terminal that provides Frequency Shift Keying (FSK) modulation and demodulation. The FM terminal is then connected to the TD-2 radio via Intermediate Frequency (IF) cables. The baseband power level at the input to the FMT and the output of the Frequency Modulation Receiver (FMR) were -16.9 dBm and -0.9 dBm, respectively.

The test radio route consisted of two 2-way, 4-GHz radio channels connecting Freehold (FH) and Cherryville (CH), NJ, as shown in Fig. 3. All radio units used in the test were TD-2, retrofitted with solid-state microwave generators. The radio channels are suitable for 1500-message-circuit loading.

3.2 Deployment of test equipment

The switch initiator of a frequency-diversity protection system de-

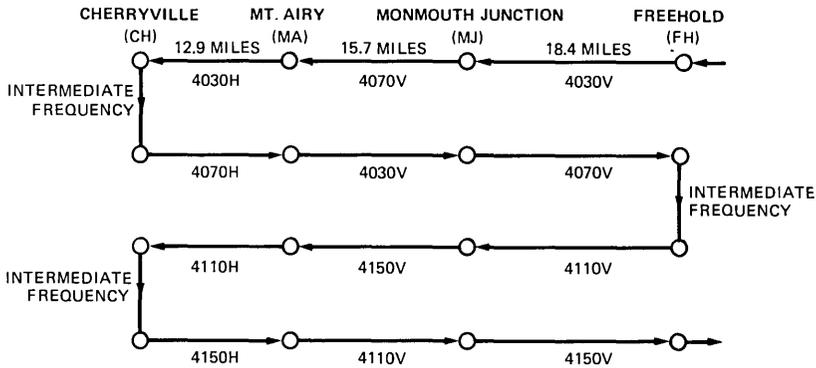


Fig. 3—Frequencies (in megahertz) of TD radio channels for a digital transmission experiment in New Jersey.

tests the noise power in a 1.74-kHz bandwidth centered at 8.9 MHz at the end of a switch section to determine if a working channel needs protection. We therefore installed a 400A protection switch initiator in the receiving IF path to measure the amount of accumulated channel fading over the entire 12-hop route (Fig. 2). For convenience, the power in the 1.74-kHz bandwidth centered at 8.9 MHz will be referred to as the 9-MHz (slot) noise power in the rest of the text.

3.3 Data acquisition system

A QUARK was installed in the Freehold station to record the Automatic Gain Control (AGC) voltage of the last receiving main amplifier and the 9-MHz slot noise of the 12-hop FM radio route, switch activities, errors in two DS-1 channels, and the parity errors in data rail one (9.8 Mb/s, 3672 bits/frame). The statistics of various inputs were accumulated in the QUARK memory. The relationships among DS-1 errors, parity errors, 9-MHz noise, and protection switch activities were studied in terms of statistics of simultaneous events as collected by the QUARK.

IV. TESTS IN THE ABSENCE OF PROPAGATION FADING

Tests were performed to characterize the system performance under normal operating conditions, as well as under controlled stressing conditions. These tests and their results are described in the following sections.

4.1 Baseband spectrum

Figure 4 shows the received baseband spectra of the 20-Mb/s signal measured at FMR-OUT with and without the 12-hop radio loop. The digital spectrum is down about 15 dB at 6 MHz. The 9-MHz noise

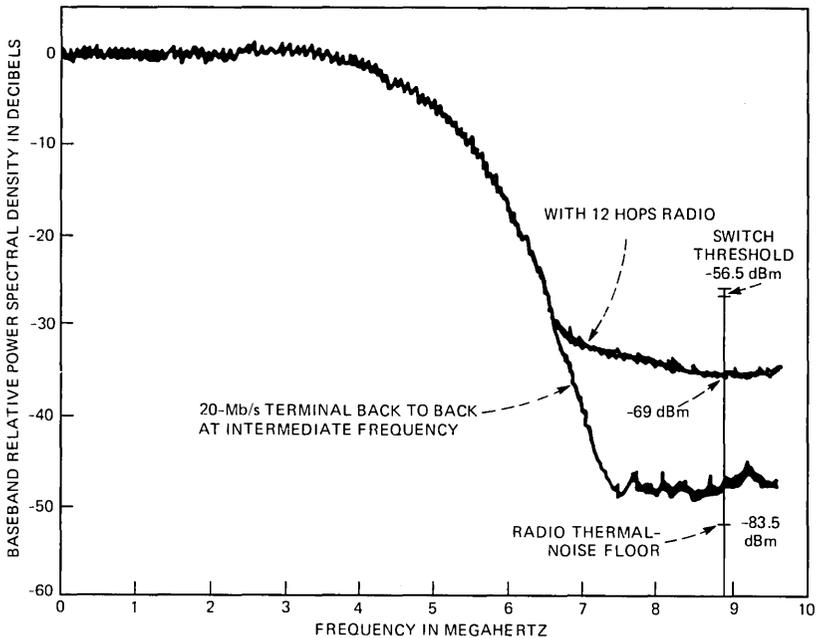


Fig. 4—Baseband spectrum.

power was -69 dBm and -83.5 dBm, with and without the 20-Mb/s modulation, respectively. The 14.5-dB difference represents the 12-hop accumulated intermodulation noise, which dominates the 9-MHz slot noise.

4.2 *S/N at baseband versus BER at DS-1 level*

Baseband noise was injected before the baseband receiver filter to stress the 20-Mb/s transmission. The main objective was to study the Bit Error Rate (BER) versus baseband signal-to-noise ratio (s/n) relationship.

Figure 5 shows the BER performance of the even- and odd-numbered DS-1 channels of the terminal. The odd-numbered channels perform better than even-numbered channels by 1 dB in s/n for equal BER. This difference is due to the circuit design of the terminal. The 20-Mb/s signal consists of two 10-Mb/s rails. At the terminal receiver, the odd-numbered DS-1 channels are derived from a 10-Mb/s rail, which is decoded from the plus-minus sign decision of the received 10-Mbaud, 4-level signal. The even-numbered DS-1 channels are derived from the other 10-Mb/s rail, which is decoded from the amplitude threshold decision of the received 4-level signal. The plus-minus sign decision is more robust than the amplitude threshold decision by 1 dB

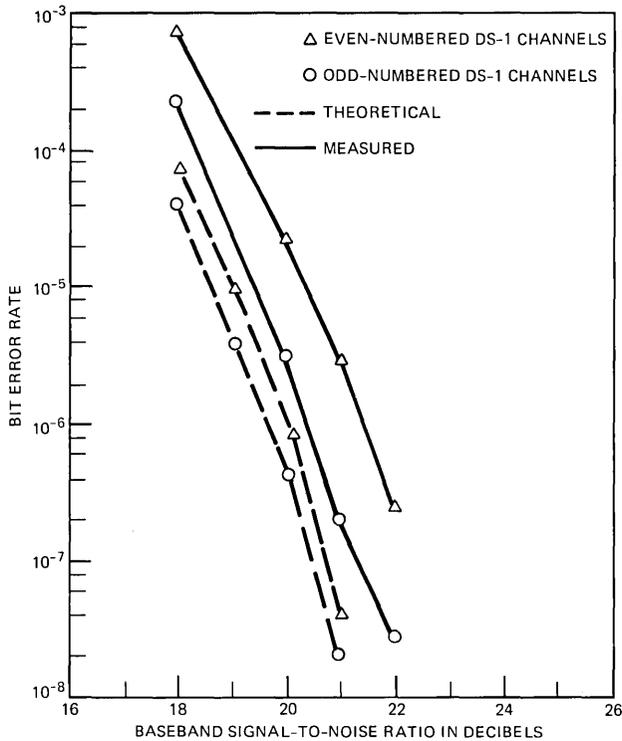


Fig. 5—Bit error rate vs. s/n_{BB} intermediate frequency loopback.

in s/n , as measured, and by 0.5 dB, as theoretically predicted. The theoretical relationship is derived in the appendix. Thus, throughout this study, a discussion of error performance in terms of an even-numbered DS-1 channel implies a conservative (lower) bound on the digital error performance.

Figure 6 shows the measured and calculated BER of the odd-numbered DS-1 channel versus baseband s/n . The 12-hop TD radio degrades the performance of the odd-numbered DS-1 channel by less than 0.5 dB. The effect of 12 hops of TD radio on even-numbered DS-1 channels is practically indiscernible. The maximum regeneration interval (ten TD hops were recommended for the field) is imposed, therefore, by intermodulation noise at the 9-MHz noise slot (Section 4.1), rather than by the digital transmission impairments.

4.3 Jitter performance

Jitter performance of the 20-Mb/s terminal satisfies the requirements of the existing digital network. Under normal operating conditions, the amount of output jitter among DS-1 channels was uniform

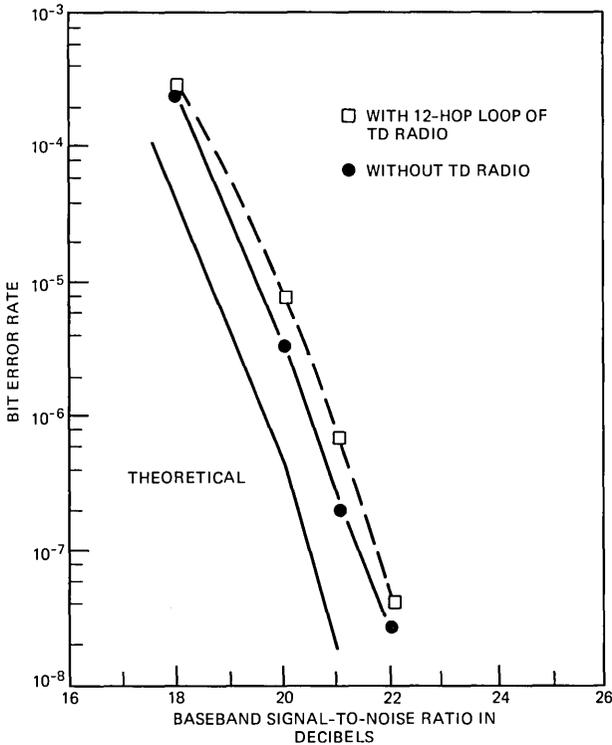


Fig. 6—Bit error rate vs. s/n_{BB} for odd-numbered DS-1 channels.

and comparable to that of the office Quasi-Random-Signal-Source (QRSS), the standard DS-1 signal. In fact, the terminal is effectively a de-jitterizer. The amount of jitter in an output signal was about half that of the input signal, as shown in Table I. Furthermore, digital error performance of two DS-1 channels was compared when the Radio Frequency (RF) signal power was severely attenuated. The channel that took a clean office QRSS as the input consistently made fewer errors than the one that took a jittered source with 11- to 13-percent rms jitter. However, the difference was so small that it was indistinguishable in terms of the fade margin.

4.4 Flat fade margin

The Bell System microwave radio plant was engineered to have adequate flat fade margins to meet the outage objective of allowing less than 0.01-percent outage for all causes over one-way, 4000-mile transmission. The outage of a digital system comprises the time when the one-second-averaged BER exceeds 1×10^{-3} . This objective need

Table I—Jitter performance

	Input to the 20-Mb/s Terminal		Output of the 20-Mb/s Terminal	
	Percent of Jitter (rms)	Percent of Jitter (Peak to Peak)	Percent of Jitter (rms)	Percent of Jitter (Peak to Peak)
1 loop*	13 to 14	60 to 75	5 to 6	25 to 30
2 loops	9 to 11	50 to 65	4 to 5	15 to 25
3 loops	14 to 16	75 to 120	5 to 7	25 to 40
4 loops	15 to 18	90 to 130	7 to 9	35 to 50

* Each loop consists of two complete M13 passes and four hops of digital radio.

Table II—Flat fade margin

Radio Hop	F_s (dB)	F_d (dB)	$F_d - F_s$
FH-MJ	43.0	43.0	0
MJ-MA	39.3	42.7	3.4
MA-CH	44.0	46.0	2.0
CH-MA	45.0	42.8	-2.2
MA-MJ	41.5	42.5	1.0
MJ-FH	41.5	40.0	-1.5
FH-MJ	39.5	41.5	2.0
MJ-MA	39.5	42.8	3.3
MA-CH	40.5	44.0	3.5
CH-MA	40.0	42.3	2.2
MA-MJ	37.8	39.5	1.7
MJ-FH	38.0	40.0	2.0

F_s = switch point fade margin.

F_d = fade margin to 10^{-6} BER.

not be met on a per-hop basis, because of the existence of severe interference conditions at some junction radio stations, as long as the prorated objective on a per switch section basis was satisfied.

This test measured the required amount of RF attenuation at each radio transmitter to reach the protection switch threshold, which is -56.5 dBm noise power at the 9-MHz slot. This test showed that the fade margin against a BER of 1×10^{-3} always exceeded the switch point fade margin, indicating that the 20-Mb/s TD system can be engineered to meet the facility outage objective. In fact, even the margin against a BER of 1×10^{-6} was generally found greater than the corresponding switch point fade margin. The 10^{-6} margin is of interest because the number of seconds at a BER of 1×10^{-6} at the DS-1 level is an estimate of the number of errored seconds (ES), which is an often referenced service performance parameter.

Table II summarizes the measured switch point fade margins and the 10^{-6} margins of the test system. There were two hops where the BER could exceed 1×10^{-6} before the radio channel would request protection. Since this condition is expected at a few junction stations of the network and we wanted to have a "typical" TD route to conduct this test, no effort was invested to identify the cause and eliminate the situation.

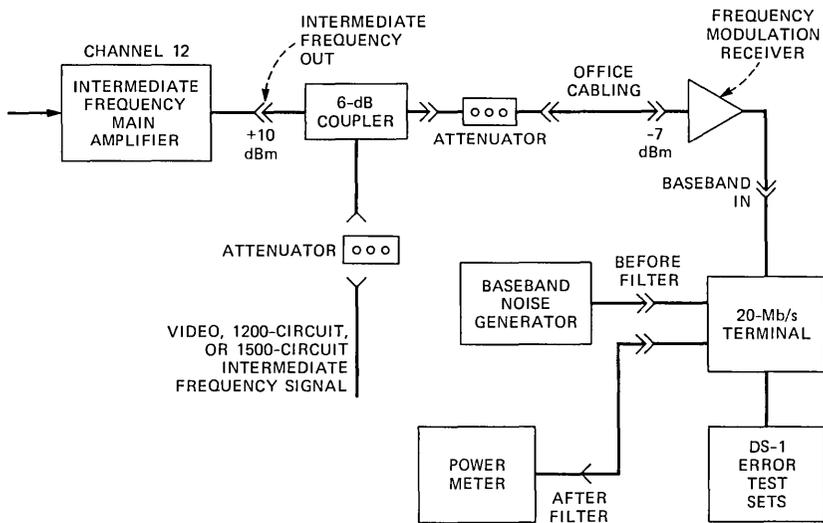


Fig. 7—CIR vs. *s/n* test setup.

4.5 Effect of cochannel interference

To study the impact of severe cochannel interference on the 20-Mb/s TD channel, an interfering signal was injected at IF. Three kinds of interference signals [i.e., 1200-circuit message, 1500-circuit message, and video (color bar)] were used. The test setup is illustrated in Fig. 7. Figure 8 shows the results for a fixed BER of 1×10^{-6} . The color-bar video signal was observed to be the most interfering; the 1500-circuit message was the least interfering among the three. This is due to the higher concentration of spectral energy near the carrier in certain signals. The baseband filter does not reduce this type of signal appreciably.

At the protection switch initiation point, the *s/n* versus CIR relationship was also measured, as shown in Fig. 8. In the region where the faded CIR exceeds 25 dB, as the *s/n* decreases, a protection switch will be initiated before BER degrades to 1×10^{-6} . However, if the faded CIR were less than 25 dB, thermal noise in the radio channel could cause BER to exceed 1×10^{-6} before the switch point. Therefore, severe cochannel interference reduces the effectiveness of the protection switch and degrades the digital error performance.

4.6 Switch system compatibility

For this portion of the test only, a two-way radio channel from Freehold to Cherryville was included in the 100A switching system between these locations. The connection thus had two one-way switch

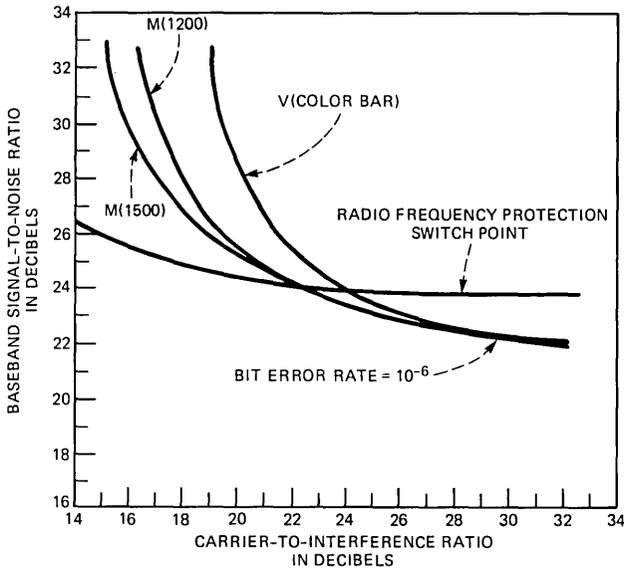


Fig. 8—CIR vs. s/n_{BB} for even-numbered DS-1 channels.

sections, each consisting of three hops of TD radio. Test results include:

(i) A switch cycle (to and back from the protection channel) consists of two transients, i.e., two interruptions in data transmission. More than one switch cycle in a second is possible, and the time span from initial loss of signal to the resynchronization of the receiving part of the 20-Mb/s terminal is approximately 30 milliseconds.

(ii) No special modification to the 100A switch equipment is necessary. The frequency-diversity protection switch is compatible with the 20-Mb/s TD radio system.

V. PERFORMANCE UNDER MULTIPATH FADING CONDITION

5.1 Amount of multipath fading activity

During the period from July 10, 1980 to August 13, 1980, the impact of multipath fading over the test route was studied. Figure 9 shows the statistics of the 9-MHz slot noise during this period. The measured distribution of 9-MHz noise displayed the inverse slope of 10 dB per decade of probability (the L^2 law¹²). This slope is a well-known characteristic of multipath fading for unprotected radio. The number of events was greater than the number of seconds when the 9-MHz noise exceeded a given abscissa. This observation suggests that multiple switch requests could occur in a second. A switch transient causes errors, and, therefore, frequency-diversity protection is expected to

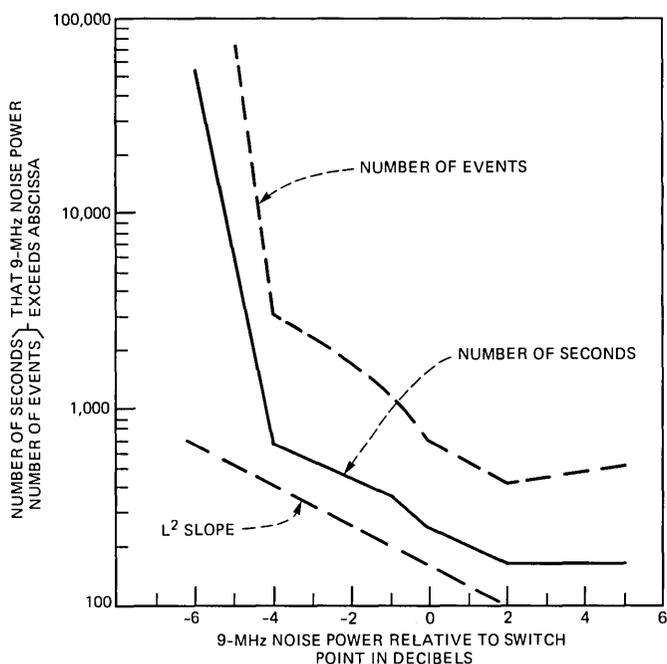


Fig. 9—Distribution of time and events of 9-MHz noise power. The time period monitored was July 10, 1980 through August 13, 1980.

offer limited reduction in the number of errored seconds during multipath fading periods.

The fading statistics of a single hop from Monmouth Junction (MJ) to Freehold are shown in Fig. 10. The solid curve in Fig. 10 also has an inverse slope of 10 dB per decade of probability. The engineering model¹² predicted 166 seconds below 30 dB for this hop during a heavy fading month. We recorded 115 seconds, which is about two thirds of a heavy fading month. Hence, we did experience a fair amount of multipath fading during the test.

5.2 Digital error performance

The measured distribution of BER of an even-numbered DS-1 channel is shown in Fig. 11. There were 432 errored seconds (ES); 263 of them have BER exceeding 10^{-3} . The BER statistics for simultaneous errored seconds of even-numbered and odd-numbered DS-1 channels were also plotted in Fig. 11. They differ in the number of low-BER ES, as expected (see Section 4.2 and the appendix). The number of ES in even-numbered DS-1 channels is approximately 25 percent more than the ES in odd-numbered DS-1 channels.

The data format of the terminal multiplexor contains one parity bit

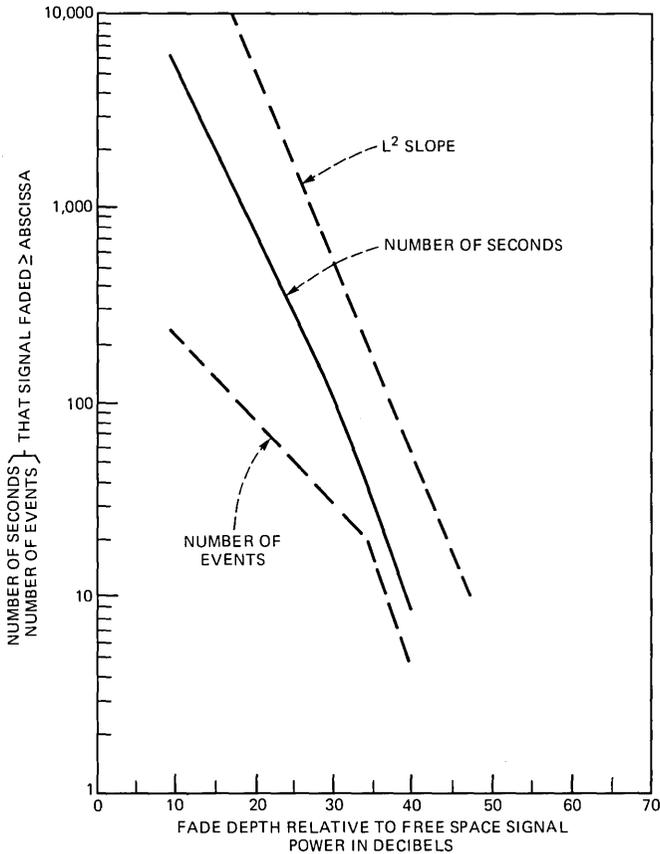


Fig. 10—Distribution of time and events of received signal power. The time period monitored was July 10, 1980 through August 13, 1980.

in a 3672-bit frame. Fault-detection algorithms internal to the terminal and a real-time error-performance monitoring plan to be used by the AT&T Long Lines operations personnel depend on the detection of the parity violations. The distribution of parity violations in a 9.8-Mb/s rail (which feeds six DS-1 channels) can be found in Fig. 12. Those parity violation seconds occurring simultaneously with the ES for the information bits of one even-numbered DS-1 channel are also plotted in Fig. 12. The difference between the two was found to be insignificant. The parity violation seconds offers an effective representation of the real DS-1 ES.

5.3 Relationship between digital errors and 9-MHz noise

Figure 13 shows the distribution of BER of an even-numbered DS-1 channel and the portion of the errored seconds with 9-MHz noise

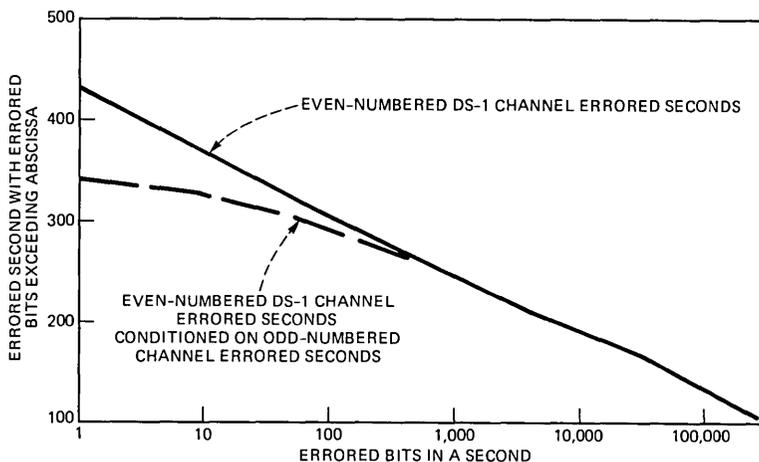


Fig. 11—DS-1 errored-second statistics for the time period July 10, 1980 through August 13, 1980.

exceeding switching point from August 14, 1980 to September 20, 1980. About 10 percent of the errored seconds occurred before the 9-MHz noise reached the switching threshold. All ES with more than 100 errors occurred at the same time the 9-MHz noise exceeded the switch point. Thus, these errored seconds are likely to be accompanied by switch activities in the plant environment.

5.4 Impact of frequency-diversity protection

Based on the assumption that the protection channel is always available for error-free transmission, this section will show that:

- (i) The switching threshold of a 1500-message-circuit loaded channel is suitable for a 20-Mb/s TD channel, and
- (ii) During multipath fading periods, the protection switch would have offered little improvement in terms of ES reduction owing to the frequent switching activities.

Under the assumption of perfect frequency-diversity protection, those errored seconds that occurred while the 9-MHz noise exceeded the switching threshold would have been prevented. The exception would be those seconds in which the 9-MHz noise passes through the switching threshold. This is because the actual switch transfer causes an errored second. Based on the observation that a switch cycle could be completed within a second (Sections 4.6 and 5.1), we estimate the number of seconds containing switching activities by the number of seconds in which the maximum value of the measured 9-MHz noise in a second exceeded the switching threshold. Hence, if the test route had been frequency-diversity protected, we would have X ES, where

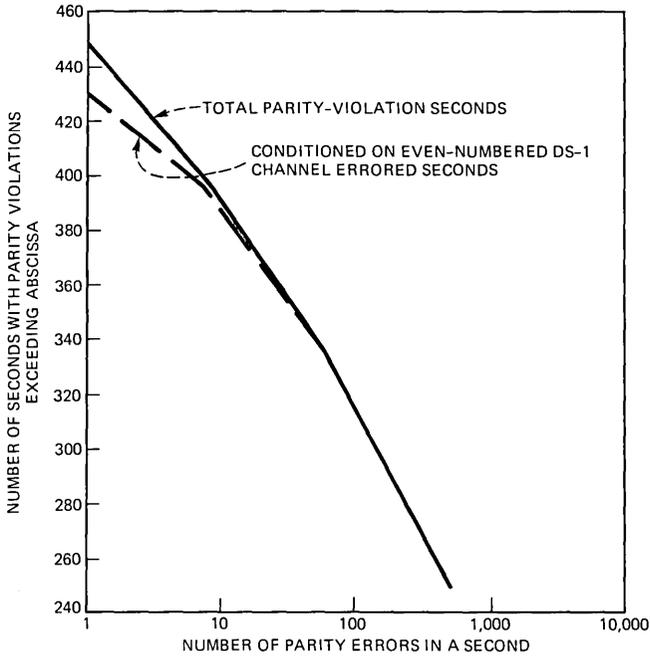


Fig. 12—Parity-violation-second statistics for the time period July 10, 1980 through August 13, 1980.

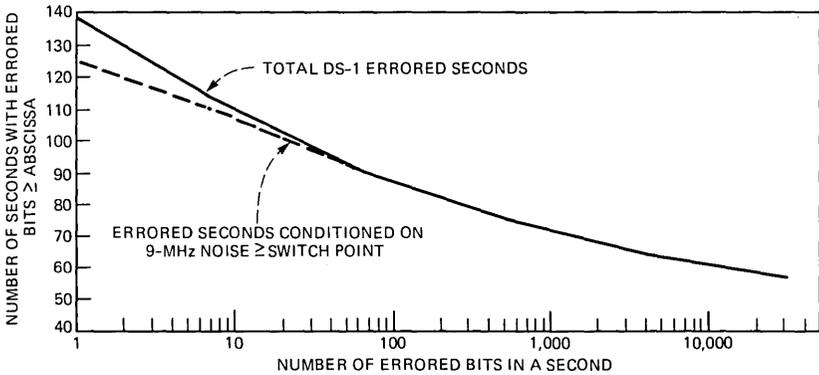


Fig. 13—DS-1 errored-second statistics for the time period August 14, 1980 through September 20, 1980.

$$\begin{aligned}
 X(\text{at a given threshold}) = & \\
 & (\text{Total ES}) - (\text{ES conditioned on 9-MHz noise power} \geq \text{threshold}) \\
 & + (\text{number of seconds with maximum 9-MHz noise power} \\
 & \geq \text{threshold}).
 \end{aligned}$$

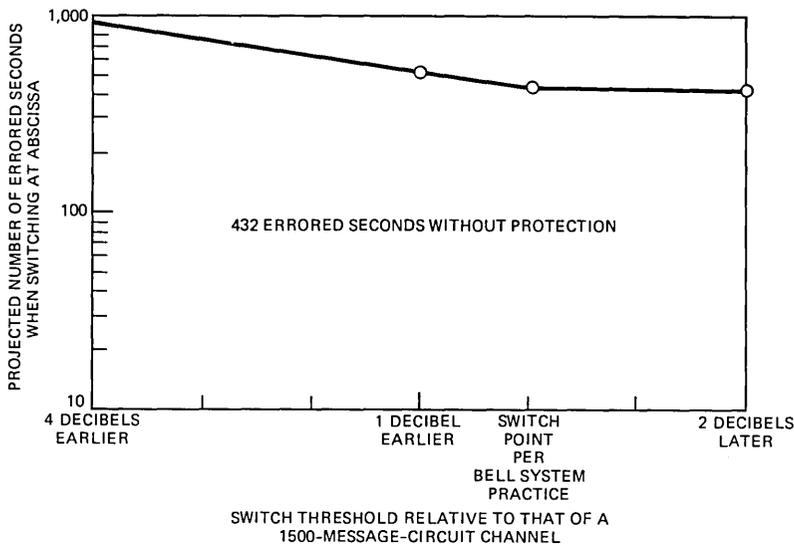


Fig. 14—Sensitivity of switch threshold to errored-second performance projection based on data collected July 10, 1980 through August 13, 1980.

Figure 14 shows the projected number of ES as a function of switching threshold. The projected number of ES decreases as the switch threshold decreases until the switch threshold reaches the level of a 1500-message-circuit loaded channel. Since the 20-Mb/s channel will coexist with 1500-message-circuit loaded channels in the plant, it is desirable from operations and maintenance considerations to use the switch threshold of a 1500-message-circuit loaded channel for a 20-Mb/s TD channel.

VI. PERFORMANCE PROJECTION FOR DDS AND PMS

A properly engineered 20-Mb/s TD system can meet the long-haul outage and quality objectives for DDS and PMS. The long-haul outage objective^{13,14} for DDS or PMS is the same as that for Message Telecommunications Service (MTS). Engineering guidelines for TD radio were developed to meet the MTS outage objective and, therefore, those for DDS and PMS as well.

The following demonstrates the ability of the 20-Mb/s TD system to meet the long-haul quality objectives for DDS and PMS.

6.1 Meeting the DDS quality objective

6.1.1 The DDS quality objective

The quality objective of the DDS transmission design requires 99.5-percent error-free-seconds at DS-0 (56-kb/s rate) for 4000-mile, one-

way, end-to-end transmission. The test results showed that 90 percent of the errored seconds would be associated with switch transients, which cause errors in all 23 DS-0 channels of a DS-1 signal. We, therefore, assume as a worst case that the DDS quality objective for the 20-Mb/s TD system is the same at DS-0 and DS-1 levels. The objective is thus to have less than 0.5-percent ES for a one-way, 4000-mile transmission.

6.1.2 Generalized performance projection

The majority (ninety percent in this test) of ES of a properly engineered 20-Mb/s TD radio system are expected to be caused by protection switch transients. Therefore, to project the performance of a 20-Mb/s TD system, the statistics of switch activities in a 4000-mile TD route must be considered.

Every switch system (e.g., 100A, 400A) has a switch register that counts the number of automatic switch completions. Any manually forced protection switches for routine maintenance are not included in these counts. AT&T Long Lines records the number of switches on a weekly basis. These records of switch completions from June 16, 1979 through June 7, 1980 on five TD routes (New York City to Boston, New York City to Philadelphia, Philadelphia to Silver Spring, Pittsburgh to Silver Spring, and Chicago to Kalamazoo) suggest that there are 13,500 switch completions per year for an average one-way radio channel for multipath fading protection in a 4000-mile TD radio system. There could be two errored seconds associated with a switch cycle; therefore, 27,000 ES due to multipath fading are expected annually.

The one-year switch register data also revealed that, in addition to the switchings due to multipath fading, there were protection switches, called the background switch activities, which were attributed to craft activities, hardware problems, and other causes. These background switches are of a transient nature. A switch cycle lasts much less than a second. According to the same database, this type of switch amounted to 0.03 switch completion per mile per week per channel on average. Thus, we project 6240 ES due to background switches.

Table III summarizes the projected number of ES in a year due to all causes. The total number of ES is 34,078 which is 0.11 percent of a year. This projection compares favorably with the objective of 0.5 percent and shows that the 20-Mb/s TD system performance is satisfactory for DDS.

6.2 Meeting the PMS quality objective

We have observed essentially error-free performance without fading. Multipath fading seldom occurs during an 8 am to 9 pm business day,¹²

Table III—Projection of errored seconds caused by switching in a one-way, 4000-mile TD route in a year

Cause	Errored seconds
FMT or 20-Mb/s terminal failure	18
Routine maintenance	720
Radio failure	100
Background switch activity	6240
Multipath fading	27,000
Total	34,078
	=0.11 percent/year

when most PMS calls will take place. Therefore, based on the rate of background switches, as discussed in the previous section, we project that the 20-Mb/s system performance is satisfactory for PMS.

VII. ACKNOWLEDGMENT

Many colleagues at Bell Laboratories and AT&T Long Lines offered their prompt assistance at every stage of this project. I am particularly grateful to L. A. Dietrich and S. H. Lin for their participation and technical guidance throughout the project. The QUARK was designed by T. G. Szekeres and G. A. Zimmerman.

REFERENCES

1. Special issue on "Digital Data System," B.S.T.J., 54, No. 5 (May-June 1975).
2. H. S. London and D. B. Menist, "A Description of the AT&T Video Teleconferencing System," Conf. Rec. Nat. Telecommun. Conf., 3 (November 1981), p. F5.4.
3. R. R. Grady and J. W. Knapp, "1A Radio Terminals Put Data Under Voice," Bell Lab. Rec., 52 (May 1974), pp. 161-6.
4. J. Gammie and J. P. Moffatt, "The AR-6A Single Sideband Microwave Radio System," Conf. Rec. Int. Conf. Commun., 1981, pp. 3.5.1-5.
5. A. A. Roetken, K. D. Smith, and R. W. Friis, "The TD-2 Microwave Radio System," B.S.T.J., 30, No. 4 (October 1951), pp. 1041-77.
6. S. D. Hathaway, W. G. Hensel, D. R. Jordan, and R. C. Prime, "TD-3D Microwave Radio Relay System," B.S.T.J., 47, No. 7 (September 1968), pp. 1143-88.
7. R. T. Cooney and S. A. Harvey, "TD-45A, a System for 45-Mb/s Digital Transmission over the TD Radio Network," GLOBECOM 1982, pp. D3.6.1-4.
8. W. P. Jaeger, P. L. Penny, and S. L. Vonrump, "Measured Performance of 45-Mb/s Transmission over the Bell System 4-GHz Radio," GLOBECOM 1982, pp. D3.7.1-4.
9. H. D. Griffiths and J. Nedelka, "100A Protection Switching System," B.S.T.J., 44, No. 10 (December 1965), pp. 2295-336.
10. H. D. Griffiths, "The 400A Microwave Radio Protection Switching System," Conf. Rec. Int. Conf. Commun., 1973, pp. 28.12-16.
11. N. F. Dinn, "Facilitating the Use of Digital Radio," Bell Lab. Rec., 57 (May 1979), pp. 135-41.
12. W. T. Barnett, "Multipath Propagation at 4, 6, and 11 GHz," B.S.T.J., 52, No. 2 (February 1972), pp. 321-61.
13. A. Vigants and M. V. Pursley, "Transmission Unavailability of Frequency-Diversity Protected Microwave FM Radio Systems Caused by Multipath Fading," B.S.T.J., 58, No. 8 (October 1979), pp. 1778-96.
14. W. Y. S. Chen, "Estimated Outage in Long-Haul Radio Systems with Protection Switching," B.S.T.J., 50, No. 4 (April 1971), pp. 1455-85.
15. R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication*, New York, NY: McGraw-Hill, 1968.

APPENDIX

Predicted Error Performance With Ideal Channel Plus Gaussian Noise

This appendix approximates a set of relationships between a 1-second bit error rate at DS-1 level and the baseband signal-to-noise ratio based on our understanding of the 20-Mb/s terminal and an idealized FM channel with Gaussian noise.

There are two multiplexor units in the main/working portion of a terminal. The number 1 multiplexor unit accepts data from the six odd-numbered DS-1 channels plus the six auxiliary (AUX) channels. The number 2 multiplexor unit accepts data from six even-numbered DS-1 channels. Each multiplexor combines its inputs into a single data-bit stream of 9.856 MHz. These two bit streams form binary groups: the first digit in the binary group is taken from the number 1 multiplexor, the second from the number 2 multiplexor. These binary groups are Gray coded and are then converted into a single 4-level signal. It has been verified in the laboratory that the four levels at the transmitter output are indeed equally spaced.

The decoder at the terminal receiver performs the inverse operation. The probabilities of error for even- and odd-numbered DS-1 channels separately can be derived in a manner similar to the analysis of Lucky¹⁵ and others under the following assumptions:

1. The amplitudes of the four symbols are equally likely to assume any of the four equally spaced values $\pm d$ and $\pm 3d$. Symbols occurring at different times are independent.
2. The bit error performance of the terminal depends on s/n but is insensitive to the spectral shape of noise. (This has been verified in the laboratory.) We assume that the additive noise is Gaussian.
3. The pulse shaping $X(w)$ is raised cosine with 50 percent roll-off:

$$X(w) = \begin{cases} T & 0 \leq w \leq \frac{\pi}{2T} \\ \frac{T}{2} (1 + \sin wT) & \frac{\pi}{2T} \leq w \leq \frac{3\pi}{2T} \end{cases}$$

where T is the baud interval.

The terminal puts all pulse shaping at the transmitter. Therefore, the signal power input to the FM channel is:

$$\begin{aligned} p_s &= \frac{\overline{\alpha^2}}{2\pi T} \int_{-\infty}^{\infty} |X(w)|^2 dw \\ &= \frac{35}{8} d^2, \end{aligned}$$

where $\overline{\alpha^2}$ is the average symbol power, $\overline{\alpha^2} = 5d^2$.

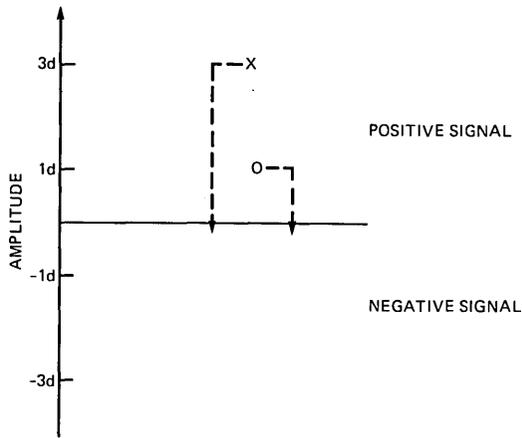


Fig. 15—Error occurrence in an odd-numbered DS-1 channel.

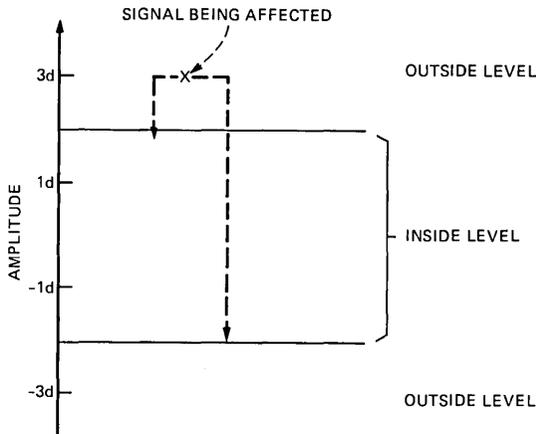


Fig. 16—Error occurrence of an outside-level signal in an even-numbered DS-1 channel.

The receiver detects the signal levels and places slicing levels at 0 and $\pm 2d$. An error occurs when the noise at a sampling time pushes the received signal amplitude (voltage) across the slicing levels. The slicing level drift due to noise is assumed insignificant. The probability of a signal being at one of the outside two levels equals that of being at one of the inside two levels. A signal at the outside level, or inside level, can only cross the zero level in one direction when the noise voltage $|y|$ exceeds $3d$ or d , respectively, as illustrated in Fig. 15. Therefore, the probability of error for an odd-numbered channel is:

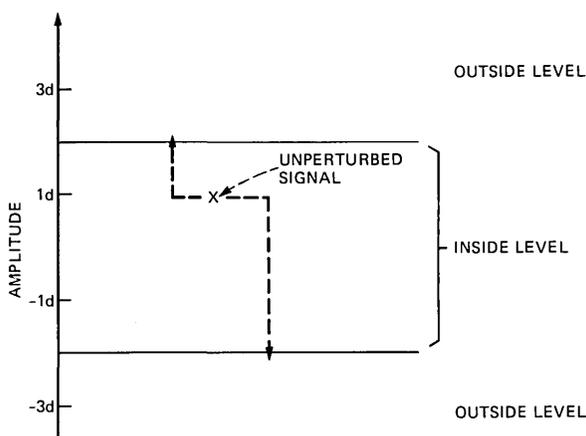


Fig. 17—Error occurrence of an inside-level signal in an even-numbered DS-1 channel.

$$p^{\text{odd}} = \frac{1}{2} \cdot \frac{1}{2}p(|y| > 3d) + \frac{1}{2} \cdot \frac{1}{2}p(|y| > d), \quad (1)$$

where $p(|y| > d)$ represents the probability of noise voltage exceeding d .

The probability of error for an even-numbered channel, p^{even} , can be derived similarly. The probability of a signal in the outside state is $\frac{1}{2}$. For that signal to be in error, the noise in one direction only with a magnitude $|y|$,

$$5d > |y| > d$$

is required, as illustrated in Fig. 16. The probability of a signal in an inside level crossing a $\pm 2d$ slicing level is $p(|y| > d) - \frac{1}{2}p(3d > |y| > d)$, as shown in Fig. 17. Therefore,

$$\begin{aligned} p^{\text{even}} &= \frac{1}{4}p(5d > |y| > d) \\ &\quad + \frac{1}{2}p(|y| > d) - \frac{1}{4}p(3d > |y| > d) \\ &= \frac{1}{4}p(|y| > d) - \frac{1}{4}p(|y| > 5d) \\ &\quad + \frac{1}{2}p(|y| > d) \\ &\quad - \frac{1}{4}p(|y| > d) + \frac{1}{4}p(|y| > 3d) \\ &= \frac{1}{2}p(|y| > d) + \frac{1}{4}p(|y| > 3d) \\ &\quad - \frac{1}{4}p(|y| > 5d). \end{aligned} \quad (2)$$

This probability is easily computed since the noise at the receiver input is assumed Gaussian. The probability of noise voltage exceeding

d is conveniently expressed in terms of the normal probability integral, $p(|y| > d) = 2Q\left(\frac{d}{\sigma}\right)$, where

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-t^2/2} dt$$

and σ is the root-mean-square (rms) noise voltage.

The signal to noise power ratio, p_s/p_N , can be expressed as

$$\frac{p_s}{p_N} = \frac{35d^2}{8\sigma^2}.$$

Hence, eqs. (1) and (2) become

$$p^{\text{odd}} = \frac{1}{2} \left\{ Q \left[3 \left(\frac{8X}{35} \right)^{1/2} \right] + Q \left[\left(\frac{8X}{35} \right)^{1/2} \right] \right\}$$

$$p^{\text{even}} = \frac{1}{2} Q \left[\left(\frac{8X}{35} \right)^{1/2} \right] \left\{ 1 - 2Q \left[5 \left(\frac{8X}{35} \right)^{1/2} \right] \right\}$$

$$+ Q \left[\left(\frac{8X}{35} \right)^{1/2} \right]$$

$$- \frac{1}{2} Q \left[\left(\frac{8X}{35} \right)^{1/2} \right] \left\{ 1 - 2Q \left[3 \left(\frac{8X}{35} \right)^{1/2} \right] \right\},$$

where $X = \frac{p_s}{p_N}$.

A digital computer was used to compute the probabilities of error as a function of the baseband signal-to-noise ratio. Results are discussed in Section 4.2.

GLOSSARY

AGC	Automatic gain control circuitry
AM	Amplitude modulation
AR-6A	A Western Electric 6-GHz single-sideband AM microwave radio system
BER	Bit error rate
CIR	Carrier-to-interference power ratio, expressed in dB
DDS	<i>Dataphone</i> [®] Digital Service, a synchronous full-duplex digital service at 2.4-, 4.8-, 9.6-, and 56-kb/s rates on point-to-point and multipoint bases
DS-0	Digital signal at the 0th level of the TDM hierarchy, the DS-0 level; a signal at the 64-kb/s rate, the DS-0 rate

DS-1	Digital signal at the 1st level of the TDM hierarchy, the DS-1 level; a signal at the 1.544-Mb/s rate, the DS-1 rate
DUV	Data under voice, a system that provides for the transmission of one DS-1 signal over an FM microwave radio link. (This system is also known as 1A radio digital system.)
ES	Errored second; a second that contains at least one errored bit
EFS	Error free second
FM	Frequency modulation
FMT/FMR	Frequency modulation terminal transmitter/receiver
FSK	Frequency shift keying
IF	Intermediate frequency (70-MHz \pm 10-MHz for TD radio)
MTS	Message Telecommunications Service
PMS	<i>Picturephone</i> [®] Meeting Service, a switched, common-user, interactive visual and audio teleconferencing service offered between two remote conference room locations
s/n	Signal-to-noise ratio in dB
TD-45A	A system for 45-Mb/s digital transmission over the TD radio network
TDM	Time division multiplexing, the process of combining a number of digital signals into a single digital stream by an orderly assignment of time slots
TD-2	A Western Electric point-to-point 4-GHz microwave radio transmission system

An Experimental Broadband Imaging Feed

By T. S. CHU and R. W. ENGLAND*

(Manuscript received November 8, 1982)

Imaging with Fresnel diffraction taken into account is applied to the design of a broadband narrow pattern feed. This feed is not only a basic building block of an imaging beam waveguide, but also essential for an offset reflector antenna of large effective F/D ratio. Furthermore, it can be used as a constant beamwidth radiometer antenna for multifrequency remote sensing. We have built and tested a practical example that consists of an offset ellipsoidal reflector and a corrugated horn. Measured amplitude and phase patterns agree with calculated results, which include truncation effects. Systematic design procedures are obtained for a given feed horn and the required reflector illumination. Necessary and sufficient conditions of the thin lens model are translated into design parameters of an offset ellipsoidal reflector with projected circular aperture. Geometrical relations of the offset ellipsoid and calculations of radiation patterns are described in the appendices.

I. INTRODUCTION

The successful performance of an offset dual-reflector antenna often depends upon illumination by a broadband feed with a narrow feed pattern.¹ For example, a broadband corrugated horn with the required pattern could be used, but it would be excessively long for most practical applications. Good illumination can also be achieved by a narrowband offset launcher,² which is essentially an offset reflector fed by a relatively short feed horn. Excellent 19/28.5 GHz dual-frequency performance was demonstrated by the Crawford Hill 7-meter antenna using a quasi-optical frequency diplexer to combine two narrowband offset launchers.¹ However, this approach is not usually cost effective, especially for lower frequency systems. In this paper we present design

* Now at AT&T Long Lines.

procedures for *broadband* offset launchers and comparisons between calculated and measured patterns of both amplitude and phase. These feeds represent a considerable simplification of the hardware, and thus improve the prospect for system application of offset dual-reflector antennas.

The following analysis will assume the thin lens approximation for an imaging offset reflector. It will also make use of the frequency insensitivity of the field distribution in the aperture of a corrugated horn.^{3,4} The principles of imaging have been discussed previously in connection with frequency-independent far-field beamwidths,⁵ and the laws of geometrical optics have been used for the design of imaging reflector arrangements.⁶ However, broadband feed illumination for reflector antennas needs not only constant beamwidth, but also constant phase center. The condition for satisfying the latter requirement is obtained in this paper by an interpretation of the additional phase shift⁷ due to Fresnel diffraction. The imaging laws of geometrical optics are thus modified to provide a theoretically frequency-independent design of offset-launcher feeds. However, the practical bandwidth will be limited by the corrugated horn and truncation effects. These effects will be examined by numerical calculations and experimental measurements.

The imaging feed discussed here has important potential for application in radio communication and other scientific explorations. Furthermore, the basic properties of single-stage imaging are of vital interest in the design of multistage-imaging beam waveguides.⁸ One notes the difficulty of performing pattern measurements of a bulky beam waveguide assembly. Both single- and double-stage imaging are also special cases of a proposed technique⁹ for broadband astigmatic compensation in which the image of a corrugated horn through two astigmatic lenses can be designed to produce a specified astigmatic illumination.

Section II will discuss the imaging feed within thin lens approximation. Section III will describe how to translate the design parameters from the thin lens model to a practical offset ellipsoidal reflector with the required projected circular aperture. Section IV will give comparisons between calculated and measured data for both amplitude and phase patterns of an experimental broadband imaging feed. Geometrical relations of an offset ellipsoid and a calculation of radiation patterns will be given in Appendices A and B, respectively.

II. THIN LENS MODEL

It is well known that the aperture distribution of a corrugated horn with radius a_0 is the Bessel function $J_0(\alpha\rho/a_0)$.^{3,4} The edge field vanishes when $\alpha = 2.405$. This normally occurs at the design frequency

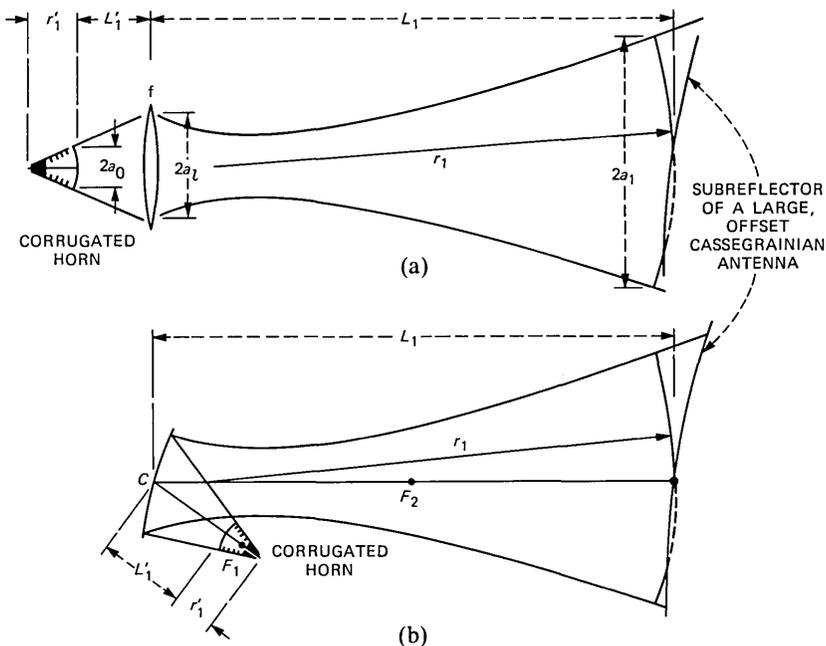


Fig. 1—Sketch of a broadband imaging feed. (a) Thin lens approximation. (b) Offset ellipsoidal reflector.

for resonant corrugations. However, α remains close to the above optimum value over a broad bandwidth if the aperture diameter is about five wavelengths or greater. If a magnified image of the corrugated-horn-aperture distribution is used to illuminate a reflector antenna, the illumination is expected to be a truncated Bessel function. For a specified reflector edge taper of T (in dB), the radius a_r of the illumination circle at the reflector aperture should have a value satisfying

$$20 \log_{10} J_0(\alpha a_r / a_1) = -T, \quad (1)$$

where $a_1 = M a_0$ is the magnified corrugated-horn-aperture radius. The magnification is given by

$$M = \frac{L_1}{L'_1}, \quad (2)$$

where L_1 and L'_1 are the distances of the illuminated reflector and the corrugated horn from the imaging lens, respectively, as shown in Fig. 1.

Now the focal length, f , of the imaging lens should obey the thin

lens formula:

$$\frac{1}{f} = \frac{1}{L_1} + \frac{L}{L'_1}. \quad (3)$$

Furthermore, the image-illumination phase-front radius of curvature r_1 can be expressed in terms of the feed-horn-aperture phase-front radius of curvature r'_1 as

$$\frac{1}{r_1} = \frac{1}{L_1} \left[1 + \frac{L'_1}{L_1} \left(1 + \frac{L'_1}{r'_1} \right) \right]. \quad (4)$$

Equation (4), which is the same as the relation for Gaussian beam imaging,^{8,9} is essentially an alternative form of eq. (7) in Ref. 7. This equation represents the additional phase shift due to Fresnel diffraction.

Equation (4) shows that r_1 will be always less than L_1 if all parameters on the right side are positive. r_1 may become greater than L_1 for negative r'_1 . One notes that r'_1 is always positive for any corrugated-feed horn unless modified by another lens or offset reflector. To obtain some feeling about the required physical spacings for a given pair of specified illumination and feed horn, normalized L_1 (with respect to r_1) has been plotted from eq. (4) versus normalized L'_1 (with respect to r'_1) for several ratios of L'_1/L_1 in Fig. 2. For a given pair of specified illumination and corrugated horn, i.e., L'_1/L_1 , r_1 and r'_1 specified, it is also convenient to rearrange eq. (4), as follows:

$$\frac{L_1}{r_1} = \frac{1 + \frac{L'_1}{L_1}}{1 - \left(\frac{L'_1}{L_1} \right)^2 \frac{r_1}{r'_1}}. \quad (5)$$

It is of interest to note a few special cases of eq. (4). The near-field gregorian configuration⁶ corresponds to $r'_1 = \infty$ in eq. (4). J. A. Arnaud's confocal feed reflector arrangement¹⁰ is also a special case in which $r'_1 = -L'_1$ and $r_1 = L_1$. However, Arnaud assumed both constant beamwidth and constant phase center for the corrugated horn to obtain a frequency-independent aperture distribution at his first reflector.

If the radiation from a corrugated horn is approximated by a Gaussian beam, the $1/e^*$ beam radius at the horn aperture has been shown⁵ to be $w'_1 = 0.64a_o$, where a_o is the horn-aperture radius. Then

* Here $e = 2.71828$, whereas e represents eccentricity of a conic in the next section and appendices.

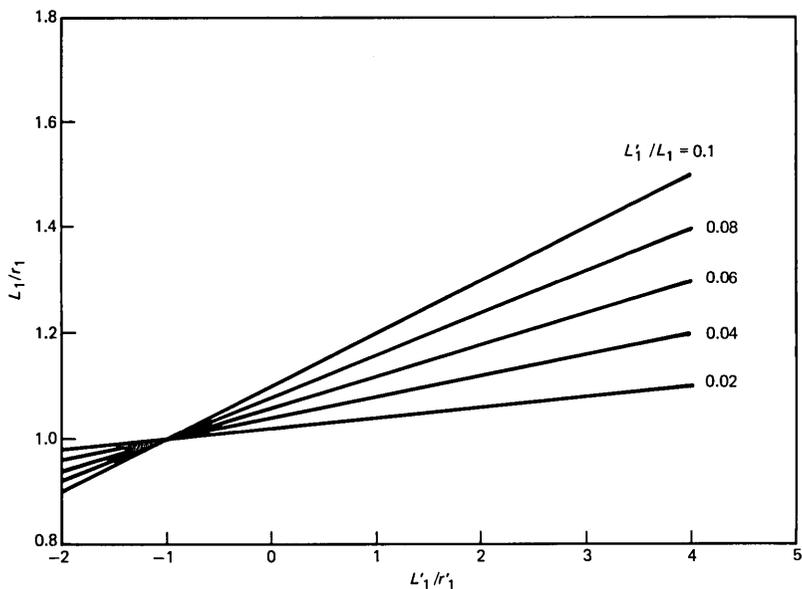


Fig. 2—Relation between image and object phase curvatures for various ratios of L_1'/L_1 .

the Gaussian beam radius w_2 at the imaging lens can be found from¹¹

$$w_2 = w_1' \sqrt{\left(\frac{L_1'}{r_1'} + 1\right)^2 + \left(\frac{\lambda L_1'}{\pi w_1'^2}\right)^2}. \quad (6)$$

The lens diameter should be 3.04 times the Gaussian beam radius w_2 for a truncation edge taper of -20 dB.

III. OFFSET ELLIPSOIDAL REFLECTOR WITH PROJECTED CIRCULAR APERTURE

In practice, the thin lens will be approximated by an offset ellipsoidal reflector, as shown in Fig. 1(b). The distances from the center of the reflector to the two foci, F_1 and F_2 , are, respectively, the incident and reflected phase-front radii of curvature. The incident phase-front radius of curvature can be obtained from the Gaussian beam propagation formula¹¹

$$R_1 = L_1 \left/ \left[1 - \frac{\frac{L_1'}{r_1'} + 1}{\left(\frac{L_1'}{r_1'} + 1\right)^2 + \left(\frac{\lambda L_1'}{\pi w_1'^2}\right)^2} \right] \right. \quad (7)$$

If the offset ellipsoid is in the far zone of the horn, R_1 becomes the distance CF_1 from the phase center of the horn to the center of the reflector. The thin lens formula yields $R_2 = CF_2$, the reflected phase-front radius of curvature

$$\frac{1}{R_1} + \frac{1}{R_2} = \frac{1}{f}. \quad (8)$$

The frequency-dependent R_1 and R_2 are often not the same as L_1 and L'_1 in eq. (3). The equivalent focal length f is identical to that in eq. (3). It has been shown⁸ that the second-order terms of the offset ellipsoidal surface are only functions of f and the angle of incidence, whereas the third-order terms are also dependent upon R_1 and R_2 .

The design of an offset ellipsoidal reflector with oversized aperture of rectangular shape was discussed in Ref. 8. Since corrugated feed horns and required reflector illuminations are often of circular shape, we shall now describe the design of offset ellipsoidal reflectors with projected circular apertures.

The intersection of an ellipsoid and a circular cone subtended at one focus is a plane ellipse subtended by another circular cone at the other focus¹ (see Appendix A for proof and other geometrical relations). However, the two circular cone axes do not intersect the ellipsoid at the same point. In an ideal approximation of the thin lens by an offset ellipsoidal reflector, one would like to have both beam axes of the incident and reflected beams intersect the ellipsoid at the center of the offset reflector. This condition can be approximately realized by locating the intersection of R_1 and R_2 midway between the intersections O and O' of two circular cone axes with the ellipsoid, as shown in Fig. 3, i.e.,

$$R_1(\theta_{p1} - \theta_o) = R_2(\theta'_o - \theta_{p2}), \quad (9)$$

where

$$\frac{\sin \theta_{p1}}{\sin 2\theta_i} = \frac{R_2}{\sqrt{R_1^2 + R_2^2 - 2R_1R_2\cos 2\theta_i}} \quad (10)$$

and

$$\frac{\sin \theta_{p2}}{\sin 2\theta_i} = \frac{R_1}{\sqrt{R_1^2 + R_2^2 - 2R_1R_2\cos 2\theta_i}}. \quad (11)$$

θ_o and θ'_o are defined in Appendix A. We also need the expressions for the eccentricity, e , and the distance, f_o , between the vertex and the near focus:

$$f_o = \frac{|R_1 + R_2| - \sqrt{R_1^2 + R_2^2 - 2R_1R_2\cos 2\theta_i}}{2} \quad (12)$$

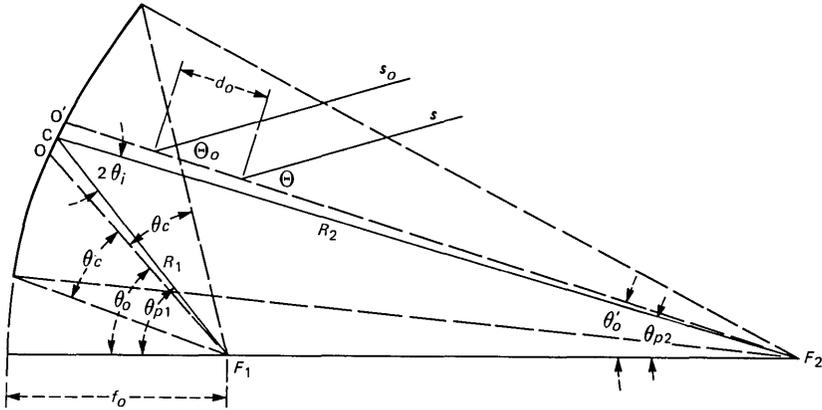


Fig. 3—Schematic diagram of an offset ellipsoidal reflector.

$$e = 1 - \frac{2f_o}{|R_1 + R_2|}. \quad (13)$$

Now the minor radius of the plane ellipse in eq. (23) should be equal to the radius of the projected circular aperture in the beam direction. (i.e., the thin lens aperture radius)

$$\rho_{\text{major}} \sqrt{1 - e^2 \sin^2 \theta_{po}} = a_l. \quad (14)$$

Substituting eqs. (20) and (22) into eq. (14), we obtain

$$e^2 \cos^2 \theta_o + 2e \cos \theta_c \cos \theta_o + 1 - \sin^2 \theta_c \left[e^2 + \frac{(1+e)^2 f^2}{a_l^2} \right] = 0. \quad (15)$$

When we are given an angle of incidence, θ_i , and a pair of radii of curvature, R_1 and R_2 , eqs. (9) and (15) can be solved simultaneously for numerical values of θ_o and θ_c . These parameters completely determine both the shape and size of the ellipsoid.

IV. NUMERICAL AND MEASURED RESULTS

To demonstrate practical feasibility of the proposed imaging feed design, we have built and tested an experimental broadband offset launcher. This feed was designed to provide the illumination required for the hyperboloidal subreflector of the Crawford Hill 7-meter offset cassegrainian antenna.

Following the method of Section II, we obtain parameters for the thin lens model of this experimental feed, as listed in Table I. Using the procedures described in Section III, we translated the lens parameters into dimensions of an offset ellipsoid, as shown in the schematic diagram of Fig. 3. A design frequency of 22 GHz was used in eqs. (7) and (8) to find R_1 and R_2 , as shown in Table II.

Table I—Design parameters of the thin lens model for a broadband imaging feed (all lengths are in cm.)

Required -15 dB illumination circle radius	47.5
Required phase-front radius of curvature	535.9
Frequency band (GHz)	19/28.5
Corrugated horn aperture radius, a_0	3.8
Horn apex to aperture distance, r_1'	15.9
Horn aperture to imaging lens distance, L_1'	47.5
Imaging lens to subreflector distance, L_1	684.6
Lens focal length, f	44.5
Lens radius (truncation edge taper ≈ 20 dB)	21
Lens to phase-center distance, Δ	148.7

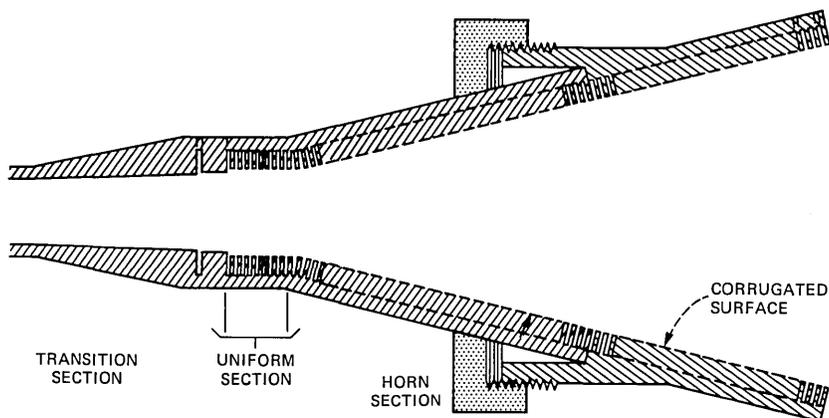


Fig. 4—Corrugated horn geometry.

A sketch of the corrugated horn is shown in Fig. 4. The corrugation depth is about a quarter wavelength at 18 GHz. The steps between smooth and corrugated sections help impedance matching because the desirable HE_{11} mode in the corrugated guide is concentrated toward the center of the cross section. The return loss of the horn is better than 20 dB for frequencies above 16 GHz. The E- and H-plane beamwidths are nearly identical to each other from 16 to 30 GHz.

Figure 5 shows a photograph of the combination of offset ellipsoidal reflector and corrugated horn. Amplitude and phase patterns have been measured in both the offset plane and the transverse plane orthogonal to the offset. Since the phase center is located about 1.5 m from the center of the offset ellipsoid, physical rotation around the phase center would cause problems of mechanical unbalance. Therefore, the center of rotation in the pattern measurements is located midway between the ellipsoid and the phase center. The measured data can be transformed into measured patterns around the phase center by the relations between the angles of rotation and between the path lengths.

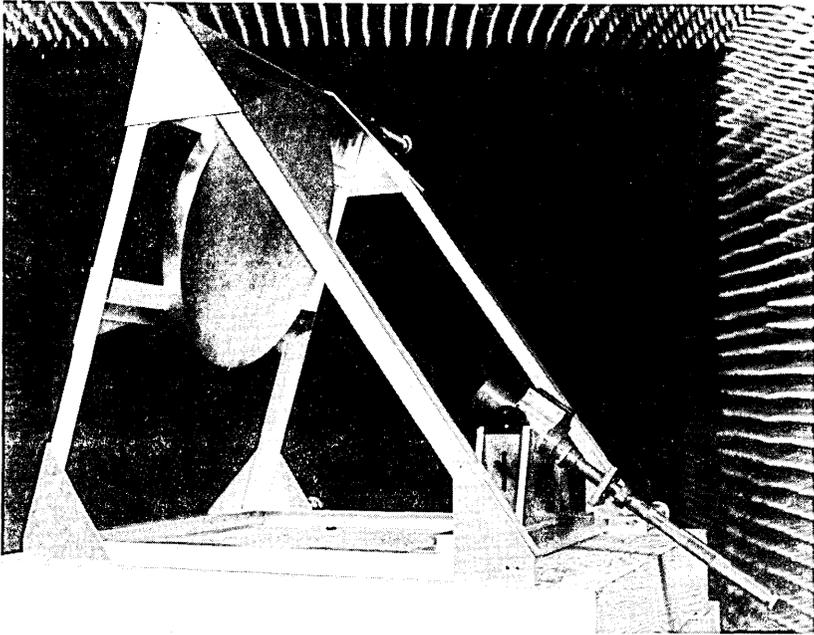


Fig. 5—Experimental broadband imaging feed.

Table II—Design parameters of the offset ellipsoidal reflector (all lengths are in cm.)

Incident phase-front radius of curvature R_1	54.36
Reflected phase-front radius of curvature R_2	244.22
Angle of incidence, θ_i	17°
Angle between incident center ray and major ellipsoidal axis, θ_{p1}	42.68°
Angle between reflected center ray and major ellipsoidal axis, θ_{p2}	8.68°
Offset angle, θ_o	41.87°
Half-cone angle, θ_c	22.09°
Offset angle at the distant focus, θ_o'	8.86°
Vertex to near focus distance, f_o	48.56
Eccentricity, e	0.67475
Semi-major-axis of plane ellipse, ρ_{major}	21.98
Semi-minor-axis of plane ellipse, ρ_{minor}	21
Semi-major-axis of ellipsoid, a	149.29
Semi-minor-axis of ellipsoid, b	110.18

The two equations below

$$\sin \Theta = \frac{s_o}{s} \sin \Theta_o \quad (16)$$

$$s = \sqrt{s_o^2 + d_o^2 - 2s_o d_o \cos \Theta_o} \quad (17)$$

are similar to eqs. (32) and (31). The parameters are explained in Fig. 3. The angular conversion of eq. (16) will be required in both

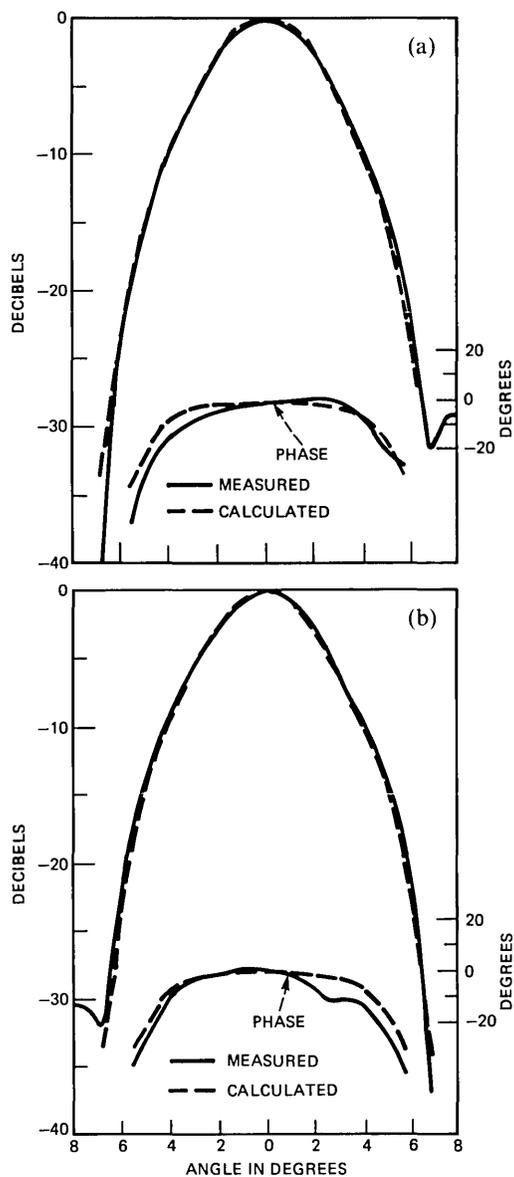


Fig. 6—19-GHz radiation pattern of transverse polarization (transverse perpendicular to offset) in (a) offset plane and (b) transverse plane.

amplitude- and phase-pattern transformations, whereas another factor of $(s + d_o - s_o)$ will be also added to the measured phase pattern.

To ensure reliable phase data, all pattern measurements were made in the azimuthal plane where a rotary joint can be used for avoiding

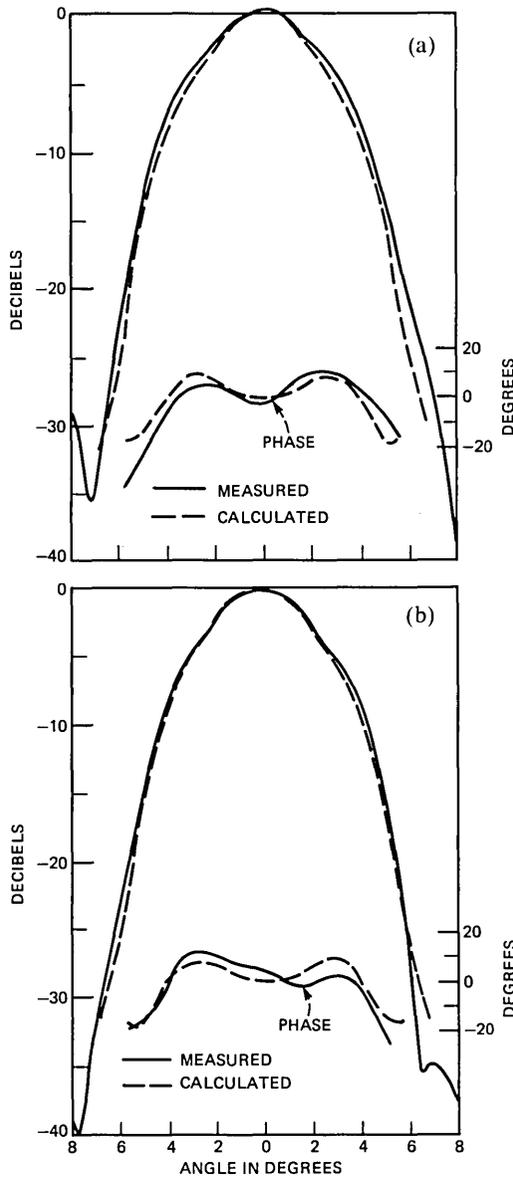


Fig. 7—23-GHz radiation pattern of transverse polarization (transverse perpendicular to offset) in (a) offset plane and (b) transverse plane.

cable motion during rotation of the turntable. Care was taken in keeping the same center of rotation by flipping over the feed assembly between measuring the transverse plane and offset plane cuts.

Calculations of radiation patterns are described in Appendix B. Both

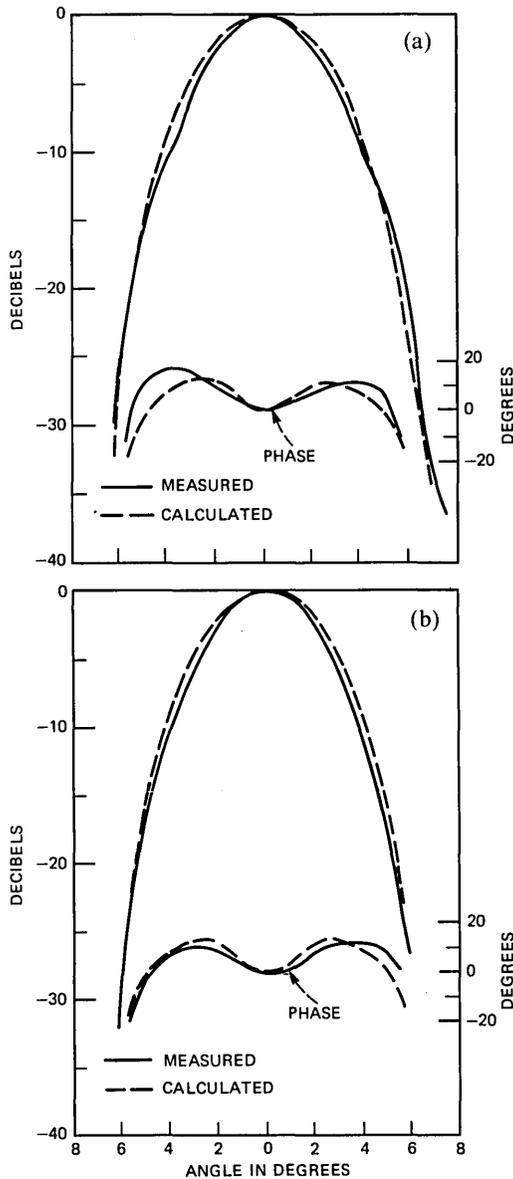


Fig. 8—28-GHz radiation pattern of transverse polarization (transverse perpendicular to offset) in (a) offset plane and (b) transverse plane.

measured and calculated patterns around the phase center are shown in Figs. 6, 7, and 8 for 19, 23, and 28 GHz, respectively. Comparisons between measured and calculated data show generally good agreement for both amplitude and phase patterns. An ideal imaging feed would

have a frequency-independent amplitude pattern of -15 dB taper at 5 degrees and a straight-line phase pattern that represents an exact spherical wave originating from a point source at the phase center. Deviations from these ideal patterns can be attributed to the imperfection of the paraxial approximation and to the truncation effect. Measured patterns also include the effect of deviation of the corrugated-horn-aperture field from the theoretical model. Figures 6 through 8 show that 28 -GHz measured pattern widths are slightly narrower than those of 19 GHz. Phase deviations remain less than 20 degrees. These results are very similar to those of a long broadband corrugated horn designed for constant beamwidth feed.

The patterns in Figs. 6 through 8 were measured with the polarization transverse to the offset plane. Measurements with the polarization parallel to the offset plane showed similar patterns. Cross-polarized pattern measurements showed a maximum cross polarization of -26 dB, which is what would be expected from the offset geometry.¹²

V. DISCUSSIONS

Theoretical and experimental studies have demonstrated the feasibility of a broadband imaging feed using the combination of an offset ellipsoid and a corrugated horn. This feed is also important for serving as a basic building block of the imaging beam waveguide.

Since a broadband feed design will avoid the need of quasi-optical frequency diplexing, a much simpler, cost-effective feed system can be built to achieve performance similar to that of the $19/28.5$ GHz feed of the Crawford-Hill 7-meter antenna.

The application of imaging of a corrugated horn by an offset ellipsoidal reflector is certainly not limited to broadband narrow feed patterns for large ground station antennas. For example, Dragone suggests¹³ its application to terrestrial microwave repeater antennas. Furthermore, it can be used as a constant beamwidth radiometer antenna for multifrequency remote sensing.

REFERENCES

1. T. S. Chu et al., "The Crawford Hill 7-Meter Millimeter Wave Antenna," *B.S.T.J.*, *57*, No. 5 (May-June 1978), pp. 1257-88.
2. M. J. Gans and R. A. Semplak, "Some Far-Field Studies of An Offset Launcher," *B.S.T.J.*, *54*, No. 7 (September 1975), pp. 1319-40.
3. T. S. Chu and W. E. Legg, "Gain of Corrugated Conical Horns," *IEEE Trans.*, *AP-30* (July 1982), pp. 698-703.
4. C. Dragone, "Reflection, Transmission, and Mode Conversion in a Corrugated Feed," *B.S.T.J.*, *56*, No. 6 (July-August 1977), pp. 835-67.
5. C. Dragone, "An Improved Antenna for Microwave Radio System Consisting of Two Cylindrical Reflectors and a Corrugated Horn," *B.S.T.J.*, *53*, No. 7 (September 1974), pp. 1351-77.
6. C. Dragone and M. J. Gans, "Imaging Reflector Arrangements to Form a Scanning Beam Using a Small Array," *B.S.T.J.*, *58*, No. 2 (February 1979), pp. 501-15.

7. H. Kogelnik, "Imaging of Optical Modes—Resonators With Internal Lenses," B.S.T.J., 44, No. 3 (March 1965), pp. 455-94.
8. T. S. Chu, "An Imaging Beam Waveguide Feed," IEEE Trans. AP-S, AP-31, No. 4 (July 1983).
9. T. S. Chu, "Broadband Astigmatic Compensation," Digest of 1981 IEEE Antenna and Propagation Symp., p. 131-4.
10. M. J. Gans, "Cross-Polarization in Reflector-Type Beam Waveguides and Antennas," B.S.T.J., 55, No. 3 (March 1976), pp. 289-316.
11. T. S. Chu, "Geometrical Representation of Gaussian Beam Propagation," B.S.T.J., 45, No. 2 (February 1966), pp. 287-99.
12. T. S. Chu and R. H. Turrin, "Depolarization of Offset Reflector Antennas," IEEE Trans., AP-21 (May 1973), pp. 339-45.
13. C. Dragone, private communication.
14. J. S. Cook, E. M. Elam, and H. Zucker, "The Open Cassegrain Antennas: Part I, Electromagnetic Design and Analysis," B.S.T.J., 44, No. 7 (September 1965), pp. 1255-1300.
15. V. Jamnejad-Dailami and Y. Rahmat-Samii, "Some Important General Features of Conic-Section-Generated Offset Reflector Antennas," IEEE Trans., AP-28 (November 1980), pp. 952-7.

APPENDIX A

Geometry of Offset Ellipsoids

The geometrical properties of offset ellipsoids can be generalized from those¹⁴ of offset paraboloids. The following formulas are derived by lengthy but straightforward algebra. The intersection of an offset ellipsoid and a circular cone subtended at the focus is a plane ellipse, which is subtended by another circular cone at the other (distant) focus. This property was indicated in Ref. 1, and also observed later in Ref. 15.

Let us define the ellipsoid in Figure 9 by

$$r = \frac{(1 + e)f_o}{1 + e \cos \theta_p}, \quad (18)$$

where the eccentricity, e , is less than unity. f_o is the distance between the origin in $X_p Y_p Z_p$ coordinates and the vertex, and θ_p is the polar angle with respect to the Z_p axis. We first find the intersection between the ellipsoid and the $x_o y_o$ plane, which is perpendicular to $x_p z_p$ plane, located at a distance r_o from the origin, and its normal makes an angle θ_{p_o} with the Z_p axis:

$$(1 - e^2 \sin^2 \theta_{p_o}) \left\{ x_o - \frac{e(\sin \theta_{p_o})}{1 - e^2 \sin^2 \theta_{p_o}} [(1 + e)f_o - er_o \cos \theta_{p_o}] \right\}^2 + y_o^2 = \frac{[er_o \cos \theta_{p_o} - (1 + e)f_o]^2}{1 - e^2 \sin^2 \theta_{p_o}} - r_o^2. \quad (19)$$

Next, the following expressions are found for θ_{p_o} and r_o :

$$\theta_{p_o} = \sin^{-1} \left(\frac{\sin \theta_o}{\sqrt{1 + 2e \cos \theta_o \cos \theta_c + e^2 \cos^2 \theta_c}} \right) \quad (20)$$

$$r_o = \frac{(1 + e)f_o \cos \theta_c}{\sqrt{1 + 2e \cos \theta_o \cos \theta_c + e^2 \cos^2 \theta_c}}, \quad (21)$$

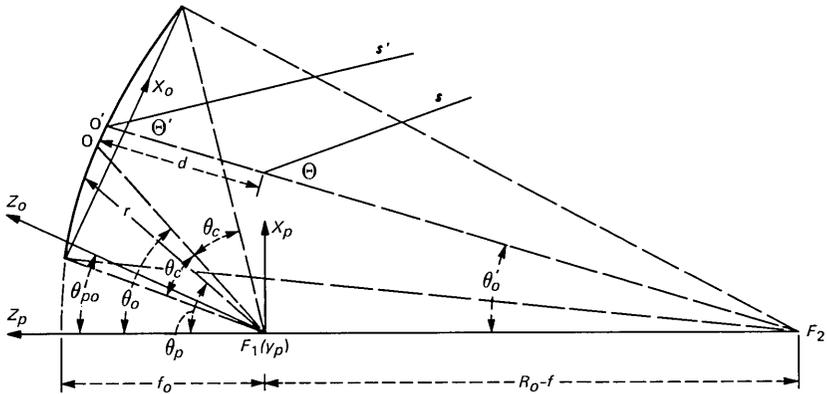


Fig. 9—Geometry of an offset ellipsoid subtended by circular cones at two foci.

where $\theta_o + \theta_c$ and $\theta_o - \theta_c$ are the polar angles of the top and bottom edges of the offset ellipsoid in the $x_p z_p$ plane. Now it can be shown that the intersection of the $x_o y_o$ plane with the circular cone, which has its axis oriented at $\theta_p = \theta_o$ and a half cone angle of θ_c , is the same ellipse as represented by eq. (19).

The major and minor radii of the ellipse are

$$\rho_{\text{major}} = \frac{(1 + e) f_o \sin \theta_c \sqrt{1 + 2e \cos \theta_o \cos \theta_c + e^2 \cos^2 \theta_c}}{1 + 2e \cos \theta_o \cos \theta_c + e^2 \cos^2 \theta_c - e^2 \sin^2 \theta_o} \quad (22)$$

$$\rho_{\text{minor}} = \rho_{\text{major}} \sqrt{1 - e^2 \sin^2 \theta_{po}}. \quad (23)$$

When $e = 1$, the above equations are reduced to those of Appendix A in Ref. 14. One notes an error in eq. (50) for the plane ellipse in Ref. 14, i.e., missing a term ($r_o \tan \theta_{po}$) in the bracket on the left-hand side of that equation.

The axis of the circular cone subtended at the second (distant) focus is oriented with respect to the major axis of the ellipsoid at

$$\theta'_o = \frac{1}{2} \left\{ \sin^{-1} \left[\frac{r_2}{R_o + f_o - r_2} \sin(\theta_o + \theta_c) \right] + \sin^{-1} \left[\frac{r_1}{R_o + f_o - r_1} \sin(\theta_o - \theta_c) \right] \right\}, \quad (24)$$

where

$$r_{1,2} = \frac{(1 + e) f_o}{1 + e \cos(\theta_o \mp \theta_c)} \quad (25)$$

are the radial distances from the (near) focus to the termini of the major axis of the ellipse represented by eq. (19) and

$$R_o = f_o \left(\frac{1+e}{1-e} \right) \quad (26)$$

is the distance from the vertex to the second focus.

APPENDIX B

Calculation of Radiation Pattern

In this appendix we shall calculate the radiation pattern of an offset ellipsoid illuminated by a corrugated feed horn. The numerical integration of the diffraction integral accomplishes computer simulation of the hardware.

When an ellipsoid is illuminated by a feed located at the near focus, the geometrical optic rays reflected from the ellipsoid are expected to converge toward the distant focus. Although the diffraction effect will take place before reaching that second focus, the ray approximation should be valid in the vicinity of the ellipsoid. To find the radiation from the ellipsoid, we will employ a spherical wave field in an equivalent plane aperture that is perpendicular to the axis of, and subtended by, the cone of reflected rays, and passes through the point, O' , of intersection of the centric reflected ray with the ellipsoid. The radius of curvature of the spherical wave is the distance, R'_o , from the distant focus to the point O' .

$$R'_o = \frac{(1+e)f_o}{1-e \cos \theta'_o} \quad (27)$$

The transverse cartesian coordinates in the plane aperture are

$$\begin{cases} x = R(\cos \theta_o \sin \theta \cos \phi + \sin \theta_o \cos \theta) \cos \theta'_o \\ \quad - [R(-\sin \theta_o \sin \theta \cos \phi + \cos \theta_o \cos \theta) + R_o - f] \sin \theta'_o \\ y = R \sin \theta \sin \phi, \end{cases} \quad (28)$$

where

$$R = \frac{(1+e)f_o}{1+e(\cos \theta_o \cos \theta - \sin \theta \sin \theta_o \cos \phi)} \quad (29)$$

is the position radius from the focus to the ellipsoid. θ and ϕ are the standard spherical coordinates with the polar axis F_1O along the axis of the circular cone subtended at the near focus, as shown in Fig. 9.

Using a small angle scalar approximation, we can determine the radiation of the offset ellipsoid as

$$E = \frac{je^{-jk(s+d)}}{\lambda(s+d)} \int \frac{F}{R} \exp \left\{ jk \left[x \sin \Theta' \cos \Phi + y \sin \Theta' \sin \Phi + \frac{x^2 + y^2}{2R'_o} - \frac{x^2 + y^2}{2(s+d)} + (s+d-s') \right] \right\} dA, \quad (30)$$

where F is the feed pattern, which will be described later. The term $(x^2 + y^2)/2(s + d)$ is a phase correction needed for the Fresnel region in which the Cassegrainian subreflector is located. The factor $(s + d - s')$ is added to give a phase pattern with respect to the phase center, which is located at a distance d in front of the ellipsoid. s is the distance between the phase center and the field point, and s' is the distance between the point O' , which is the effective center of the ellipsoid aperture, and the field point

$$s' = \sqrt{s^2 + d^2 + 2sd \cos \Theta}. \quad (31)$$

The angle Θ' , between s' and the cone axis $O'F_2$, is related by the sine law to the angle Θ between s and the cone axis, as shown below:

$$\sin \Theta' = \frac{s}{s'} \sin \Theta. \quad (32)$$

ϕ is the azimuth coordinate of the field point. The surface element dA is

$$dA = \frac{R^2 \sin \theta d\theta d\phi}{\cos \theta_i} \cos \theta_r, \quad (33)$$

where θ_i is the angle of incidence between the incident ray and the unit vector normal to the ellipsoidal surface, and θ_r is the angle between this unit normal and the beam axis. Since the ellipsoids under consideration are only slight perturbations of paraboloids, $\cos \theta_i \approx \cos \theta_r$. Noting the symmetry of the aperture field, eq. (30) becomes

$$E = \frac{je^{-jk(s+d)}}{\lambda(s+d)} \int_0^\pi \int_0^{\theta_c} F \exp jk \left[x \sin \Theta' \cos \Phi + \frac{x^2 + y^2}{2R'_o} - \frac{x^2 + y^2}{2(s+d)} + (s+d-s') \right] \cos[ky \sin \Theta' \sin \Phi] R \sin \theta d\theta d\phi. \quad (34)$$

The feed pattern of a corrugated horn is given by

$$F(\Theta') = \int_0^1 J_o(\alpha r) J_o \left(ka \sin \theta' \frac{R}{R'} r \right) e^{-jk \left[ar^2 \tan \frac{\beta}{2} + \frac{(ar)^2}{2R'} + \ell(1 - \cos \theta') \right]} r dr, \quad (35)$$

where a is the aperture radius, β is the half cone angle of the horn, $\alpha = 2.405$ for frequencies close to the resonance of the corrugated depth, r is the normalized radial coordinate of the horn aperture, and $\sin \theta'$ is multiplied by R/R' because θ' is referred to the ellipsoid focus (the phase center of the horn), which is located at a distance ℓ behind the horn aperture. R' is the distance from the center of the horn aperture to a point on the ellipsoid:

$$R' = \sqrt{R^2 + \ell^2 - 2R\ell \cos \theta'}. \quad (36)$$

The first term $kar^2 \tan(\beta/2)$ in the exponential bracket of eq. (35) is the phase deviation in the horn aperture, the second term is a phase correction in the Fresnel region, and the third term is added to give a phase pattern with respect to the horn phase center (ellipsoidal focus) instead of the horn aperture.

To compensate for the difference in length between the radii from the focus toward the top and bottom edges of the offset ellipsoid, the axis of the feed horn is offset by an angle θ_1 from the axis of the circular cone subtended by the ellipsoid at the focus. θ' can be expressed in terms of the angular coordinates (θ, ϕ) of the offset ellipsoid:

$$\cos \theta' = \cos \theta \cos \theta_1 + \sin \theta \sin \theta_1 \cos \phi. \quad (37)$$

$\theta_1 = 1.6$ degrees is used in the calculated patterns of Figs. 6, 7, and 8.

An M/G/1 Queue With a Hybrid Discipline

By B. T. DOSHI

(Manuscript received October 27, 1982)

In this paper we analyze the delay in a single-server queue in which the server, when it becomes free, selects for the next service the oldest customer with current delay smaller than T . If no such customer is present, then it selects the youngest customer with the current delay in excess of T . This service discipline is desirable in applications where the success or failure of a service depends on the delay in providing the service. Telephone call processing and steel rolling are two of these applications. We obtain the delay distribution for this service discipline using a combination of level-crossing arguments and renewal theory, and compare this performance with that of the last-in-first-out discipline with respect to the throughput of successfully served customers.

I. INTRODUCTION

The following situation is common in telephone call processing or data-processing systems. When a customer requests service, an entry is made in the queue that is serviced by a processor. The processor serves the queue of entries according to some specified service discipline. When an entry is served, the corresponding customer is notified of the completion of the service. The customer, however, does not wait forever for the completion of the service. At some random time, R , after its arrival, the customer will renege if service is not completed. The associated entry remains in the queue, and the server does not know that the customer has reneged until after it completes the service and attempts to notify the customer. Such a service is wasted. The customer may make the situation worse by reattempts at getting the desired service, thereby increasing the load on the server. It is therefore necessary to keep the proportion of renegeing customers as small as possible. This can be done by selecting an appropriate discipline.

In this paper we study a specific model of the above situation. In

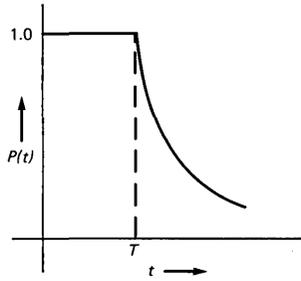


Fig. 1—Customer behavior function 1.

particular, we have a single-server queue with Poisson arrivals at rate $\lambda > 0$.

Let G be the distribution of the service time. Let $\bar{P}(t)$ be the probability that the customer does not renege before time t . We may think of $\bar{P}(t)$ as the expected reward obtained by completing the service of a customer at t time units after its arrival. With this general interpretation $\bar{P}(t)$ need not be restricted to be between 0 and 1. Let

$$P(t) = \int_{0^-}^{\infty} \bar{P}(t + y) dG(y) \quad (1)$$

for $0 \leq t < \infty$. Then $P(t)$ is the expected reward from a customer whose *waiting time* (excluding the service time) is t . Let W_{π} be the distribution of the waiting time under the specified service discipline, π . Then the expected reward from an arbitrary customer is

$$V_{\pi} = \int_{0^-}^{\infty} P(t) dW_{\pi}(t). \quad (2)$$

We want to select a service discipline that maximizes V_{π} . It was shown in Doshi and Lipper¹ that if $P(t)$ is convex (respectively, concave), then the last-in-first-out (LIFO) [respectively, first-in-first-out (FIFO)] discipline is optimal. More realistic functions $P(t)$, however, are of the forms given in Figs. 1 and 2.

For such functions, $P(t)$, an optimal service discipline, is not known. However, our results for concave and convex $P(t)$ indicate that a hybrid discipline may provide better performance than either the FIFO or the LIFO discipline does. In this hybrid discipline the server, when it completes a service, first looks at the customers with the current waiting time less than T and selects the oldest waiting customer for the next service. If no such customer is waiting, then the server looks at the customers with the current waiting time in excess of T and selects the youngest customer. Note that this hybrid discipline includes FIFO ($T = \infty$) and LIFO ($T = 0$) as special cases. Since $P(t)$

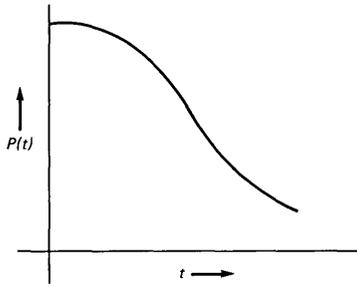


Fig. 2—Customer behavior function 2.

is assumed to be given, we only need $W_\pi(t)$, $0 \leq t < \infty$, for this hybrid discipline π , to calculate V_π . In this paper we obtain W_π for general service time distribution. We do this in three steps. First we describe another queueing system for which the distribution of the waiting time is the same as for the original system. Moreover, for this equivalent queueing system the distribution of the waiting time can easily be expressed in terms of the distribution, $F(x)$, of the work in a subsystem. We then use level-crossing arguments to derive an integral equation satisfied by $f(x) = F'(x)$. Finally, we use some results from renewal theory to solve this integral equation.

Some comments about the model are in order before we proceed to give an outline of the rest of this paper. Models similar to the one studied here can be useful in a variety of other applications. Some of these are the management of steel-rolling operation and the management of blood bank. Also, in many applications the customers do not necessarily renege. They simply take actions (start to dial, become cold, etc.) which make any subsequent service worthless.

This paper is organized as follows: In Section II we formally define the queueing system under consideration. We describe an equivalent queueing system in Section III. There we also show the relationship between the distributions of the waiting time in the original system and of the work in a subsystem of the equivalent queueing system. In Section IV we derive an integral equation for the steady-state density of the work in the subsystem. We give the solution of this integral equation in Section V. There we also derive the steady-state distribution of the waiting time in the original system. Finally, we give some numerical results in Section VI.

II. MODEL

The queueing system and the hybrid service discipline discussed in Section I can be formally described as follows: We have a queueing system with a single server and two queues, Q1 and Q2. Customers in

Q1 have a nonpreemptive priority over those in Q2. The order of service is first-in-first-out (FIFO) in Q1 and last-in-first-out (LIFO) in Q2. Customers arrive according to a Poisson process at rate $\lambda > 0$. On arrival the customer is put in Q1. If its service has not started within T seconds after its arrival, then at that time it is transferred to Q2. The service times of the customers are independent and identically distributed with distribution function G with continuous density g .

Let μ_K denote the K th moment of the service time and let $\rho = \lambda\mu_1$. Assume that the waiting-time process is in the steady state. Let W denote the distribution function of the waiting time seen by an arbitrary customer. Since G has a continuous density, W is differentiable on $(0, \infty)$. Let

$$w(x) = W'(x) \quad 0 < x < \infty.$$

We are interested in obtaining an expression for $W(x)$ or, equivalently, for $W(0)$ and $w(x)$, $0 < x < \infty$.

III. AN EQUIVALENT QUEUEING SYSTEM

We now describe a queueing system that is equivalent to the one described in Section II as far as the waiting times of the customers are concerned. However, the number of customers in Q1 and Q2 at a given time may be different in the two systems.

Consider the subsystem consisting of the server and Q1. Let X_t denote the work, at time t , in this subsystem. Thus, X_t is the sum of the remaining service time of the customer, if any, being served and the service times of all the customers in Q1. If a customer arriving at time t finds $X_t \leq T$, then it joins Q1; otherwise it joins Q2. Recall that Q1 has a nonpreemptive priority over Q2 and that Q1 is served FIFO and Q2 is served LIFO. A little reflection shows that the waiting time of a customer is the same in this system as in the one described in Section II.

We now relate the waiting-time distribution, W , to the steady-state distribution, F , of the work, X , in the subsystem consisting of the server and Q1. If an arriving customer sees $X \leq T$, then its waiting time will be X because the service in Q1 is FIFO and because Q1 has a nonpreemptive priority over Q2. Thus,

$$W(x) = F(x) \quad 0 \leq x \leq T. \quad (3)$$

In particular,

$$W(0) = F(0), \quad (4)$$

and

$$w(x) = f(x) = F'(x) \quad 0 < x < T. \quad (5)$$

If an arriving customer sees $X > T$, then it joins Q2. Since Q1 has priority over Q2 and since Q2 is served LIFO, this customer has to wait for the duration of a busy period in the M/G/1 queue with initial work, X . Let B_y denote the distribution of the busy period in the M/G/1 queue started by initial work, y . Since G has a continuous density,

$$b_y(x) = B'_y(x) \quad (6)$$

exists for all $x > y$. Then

$$w(x) = \int_T^x f(y)b_y(x)d(y) \quad x > T. \quad (7)$$

Thus, it is sufficient to find the distribution of X .

IV. INTEGRAL EQUATION FOR $f(x)$

We now use level-crossing arguments (see Ref. 2) to derive an integral equation for f . Figure 3 shows a typical sample function for the process $\{X_t\}$. Assume that at time 0 the queues are empty, the server is idle, and a customer arrives and enters service. If $X_t > 0$, then it decreases at unit rate as in a M/G/1 queue. Symbols \circ represent arrivals which see $X_t \leq T$ and join the subsystem, thus increasing X_t by a service time. A symbol in the shape of a dot (\bullet) represents arrivals that see $X_t > T$, and join Q2 without affecting X_t on their arrival. When X_t reaches zero, two things can happen: Q2 is empty and the server remains idle until the next arrival, or Q2 is nonempty and a customer from Q2 enters service, thus increasing X_t by its service time. Such arrivals from Q2 into the server are denoted by a symbol in the shape of a square (\blacksquare).

The stochastic process $\{X_t\}$ is not Markovian because what happens

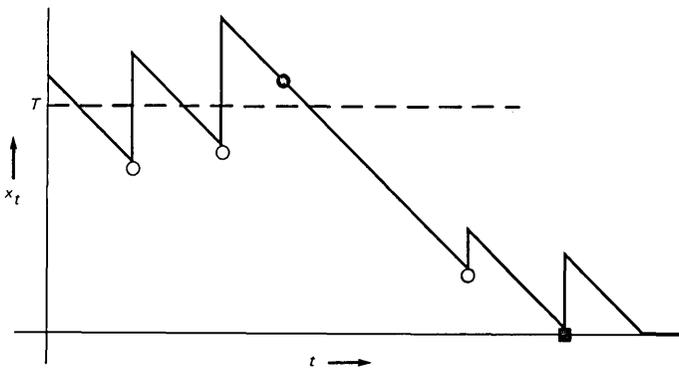


Fig. 3—Typical sample function for the process $\{X_t\}$.

when X_t reaches zero depends on the past. On the other hand, the vector-valued process $\{(X_t, N_t)\}$, where N_t is the number in Q2 at time t , is a Markov process. We can derive the steady-state distribution of (X_t, N_t) using the standard results and use that to obtain the marginal steady-state distribution of X . We, however, use a simpler approach here. Recall that $\mu_2 < \infty$. Consider the following two cases:

(i) $\rho < 1$. In this case the process $\{X_t\}$ is regenerative and with the regeneration points corresponding to the external arrivals that make an idle server busy. Denote this event by E . Then E is a positive recurrent, regenerative event.

(ii) $\rho \geq 1$. In this case the queue length in Q2 grows without bound and, in the steady state, can be assumed to be ∞ . Thus, a customer is removed from Q2 to enter service every time X_t reaches 0. This event, E' , is then positive recurrent.

Standard regenerative arguments now show that $\{X_t\}$ has a steady-state distribution and that

$$X_t \xrightarrow{D} X,$$

where the distribution of X is the steady-state distribution of $\{X_t\}$. Moreover, for any x , $0 < x < \infty$, the steady-state rate, $D(x)$, at which X_t crosses x from above, equals the rate $U(x)$, at which X_t jumps from below x to above x . We now express $D(x)$ and $U(x)$ in terms of $f(x)$ and get the desired integral equation by equating these expressions.

X_t decreases at unit rate until an arrival occurs or until $X_t = 0$. Thus, during every downcrossing of level x , the $\{X_t\}$ process spends dx units of time in the interval $(x - dx, x)$. Hence,

$$D(x) = f(x) \quad 0 < x < \infty. \quad (8)$$

Before deriving an expression for $U(x)$ we introduce some notation. Let p denote the rate at which X_t jumps from 0 to some positive value. These jumps may be due to either external arrivals coming to an idle system or to customers from Q2 moving to the server. Also, let \bar{G} denote the complementary service time distribution defined by

$$\bar{G}(x) = 1 - G(x) \quad 0 \leq x < \infty. \quad (9)$$

Assume that $G(0) = 0$, $\bar{G}(0) = 1$. Then, for $x \leq T$

$$U(x) = \lambda \int_0^x f(y) \bar{G}(x - y) dy + p \bar{G}(x). \quad (10)$$

Since an external arrival causes a jump in X_t only when $X_t \leq T$, we have, for $x > T$,

$$U(x) = \lambda \int_0^T f(y) \bar{G}(x - y) dy + p \bar{G}(x). \quad (11)$$

Thus,

$$U(x) = \lambda \int_0^{T \wedge x} f(y) \bar{G}(x-y) dy + p \bar{G}(x) \quad 0 < x < \infty. \quad (12)$$

Let

$$f(0) = \lim_{x \downarrow 0} f(x). \quad (13)$$

Then $f(0)$ is the rate at which $\{X_t\}$ hits level zero. In the steady state, this must equal the rate at which $\{X_t\}$ jumps from 0 to some positive value. Thus

$$p = f(0). \quad (14)$$

We now have for $0 < \rho < \infty$, $0 < x < \infty$,

$$\begin{aligned} f(x) &= D(x) \\ &= U(x) \\ &= \lambda \int_0^{x \wedge T} f(y) \bar{G}(x-y) dy + f(0) \bar{G}(x). \end{aligned} \quad (15)$$

Let

$$q = \frac{f(0)}{\lambda}.$$

Then

$$f(x) = \lambda \int_0^{x \wedge T} f(y) \bar{G}(x-y) dy + q \lambda \bar{G}(x). \quad (16)$$

This is the desired integral equation for f . The additional conditions needed to solve this completely depend on whether $\rho < 1$ or $\rho \geq 1$.

First consider the case where $\rho < 1$. Let P_2 be the probability that an arriving customer sees $X_t > T$ and joins Q2. Since, for $\rho < 1$, every arriving customer is eventually served, the rate at which customers enter the server from Q2 is λP_2 . Also, the rate of arrivals coming to an empty system is $\lambda F(0)$. Thus,

$$q = \frac{f(0)}{\lambda} = \frac{\lambda[F(0) + P_2]}{\lambda} = F(0) + P_2. \quad (17)$$

Also,

$$P_2 = \int_T^\infty f(x) dx, \quad (18)$$

and

$$F(0) + \int_0^{\infty} f(x)dx = 1. \quad (19)$$

For $\rho \geq 1$, Q_2 is always nonempty in steady state. Hence $F(0) = 0$,

$$q = \frac{f(0)}{\lambda}, \quad (20)$$

and

$$\int_0^{\infty} f(x)dx = 1. \quad (21)$$

Equation (16), together with either conditions [eqs. (17) to (19) or 20], characterizes f completely. We solve this equation in the next section.

V. SOLUTION OF THE INTEGRAL EQUATION

We now solve eq. (15) to obtain an expression for $f(x)$. For $0 < x < \infty$, let

$$h(x) = \lambda \bar{G}(x) \quad (22)$$

and let $m(x)$ be the renewal density function for $h(x)$. Then $m(x)$ satisfies (see Ref. 3):

$$m(x) = h(x) + \int_0^x h(x-y)m(y)dy. \quad (23)$$

Equation (15) can now be rewritten as

$$f(x) = qh(x) + \int_0^x f(x-y)h(y)dy \quad (0 < x < T), \quad (24)$$

and

$$f(x) = qh(x) + \int_0^T f(y)h(x-y)dy \quad T \leq x < \infty. \quad (25)$$

Equation (24) is a renewal equation and its solution is given by

$$\begin{aligned} f(x) &= qh(x) + q \int_0^x h(x-y)m(y)dy \\ &= q \left[h(x) + \int_0^x h(x-y)m(y)dy \right] \\ &= qm(x), \end{aligned} \quad (26)$$

where the last equality follows from (23). Note that $0 < h(x) < \infty$, and

$$\int_0^{\infty} h(x)dx = \lambda\mu_1 = \rho.$$

Thus, h is like a probability density with total mass, ρ . The function $m(x)$ is well defined for any finite x irrespective of the value of ρ , $0 < \rho < \infty$. To obtain an expression for $f(x)$, $x \geq T$, we note that the right-hand side of (25) involves $f(y)$ for y only in the interval $(0, T)$, which we have obtained in (26). Thus, replacing $f(y)$ on the right-hand side of (25) by $qm(y)$, we get

$$\begin{aligned} f(x) &= \left[qh(x) + \int_0^T qm(y)h(x-y)dy \right] \\ &= q \left[h(x) + \int_0^T h(x-y)m(y)dy \right] \quad x \geq T. \end{aligned} \quad (27)$$

We now use conditions (17) through (19) or (20) through (21) to evaluate q and thus completely characterize f . First, consider the case $\rho < 1$. We have

$$q = F(0) + P_2, \quad (28)$$

and

$$F(0) + q \int_0^T m(x)dx + P_2 = 1. \quad (29)$$

Also, equating the rate of customers coming to the system with the rate of customers leaving the system, we get

$$\lambda = \frac{1}{\mu_1} [1 - F(0)],$$

or

$$F(0) = 1 - \rho. \quad (30)$$

We can now solve (28) through (30) for q and P_2 to get

$$P_2 = \frac{\rho - (1 - \rho) \int_0^T m(x)dx}{1 + \int_0^T m(x)dx},$$

and

$$q = F(0) + P_2 = \frac{1}{1 + \int_0^T m(x)dx}.$$

Thus,

$$F(0) = 1 - \rho \tag{31}$$

$$f(x) = \frac{m(x)}{1 + \int_0^T m(x)dx} \quad x < T \tag{32}$$

$$= \frac{h(x) + \int_{y=0}^T h(x-y)m(y)dy}{1 + \int_0^T m(x)dx} \quad x \geq T. \tag{33}$$

Next, consider the case where $\rho \geq 1$. Here,

$$q = \frac{f(0)}{\lambda},$$

and

$$\int_0^T f(x)dx + \int_T^\infty f(x)dx = 1.$$

Thus,

$$f(0) \left\{ \int_0^T m(x)dx + \rho \left[1 + \int_0^T m(x)dx \right] - \int_0^T m(x)dx \right\} = 1,$$

or

$$f(0) = \frac{1}{\rho \left[1 + \int_0^T m(x)dx \right]}, \tag{34}$$

$$f(x) = \frac{m(x)}{\rho \left[1 + \int_0^T m(x)dx \right]} \quad x < T, \tag{35}$$

and

$$f(x) = \frac{h(x) + \int_0^T h(x-y)m(y)dy}{\rho \left[1 + \int_0^T m(x)dx \right]} \quad x \geq T. \quad (36)$$

We now consider a special case where $\bar{G}(x) = e^{-\mu x}$, $\mu = 1/\mu_1$. Then,

$$\begin{aligned} h(x) &= \lambda e^{-\mu x} & 0 < x < \infty, \\ m(x) &= \lambda e^{(\lambda-\mu)x} & 0 < x < \infty, \\ 1 + \int_0^T m(x)dx &= \frac{1}{(1-\rho)} [1 - \rho e^{-\mu(1-\rho)T}], \end{aligned}$$

and, for $x \geq T$,

$$\begin{aligned} h(x) + \int_0^T h(x-y)m(y)dy &= \lambda e^{-\mu x} + \lambda^2 \int_0^T e^{-\mu(x-y)} e^{-(\mu-\lambda)y} dy \\ &= \lambda e^{-\mu x} + \lambda^2 e^{-\mu x} \int_0^T e^{+\lambda y} dy \\ &= \lambda e^{-\mu x + \lambda T}. \end{aligned}$$

Thus, for $\rho < 1$

$$\begin{aligned} F(0) &= 1 - \rho \\ f(x) &= \frac{1 - \rho}{1 - \rho e^{-\mu(1-\rho)T}} e^{-\mu(1-\rho)x} \quad x < T, \end{aligned}$$

and

$$f(x) = \frac{1 - \rho}{1 - \rho e^{-\mu(1-\rho)T}} e^{-\mu x + \lambda T} \quad x \geq T.$$

For $\rho \geq 1$

$$\begin{aligned} F(0) &= 0, \\ f(x) &= \frac{1 - \rho}{\rho [1 - \rho e^{\mu(\rho-1)T}]} e^{\mu(\rho-1)x} \quad x < T, \end{aligned}$$

and

$$f(x) = \frac{1 - \rho}{\rho[1 - \rho e^{\mu(\rho-1)T}]} e^{-\mu x + \lambda T} \quad x \geq T.$$

VI. WAITING-TIME DISTRIBUTION

We now use the results of Section V to obtain an expression for the waiting-time distribution and its Laplace Stieltjes Transform. From eqs. (4), (5) and (6) we get

$$W(0) = F(0) = \begin{cases} 1 - \rho & \rho < 1 \\ 0 & \rho \geq 1, \end{cases} \quad (37)$$

$$w(x) = f(x) = \begin{cases} \frac{m(x)}{\left[1 + \int_0^T m(x) dx\right]} & \rho < 1 \\ \frac{m(x)}{\rho \left[1 + \int_0^T m(x) dx\right]} & \rho \geq 1, \end{cases} \quad (38)$$

and

$$w(x) = \begin{cases} \frac{\int_T^x \left[h(y) + \int_0^T h(y-z)m(z) dz \right] b_y(x) dy}{1 + \int_0^T m(x) dx} & \rho < 1 \\ \frac{\int_T^x \left[h(y) + \int_0^y h(y-z)m(z) dz \right] b_y(x) dy}{\rho \left[1 + \int_0^T m(x) dx\right]} & \rho \geq 1. \end{cases} \quad (39)$$

Let

$$W^*(\theta) = E[e^{-\theta W}] = W(0) + \int_0^\infty e^{-\theta x} w(x) dx \quad \text{Re } \theta > 0. \quad (40)$$

Also, for $\text{Re } \theta > 0$ let

$$g^*(\theta) = \int_0^\infty e^{-\theta x} g(x) dx \quad (41)$$

$$\begin{aligned}
 h^*(\theta) &= \int_0^\infty e^{-\theta x} h(x) dx \\
 &= \frac{\lambda[1 - g^*(\theta)]}{\theta},
 \end{aligned} \tag{42}$$

and

$$\begin{aligned}
 m^*(\theta) &= \int_0^\infty e^{-\theta x} m(x) dx \\
 &= \frac{h^*(\theta)}{1 - h^*(\theta)} \\
 &= \frac{\lambda[1 - g^*(\theta)]}{\theta - \lambda[1 - g^*(\theta)]}.
 \end{aligned} \tag{43}$$

Let $B^*(\theta)$ be the Laplace Stieltjes Transform of the ordinary busy period in an M/G/1 queue with the arrival rate $\lambda > 0$ and the service distribution, G . Then B^* satisfies

$$B^*(\theta) = g^*\{\theta + \lambda[1 - B^*(\theta)]\}. \tag{44}$$

Also, let $B_y^*(\theta)$ be the Laplace Stieltjes Transform of B_y . Then

$$B_y^*(\theta) = e^{-y(\theta + \lambda[1 - B^*(\theta)])}. \tag{45}$$

We now express $W^*(\theta)$ in terms of m , m^* and B^* . First, consider the case $\rho < 1$, where

$$\begin{aligned}
 W^*(\theta) &= 1 - \rho + \frac{1}{1 + \int_0^T m(x) dx} \left\{ \int_0^T e^{-\theta x} m(x) dx \right. \\
 &\quad \left. + \int_T^\infty e^{-\theta x} \int_{y=T}^x \left[h(y) + \int_0^T h(y-z)m(z) dz \right] b_y(x) dy dx \right\} \\
 &= 1 - \rho + \frac{1}{1 + \int_0^T m(x) dx} \left(\int_0^T e^{-\theta x} m(x) dx \right. \\
 &\quad \left. + \frac{\theta}{\theta + \lambda[1 - B^*(\theta)]} \int_T^\infty e^{-(\theta + \lambda[1 - B^*(\theta)])x} m(x) dx \right).
 \end{aligned} \tag{46}$$

For $\rho > 1$,

$$W^*(\theta) = \frac{1}{\rho \left[1 + \int_0^T m(x) dx \right]} \left\{ \int_0^T e^{-\theta x} m(x) dx + \frac{\theta}{\theta + \lambda[1 - B^*(\theta)]} \int_T^\infty e^{-[\theta + \lambda(1 - B^*(\theta))]x} m(x) dx \right\}. \quad (47)$$

6.1 Mean value of the waiting time

Let \bar{W} denote the mean value of the waiting time. For $\rho < 1$, every customer is eventually served. Hence, \bar{W} is the average over all the customers. For $\rho > 1$, some of the arriving customers do not get served and, in this case, \bar{W} is the average waiting time of those who do get served. In the first case,

$$\bar{W} = -W^{*'}(0^+),$$

and for the second case,

$$\bar{W} = \frac{-W^{*'}(0^+)}{W^*(0^+)}.$$

For $\rho < 1$, all customers are served and the order of service does not affect the mean waiting time. Thus, in this case the mean waiting time is the same as that for an M/G/1 queue with the FIFO discipline. That is,

$$\bar{W} = \frac{\lambda\mu_2}{2(1 - \rho)}. \quad (48)$$

For $\rho > 1$, the busy period distribution is defective. Let

$$b_o = P\{\text{Busy period} < \infty\}.$$

Then b_o is the unique solution in $(0, 1)$ of

$$B^*(0^+) = b_o = g^*[\lambda(1 - b_o)]. \quad (49)$$

Also,

$$B^{*'}(0^+) = \frac{g^{*'}[\lambda(1 - b_o)]}{1 + \lambda g^{*'}[\lambda(1 - b_o)]} \quad (50)$$

and

$$B^{*''}(0^+) = \frac{g^{*''}[\lambda(1 - b_o)]}{\{1 + \lambda g^{*'}[\lambda(1 - b_o)]\}^3}. \quad (51)$$

Now, from eq. (47), we get

$$\begin{aligned}
 W^*(0^+) &= \frac{1}{\rho \left[1 + \int_0^T m(x) dx \right]} \\
 &\quad \cdot \left[\int_0^T m(x) dx + \frac{1}{\lambda(1-b_0)} \lim_{\theta \rightarrow 0^+} \int_T^\infty \theta e^{-(\theta + \lambda[1-B^*(\theta)]x)} m(x) dx \right] \\
 &= \frac{1}{\rho \left[1 + \int_0^T m(x) dx \right]} \left[\int_0^T m(x) dx + 1 \right] \\
 &= \frac{1}{\rho} \tag{52}
 \end{aligned}$$

and

$$\begin{aligned}
 -W^{*'}(0^+) &= \frac{1}{\rho \left[1 + \int_0^T m(x) dx \right]} \\
 &\quad \cdot \left\{ \int_0^T xm(x) dx + \frac{1}{\lambda(1-b_0)} \left[1 + \int_0^T m(x) dx \right] \right\} \\
 &= \frac{1}{\rho} \left[\frac{\int_0^T xm(x) dx}{1 + \int_0^T m(x) dx} + \frac{1}{\lambda(1-b_0)} \right]. \tag{53}
 \end{aligned}$$

From (52) and (53) we have

$$\bar{W} = \frac{-W^{*'}(0^+)}{W^*(0^+)} = \frac{\int_0^T xm(x) dx}{1 + \int_0^T m(x) dx} + \frac{1}{\lambda(1-b_0)}. \tag{54}$$

Equation (54) shows that the mean waiting time of the customers who get served is minimized by the LIFO discipline ($T = 0$).

VII. NUMERICAL RESULTS

In this section we present some numerical results. Instead of calcu-

lating the waiting-time distribution, we calculate the quantity of interest, namely

$$V = \int_{0^-}^{\infty} P(t) dW(t)$$

for a specific P . Let

$$P(t) = \begin{cases} 1 & t \leq T \\ e^{-\sigma(t-T)} & t > T. \end{cases} \quad (55)$$

We now evaluate V for the hybrid discipline with parameter T and also for the LIFO discipline. For the hybrid discipline we have

$$V_T = q \cdot \left[\int_0^T m(x) dx + e^{\sigma T} m^* \{ \sigma + \lambda [1 - B^*(\sigma)] \} \right. \\ \cdot (1 + h^* \{ \sigma + \lambda [1 - B^*(\sigma)] \}) \\ \left. - e^{\sigma T} (1 + h^* \{ \sigma + \lambda [1 - B^*(\sigma)] \}) \right. \\ \left. \cdot \int_0^T m(y) e^{-y(\sigma+\lambda)[1-B^*(\sigma)]} dy \right], \quad (56)$$

where

$$\left\{ \begin{array}{l} \frac{1}{1 + \int_0^T m(x) dx} \quad \rho < 1 \end{array} \right. \quad (57)$$

$$\left\{ \begin{array}{l} \frac{1}{\rho \left[1 + \int_0^T m(x) dx \right]} \quad \rho > 1. \end{array} \right. \quad (58)$$

We need some more notation before writing an expression for V_L . Let $b_{FR}(\cdot)$ denote the density function for the busy period started by the forward recurrence time of the service time. Thus,

$$b_{FR}(x) = \int_0^x b_y(x) \frac{1 - G(y)}{\mu_1} dy. \quad (59)$$

Then, for $\rho < 1$,

$$V_L = 1 - \rho + \rho \int_0^T b_{FR}(x) dx + \rho e^{\sigma T} \int_T^{\infty} e^{-\sigma x} b_{FR}(x) dx \\ = 1 - \rho + \rho \left[\int_0^T b_{FR}(x) dx - e^{\sigma T} \int_0^T e^{-\sigma x} b_{FR}(x) dx \right] \\ + e^{\sigma T} \frac{\lambda [1 - B^*(\sigma)]}{\sigma + \lambda [1 - B^*(\sigma)]}, \quad (60)$$

and, for $\rho > 1$,

$$V_L = \int_0^T b_{FR}(x) dx - e^{\sigma T} \int_0^T e^{-\sigma x} b_{FR}(x) dx + \frac{1 - B^*(\sigma)}{\mu_1 \{ \sigma + \lambda [1 - B^*(\sigma)] \}}. \quad (61)$$

For numerical calculations we considered all integrals in eqs. (56) through (58) and (60) through (61) as functions of T , obtained their Laplace transforms, and inverted the transform at the specified value of T using the inversion method of D. Jagerman.³ Thus, let

$$R_1(t) = \int_0^t m(x) dx,$$

$$R_2(t) = \int_0^t m(x) e^{-x(\sigma + \lambda [1 - B^*(\sigma)])} dx,$$

$$R_3(t) = \int_0^t b_{FR}(x) dx,$$

and

$$R_4(t) = \int_0^t e^{-\sigma x} b_{FR}(x) dx.$$

Also, for $i = 1, 2, 3, 4$, and θ in the appropriate domain, let

$$R_i^*(\theta) = \int_0^\infty e^{-\theta t} R_i(t) dt.$$

Then

$$R_1^*(\theta) = \frac{m^*(\theta)}{\theta},$$

$$R_2^*(\theta) = \frac{m^* \{ \theta + \sigma + \lambda [1 - B^*(\sigma)] \}}{\theta},$$

$$R_3^*(\theta) = \frac{1 - B^*(\theta)}{\theta \{ \theta + \lambda [1 - B^*(\theta)] \}},$$

and

$$R_4^*(\theta) = \frac{1 - B^*(\theta + \sigma)}{\sigma + \theta + \lambda [1 - B^*(\theta + \sigma)]}.$$

For numerical examples we had the service time distribution gamma with mean 1 and variance $1/K$. We used two different values of K , $K = 1$ (exponential distribution), and $K = 10$. We used two different

values of T , 1 and 3. Finally, we used two values of σ , 2.0 and 0.15. These give us eight parameter sets. The values of V_T and V_L as functions of the load $\lambda = \rho$ are given in Figs. 4 through 11. From these figures we observe:

(i) For both service disciplines, the throughput of good calls is larger for larger T , smaller σ , and larger K . The behavior of the throughput with respect to T and σ is obvious. Larger K implies smaller variability in the service time, thus reducing the probability of a customer getting served after a long wait. This, in turn, results in a higher throughput.

(ii) For the assumed customer behavior, the hybrid discipline always provides higher throughput than the LIFO discipline does. The difference is larger for larger T , larger σ , and larger K .

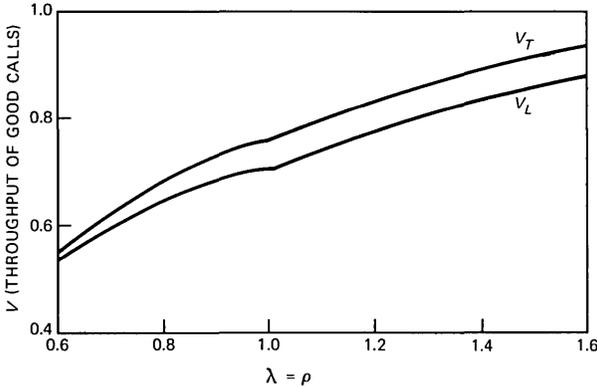


Fig. 4—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 3.0$, $\sigma = 2.0$, and $K = 1$.

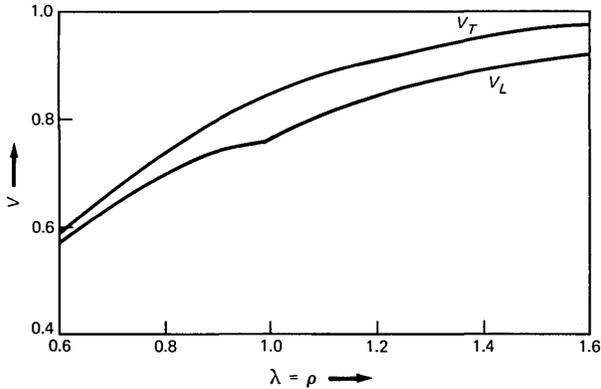


Fig. 5—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 3.0$, $\sigma = 2.0$, and $K = 10$.

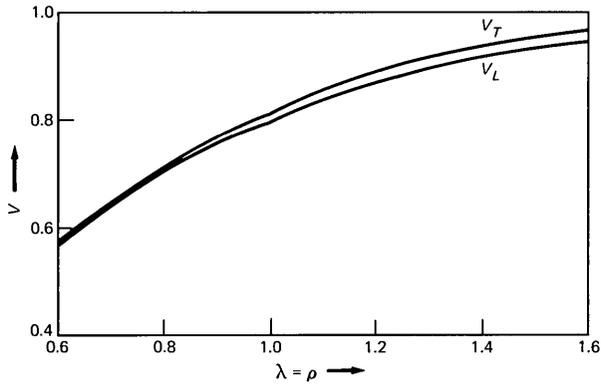


Fig. 6—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 3.0$, $\sigma = 0.15$, and $K = 1$.

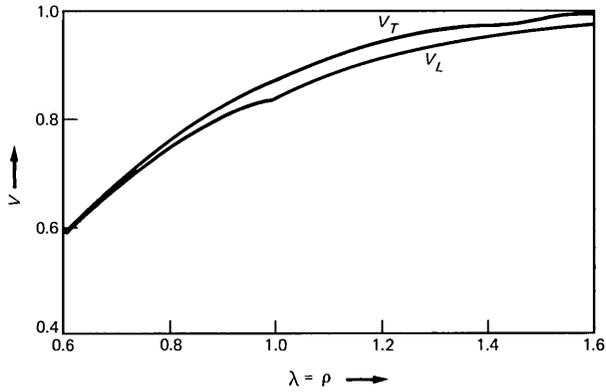


Fig. 7—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 3.0$, $\sigma = 0.15$, and $K = 10$.

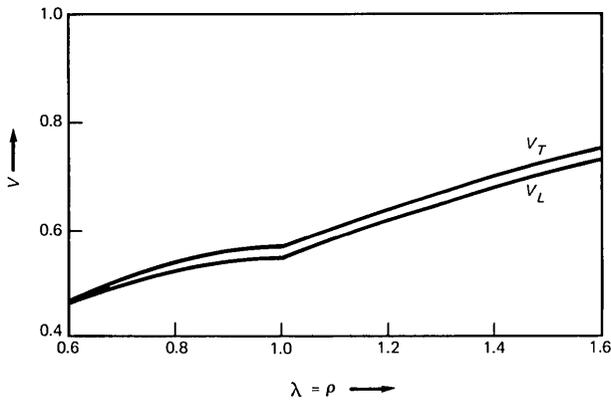


Fig. 8—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 1.0$, $\sigma = 2.0$, and $K = 1$.

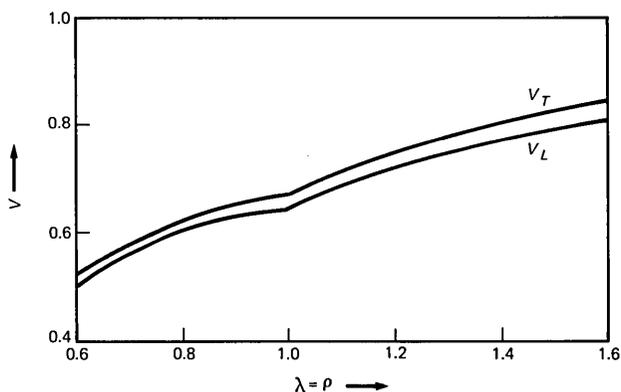


Fig. 9—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 1.0$, $\sigma = 2.0$, and $K = 10$.

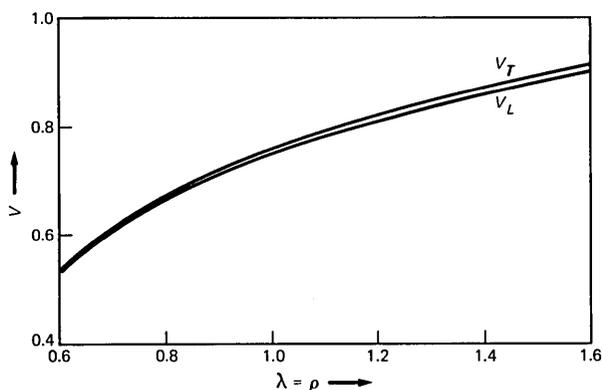


Fig. 10—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 1.0$, $\sigma = 0.15$, and $K = 1$.

Of course, our knowledge of the customer behavior may be more or less accurate, depending on the application. An issue of interest then is the sensitivity of the throughput to the assumed customer behavior. This was studied for a special case ($K = 1$) in Ref. 1. The analysis in this paper can be used to answer such issues for more general service time distributions. Qualitatively, however, the conclusions will remain the same: the last-in-first-out (LIFO) discipline is robust with respect to the knowledge of customer behavior. The hybrid discipline, on the other hand, is very sensitive to the customer behavior and should be used only when the customer behavior is adequately known and does not change in time, or when the parameters of the customer behavior can be estimated and used to change the control parameters dynamically.

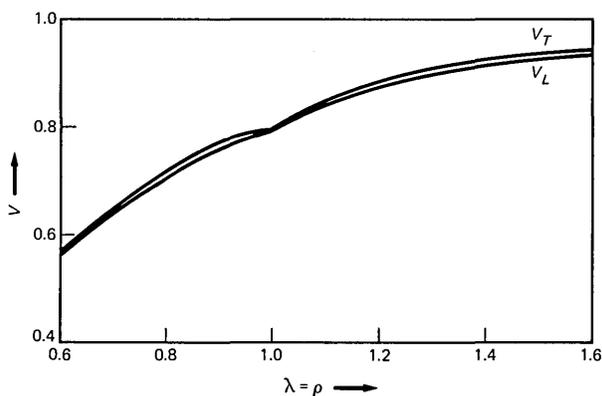


Fig. 11—The values V_T and V_L as functions of the load $\lambda = \rho$ for $T = 1.0$, $\sigma = 0.15$, and $K = 10$.

REFERENCES

1. B. T. Doshi and E. H. Lipper, "Comparisons of Service Disciplines in a Queueing System with Delay Dependent Customer Behavior," Proceedings of "Applied Probability, Computer Science, The Interface" Conf., II, Birkhauser, 1982, pp. 269-301.
2. P. H. Brill and M. J. M. Posner, "Level Crossings in Point Processes Applied to Queues: Single Server Case," *Opns. Res.*, 25 (July 1977), pp. 662-74.
3. S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, New York, NY: Academic Press, 1975, p. 183.
4. D. Jagerman, "An Inversion Technique for the Laplace Transform with Applications to Approximation," *B.S.T.J.*, 61 (October 1982), pp. 1995-2002.

Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design

By M. S. PHADKE, R. N. KACKAR, D. V. SPEENEY, and
M. J. GRIECO

(Manuscript received March 25, 1982)

In this paper we describe the off-line quality control method and its application in optimizing the process for forming contact windows in 3.5- μm complementary metal-oxide semiconductor circuits. The off-line quality control method is a systematic method of optimizing production processes and product designs. It is widely used in Japan to produce high-quality products at low cost. The key steps of off-line quality control are: (i) Identify important process factors that can be manipulated and their potential working levels; (ii) perform fractional factorial experiments on the process using orthogonal array designs; (iii) analyze the resulting data to determine the optimum operating levels of the factors (both the process mean and the process variance are considered in this analysis); (iv) conduct an additional experiment to verify that the new factor levels indeed improve the quality control.

I. INTRODUCTION AND SUMMARY

This paper describes and illustrates the off-line quality control method, which is a systematic method of optimizing a production process. It also documents our efforts to optimize the process for forming contact windows in 3.5- μm technology complementary metal-oxide semiconductor (CMOS) circuits fabricated in the Murray Hill Integrated Circuit Design Capability Laboratory (MH ICDCL). Here, by optimization we mean minimizing the process variance while keeping the process mean on target.

A typical very large scale integrated circuit (IC) chip has thousands of contact windows (e.g., a *BELLMAC**-32 microprocessor chip has

* Trademark of Bell Laboratories.

250,000 windows on an approximately 1.5-cm² area), most of which are not redundant. It is critically important to produce windows of size very near the target dimension. (In this paper windows mean contact windows.) Windows that are not open or are too small result in loss of contact to the devices, while excessively large windows lead to shorted device features. The application of the off-line quality control method has reduced the variance of the window size by a factor of four. Also, it has substantially reduced the processing time required for the window-forming step.

This study was inspired by Professor Genichi Taguchi's visit to the Quality Theory and Systems Group in the Quality Assurance Center at Bell Laboratories during the months of August, September, and October, 1980. Professor Taguchi, director of the Japanese Academy of Quality and a recipient of the Deming award, has developed the method of off-line quality control during the last three decades. It is used routinely by many leading Japanese industries to produce high-quality products at low cost. An overview of Professor Taguchi's off-line and on-line quality control methods is given in Taguchi,¹ and Kacker and Phadke.² This paper documents the results of the first application of Professor Taguchi's off-line quality control method in Bell Laboratories.

The distinctive features of the off-line quality control method are experimental design using orthogonal arrays and the analysis of signal-to-noise ratios (s/n). The orthogonal array designs provide an economical way of simultaneously studying the effects of many production factors on the process mean and variance. Orthogonal array designs are fractional factorial designs with the orthogonality property defined in Section IV. The s/n is a measure of the process variability. According to Professor Taguchi,³ by optimizing the process with respect to the s/n, we ensure that the resulting optimum process conditions are robust or stable, meaning that they have the minimum process variation.

The outline of this paper is as follows: Section II gives a brief description of the window-forming process, which is a critical step in IC fabrication. The window-forming process is generally considered to be one of the most difficult steps in terms of reproducing and obtaining uniform-size windows. Nine key process factors were identified and their potential operating levels were determined. A description of the factors and their levels is given in Section III. The total number of possible factor-level combinations is about six thousand.

The aim of the off-line quality control method is to determine a factor-level combination that gives the least variance for the window size while keeping the mean on target. To determine such a factor-level combination we performed eighteen experiments using the L_{18}

orthogonal array. The experimental setup is given in Section IV. These eighteen experiments correspond to eighteen factor-level combinations among the possible six thousand combinations. For each experiment, measurements were taken on the line width and the window-size control features. The resulting data were analyzed to determine the optimum factor-level combination. The measurements and the data analysis are presented in Sections V through IX.

The optimum factor levels, inferred from the data analysis, were subsequently used in fabricating the *BELLMAC-32* microprocessor, the *BELLMAC-4* microcomputer, and some other chips in the Murray Hill ICDCL. The experience of using these conditions is discussed in Section X.

The experiment was designed and preliminary analysis of the experimental data was performed under Professor Taguchi's guidance and collaboration.

II. THE WINDOW-FORMING PROCESS

Fabrication of integrated circuits is a complex, lengthy process.⁴ Window forming is one of the more critical steps in fabricating state of the art CMOS integrated circuits. It comes after field and gate oxides are grown; polysilicon lines have been formed; and the gate, source, and drain areas are defined by the process of doping. Figure 1 shows the windows in a cross section of a wafer. A window is a hole of about $3.5\ \mu\text{m}$ diameter etched through an oxide layer of about $2\ \mu\text{m}$ thickness. The purpose of the windows is to facilitate the interconnections between the gates, sources, and drains. For this reason these windows are called contact windows.

The process of forming windows through the oxide layers involves photolithography. First the *P*-glass surface is prepared by depositing

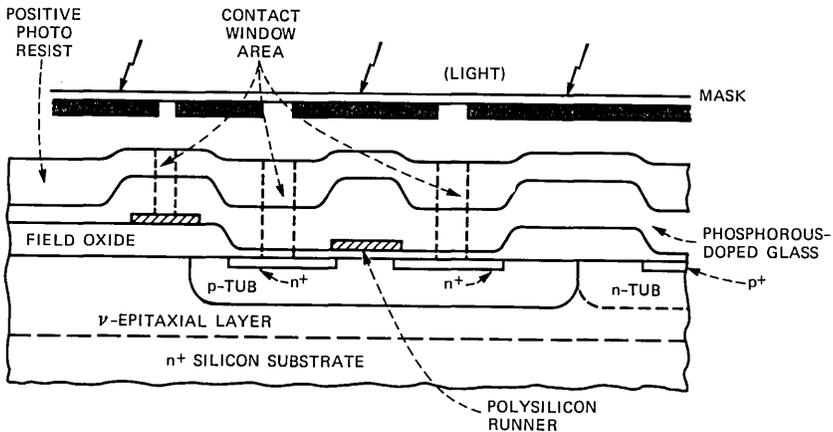


Fig. 1—Cross section of a wafer.

undoped oxide on it and prebaking it. The window-forming process is described below.

(i) Apply Photoresist: A wetting agent is sprayed on the wafer to promote adhesion of photoresist to the oxide surface. Then an appropriate photoresist is applied on the wafer and the wafer is rotated at high speed so that the photoresist spreads uniformly.

(ii) Bake: The wafer is baked to dry the photoresist layer. The thickness of the photoresist layer at this stage is about 1.3 to 1.4 μm .

(iii) Expose: The photoresist-coated wafer is exposed to ultraviolet radiation through a mask. The windows to be printed appear as clear areas on the mask. In addition to the windows, which are parts of the desired circuits, the mask has some test patterns. Light passes through these areas and causes the photoresist in the window areas and the test pattern areas to become soluble in an appropriate solvent (developer). The areas of the photoresist where light does not strike remain insoluble.

(iv) Develop: The exposed wafer is dipped in the developer, which dissolves only the exposed areas. In properly printed windows, the exposed photoresist is removed completely and the oxide surface is revealed.

(v) Plasma Etch: The wafers are placed in a high-vacuum chamber wherein a plasma is established. The plasma etches the exposed oxide areas faster than it etches the photoresist. So at the places where the windows are printed, windows are cut through the oxide layers down to the silicon surface.

(vi) Remove Photoresist: The remaining photoresist is now removed with the help of oxygen plasma and wet chemicals.

In the formation of the final contact windows there are additional steps: (vii) removal of cap-oxide, (viii) oxidation of the contact area to prevent diffusion of phosphorus in the subsequent step, (ix) reflow of the *P*-glass to round the window corners, (x) hydrogen annealing, and (xi) pre-metal wet-etching to remove any remaining oxides from the contact window areas.

At the time we started this study, the target window size at step 6 was considered to be 3.0 μm . The final target window size (after step xi) was 3.5 μm .

III. SELECTION OF FACTORS AND FACTOR LEVELS

For the present study only the steps numbered (i) through (v) were chosen for optimization. Discussions with process engineers led to the selection of the following nine factors for controlling the window size. The factors are shown next to the appropriate fabrication steps.

(i) Apply Photoresist: Photoresist viscosity (*B*) and spin speed (*C*).

- (ii) Bake: Bake temperature (*D*) and bake time (*E*).
- (iii) Expose: Mask dimension (*A*), aperture (*F*), and exposure time (*G*).
- (iv) Develop: Developing time (*H*).
- (v) Plasma etch: Etch time (*I*).

No factor was chosen corresponding to the photoresist removal step because it does not affect the window size.

The standard operating levels of the nine factors are given in Table I. Under these conditions, which prevailed in September 1980, the contact windows varied substantially in size and on many occasions even failed to print and open. Figure 2 shows a typical photograph of the programmed logic array (PLA) area of a microcomputer chip. The wide variation in window size and the presence of unopened windows is obvious from the figure.

The principle of off-line quality control is to systematically investigate various possible levels for these factors with an aim of obtaining uniform-size windows.

In the window-forming experiment a number of alternate levels were considered for each of the nine factors. These levels are also listed in Table I. Six of these factors have three levels each. Three of the factors have only two levels.

Table I—Test levels

Label	Factors Name	Levels			
		Standard Levels			
A	Mask Dimension (μm)		2	2.5	
B	Viscosity		204	206	
C	Spin Speed (rpm)	Low	Normal	High	
D	Bake Temperature ($^{\circ}\text{C}$)		90	105	
E	Bake Time (min)		20	30	40
F	Aperture		1	2	3
G	Exposure Time	20% Over	Normal	20% Under	
H	Developing Time (s)		30	45	60
I	Plasma Etch Time (min)		14.5	13.2	15.8

Dependence of spin speed on viscosity

		Spin Speed (rpm)		
		Low	Normal	High
Viscosity	204	2000	3000	4000
	206	3000	4000	5000

Dependence of exposure on aperture

		Exposure (PEP-Setting)		
		20% Over	Normal	20% Under
Aperture	1	96	120	144
	2	72	90	108
	3	40	50	60

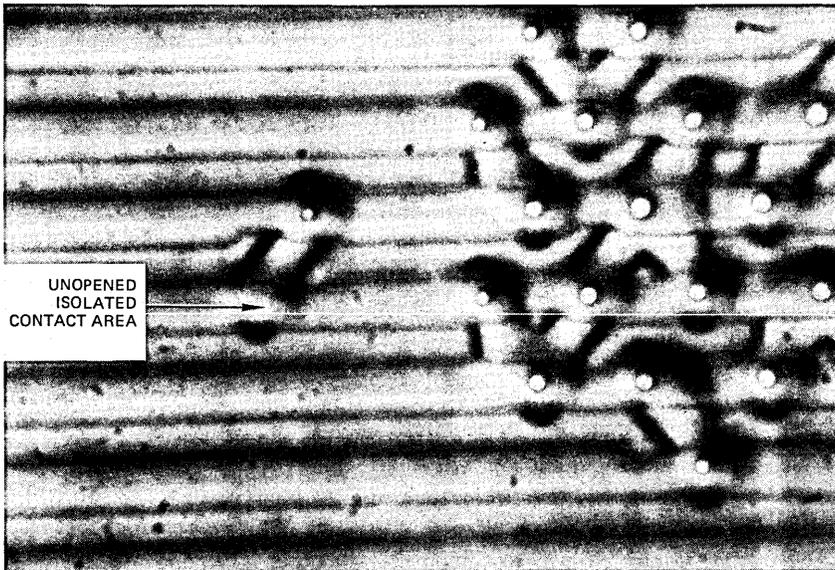


Fig. 2—Example of nonuniform contact window sizes and an isolated, unopened contact window. Both are typical results obtained in August 1980, for the PLA area of microprocessor and microcomputer chips. The contact windows are round shaped.

The levels of spin speed are tied to the levels of viscosity. For the 204 photoresist viscosity the low, normal, and high spin speeds mean 2000 rpm, 3000 rpm, and 4000 rpm, respectively. For the 206 photoresist viscosity the spin speed levels are 3000 rpm, 4000 rpm, and 5000 rpm. Likewise, the exposure setting depends on the aperture. These relationships are also shown in Table I.

IV. THE ORTHOGONAL ARRAY EXPERIMENT

The full factorial experiment to explore all possible factor-level combinations would require $3^6 \times 2^3 = 5832$ experiments. Considering the cost of material, the time, and the availability of facilities, the full factorial experiment is prohibitively large. Also from statistical considerations it is unnecessary to perform the full factorial experiment because processes can usually be adequately characterized by a relatively few parameters.

The fractional factorial design used for this study is given in Table II. It is the L_{18} orthogonal array design consisting of 18 experiments taken from Taguchi and Wu.³ The rows of the array represent runs while the columns represent the factors. Here we treat BD as a joint factor with the levels 1, 2, and 3 representing the combinations B_1D_1 , B_2D_1 , and B_1D_2 , respectively. This is done so that we can study all the

Table II—The L_{18} orthogonal array

Experiment Number	Column Number & Factor							
	1 A	2 BD	3 C	4 E	5 F	6 G	7 H	8 I
1	1	1	1	1	1	1	1	1
2	1	1	2	2	2	2	2	2
3	1	1	3	3	3	3	3	3
4	1	2	1	1	2	2	3	3
5	1	2	2	2	3	3	1	1
6	1	2	3	3	1	1	2	2
7	1	3	1	2	1	3	2	3
8	1	3	2	3	2	1	3	1
9	1	3	3	1	3	2	1	2
10	2	1	1	3	3	2	2	1
11	2	1	2	1	1	3	3	2
12	2	1	3	2	2	1	1	3
13	2	2	1	2	3	1	3	2
14	2	2	2	3	1	2	1	3
15	2	2	3	1	2	3	2	1
16	2	3	1	3	2	3	1	2
17	2	3	2	1	3	1	2	3
18	2	3	3	2	1	2	3	1

nine factors with the L_{18} orthogonal array. Thus, experiment 2 would be run under level 1 of factors A , B , and D , and level 2 of the remaining factors. In terms of the actual settings, these conditions are: 2- μ m mask dimension, 204 viscosity, 90°C bake temperature, 3000-rpm spin speed, bake time of 30 minutes, aperture 2, exposure PEP setting 90, 45-second developing time, and 13.2 minutes of plasma etch. The other rows are interpreted similarly.

Here are some of the properties and considerations of this design:

(i) This is a main-effects-only design; i.e., the response is approximated by a separable function. A function of many independent variables is called separable if it can be written as a sum of functions where each component function is a function of only one independent variable.

(ii) For estimating the main effects there are two degrees of freedom associated with each three-level factor, one degree of freedom for each two-level factor, and one degree of freedom with the overall mean. We need at least one experiment for every degree of freedom. Thus, the minimum number of experiments needed is $2 \times 6 + 1 \times 3 + 1 = 16$. Our design has 18 experiments. A single-factor-by-single-factor experiment would need only 16 experiments, two fewer than 18. But such an experiment would yield far less precise information compared with the orthogonal array experiment.^{3,5}

(iii) The columns of the array are pairwise orthogonal. That is, in every pair of columns, all combinations of levels occur and they occur an equal number of times.

(iv) Consequently, the estimates of the main effects of all factors

as shown in Table II and their associated sums of squares are independent under the assumption of normality and equality of error variance. So the significance tests for these factors are independent. Though BD is treated as a joint factor, the main effects and sums of squares of B and D can be estimated separately under the assumption of no interaction. In general, these estimates would be correlated with each other. However, these estimates are not correlated with those for any of the other seven factors.

(v) The estimates of the main effects can be used to predict the response for any combination of the parameter levels. A desirable feature of this design is that the variance of the prediction error is the same for all parameter-level combinations covered by the full factorial design.

(vi) It is known that the main-effect-only models are liable to give misleading conclusions in the presence of interactions. However, in the beginning stages of this study the interactions are assumed to be negligible. If we wished to study all two-factor interactions, with no more than 18 experiments we would have enough degrees of freedom for studying only two three-level factors, or five two-level factors! That would mean in the present study we would have to eliminate half of the process factors without any experimental evidence. Alternately, if we wished to study all the nine process factors and their two-factor interactions, we would need at least 109 experiments! Orthogonal array designs can, of course, be used to study interactions.³

(vii) Optimum conditions obtained from such an experiment have to be verified with an additional experiment. This is done to safeguard us against the potential adverse effects of ignoring the interactions among the manipulatable factors.

In conducting experiments of this kind, it is common for some wafers to get damaged or broken. Also, the wafer-to-wafer variability of window sizes is typically large. So we decided to run each experiment with two wafers.

4.1 Analysis of variance

Data collected from such experiments are analyzed by a method called analysis of variance (ANOVA).⁶ The purpose of ANOVA is to separate the total variability of the data, which is measured by the sum of the squared deviations from the mean value, into contributions by each of the factors and the error. This is analogous to the use of Parseval's theorem to separate the signal strength into contributions by the various harmonics.³ To see which of the factors have a significant effect, F -tests are performed. In performing the standard F -test we assume that the errors are normally distributed with equal variance and are independent. The results of the F -test are indicated by the significance level. When we say that a factor is significant at 5-percent

level we mean that there is 5 percent or less chance that, if we change the level of the factor, the response will remain the same. If the F -test indicates that a factor is not significant at the 5-percent level it means that, if we change the level of that factor, there is more than a 5-percent chance that the response will remain the same.

The levels of factors which are identified as significant are then set to obtain the best response. The levels of the other factors can be set at any levels within the experimental range. We choose to leave them at the starting levels.

If the assumptions of the F -test are not completely satisfied, the quoted significances are not accurate. However, the standard F -test is relatively insensitive to deviations from the assumptions used in its derivation. Thus, for making engineering decisions about which factor levels to change, the accuracy of the significance level is an adequate guide. In this paper we will use the standard F -test even though some of the assumptions are not strictly satisfied.

V. QUALITY MEASURES

The window size is the relevant quality measure for this experiment. The existing equipment does not give reproducible measurements of the sizes of windows in the functional circuits on a chip. This is because of the small size of these windows and their close proximity to one another. Therefore, test patterns—a line-width pattern and a window pattern—are provided in the upper left-hand corner of each chip. The following measurements were made on these test patterns to indicate the quality.

(i) Line width after step (iv), called the pre-etch line width or photo-line width.

(ii) Line width after step (vi), called the post-etch line width.

(iii) Size of the window test pattern after step (vi), called the post-etch window size.

Five chips were selected from each wafer for making the above measurements. These chips correspond to specific locations on a wafer—top, bottom, left, right, and center.

All three quality measures are considered to be good indicators of the size of the functional windows. However, between the geometries of the window-size pattern and the line-width pattern, the geometry of the window-size pattern is closer to the geometry of the functional windows. So, among the three quality measures, the post-etch window size may be expected to be better correlated with the size of the functional windows.

VI. EXPERIMENTAL DATA

Only thirty-four wafers were available for experimentation. So experiments 15 and 18 were arbitrarily assigned only one wafer each. One

Table III—Experimental data

Experiment No.	Line-Width Control Feature Photoresist—Nanoline Tool (Micrometers)					Comments
	Top	Center	Bottom	Left	Right	
1	2.43	2.52	2.63	2.52	2.5	
1	2.36	2.5	2.62	2.43	2.49	
2	2.76	2.66	2.74	2.6	2.53	
2	2.66	2.73	2.95	2.57	2.64	
3	2.82	2.71	2.78	2.55	2.36	
3	2.76	2.67	2.9	2.62	2.43	
4	2.02	2.06	2.21	1.98	2.13	
4	1.85	1.66	2.07	1.81	1.83	
5	—	—	—	—	—	Wafer Broke
5	1.87	1.78	2.07	1.8	1.83	
6	2.51	2.56	2.55	2.45	2.53	
6	2.68	2.6	2.85	2.55	2.56	
7	1.99	1.99	2.11	1.99	2.0	
7	1.96	2.2	2.04	2.01	2.03	
8	3.15	3.44	3.67	3.09	3.06	
8	3.27	3.29	3.49	3.02	3.19	
9	3.0	2.91	3.07	2.66	2.74	
9	2.73	2.79	3.0	2.69	2.7	
10	2.69	2.5	2.51	2.46	2.4	
10	2.75	2.73	2.75	2.78	3.03	
11	3.2	3.19	3.32	3.2	3.15	
11	3.07	3.14	3.14	3.13	3.12	
12	3.21	3.32	3.33	3.23	3.10	
12	3.48	3.44	3.49	3.25	3.38	
13	2.6	2.56	2.62	2.55	2.56	
13	2.53	2.49	2.79	2.5	2.56	
14	2.18	2.2	2.45	2.22	2.32	
14	2.33	2.2	2.41	2.37	2.38	
15	2.45	2.50	2.51	2.43	2.43	
15	—	—	—	—	—	No wafer
16	2.67	2.53	2.72	2.7	2.6	
16	2.76	2.67	2.73	2.69	2.6	
17	3.31	3.3	3.44	3.12	3.14	
17	3.12	2.97	3.18	3.03	2.95	
18	3.46	3.49	3.5	3.45	3.57	
18	—	—	—	—	—	No wafer

of the wafers assigned to experiment 5 broke in handling. So experiments 5, 15, and 18 have only one wafer.

The experimental data are shown in Table III.

The data arising from such experiments can be classified as two types—continuous data and categorical data. Here, the pre-etch and the post-etch line-width data are of the continuous type. The post-etch window size data are mixed categorical-continuous type, because some windows are open while some are not. The two types of data are analyzed somewhat differently, as we explain the following two sections.

VII. ANALYSIS OF THE LINE-WIDTH DATA

Both the pre-etch and the post-etch line widths are continuous variables. For each of these variables the statistics of interest are the

Table III—Experimental data (Continued)

Experiment No.	Line-Width Control Feature Etched—Nanoline Tool (Micrometers)					Comments
	Top	Center	Bottom	Left	Right	
1	2.95	2.74	2.85	2.76	2.7	
1	3.03	2.95	2.75	2.82	2.85	
2	3.05	3.18	3.2	3.16	3.06	
2	3.25	3.15	3.09	3.11	3.16	
3	3.69	3.57	3.78	3.55	3.40	
3	3.92	3.62	3.71	3.71	3.53	
4	2.68	2.62	2.9	2.45	2.7	
4	2.29	2.31	2.77	2.46	2.49	
5	—	—	—	—	—	Wafer Broke
5	1.75	1.15	2.07	2.12	1.53	
6	3.42	2.98	3.22	3.13	3.17	
6	3.34	3.21	3.23	3.25	3.28	
7	2.62	2.49	2.53	2.41	2.51	
7	2.76	2.94	2.68	2.62	2.51	
8	4.13	4.38	4.41	4.03	4.03	
8	4.0	4.02	4.18	3.92	3.91	
9	3.94	3.82	3.84	3.57	3.71	
9	3.44	3.30	3.41	3.28	3.20	
10	3.17	2.85	2.84	3.06	2.94	
10	3.70	3.34	3.45	3.41	3.29	
11	4.01	3.91	3.92	3.80	3.90	
11	3.67	3.31	2.86	3.41	3.23	
12	4.04	3.80	4.08	3.81	3.94	
12	4.51	4.37	4.45	4.24	4.48	
13	3.40	3.12	3.11	3.25	3.06	
13	3.22	3.03	2.89	2.92	2.98	
14	3.18	3.03	3.4	3.17	3.32	
14	3.18	2.83	3.17	3.07	3.02	
15	2.86	2.46	2.3	2.6	2.55	
15	—	—	—	—	—	No wafer
16	2.85	2.14	1.22	2.8	3.03	
16	3.4	2.97	2.96	2.87	2.88	
17	4.06	3.87	3.90	3.94	3.87	
17	4.02	3.49	3.51	3.69	3.47	
18	4.49	4.28	4.34	4.39	4.25	
18	—	—	—	—	—	No wafer

mean and the standard deviation. The objective of our data analysis is to determine the factor-level combination such that the standard deviation is minimum while keeping the mean on target. We will call this the optimum factor-level combination. Professor Taguchi's method for obtaining the optimum combination is given next.

7.1 Single response variable

Let us first consider the case where there is only one response variable. Instead of working with the mean and the standard deviation, it is preferable to work with the transformed variables—the mean and the signal-to-noise ratio (s/n). The s/n is defined as

$$\begin{aligned}
 s/n &= \log_{10} \left(\frac{\text{Mean}}{\text{Standard Deviation}} \right) \\
 &= -\log_{10}(\text{coefficient of variation}).
 \end{aligned}$$

Table III—Experimental data (Continued)

Experiment No.	Window-Control Feature Etched—Vickers Tool (Micrometers)					Comments
	Top	Center	Bottom	Left	Right	
1	WNO*	WNO	WNO	WNO	WNO	
1	WNO	WNO	WNO	WNO	WNO	
2	2.32	2.23	2.30	2.56	2.51	
2	2.22	2.33	2.34	2.15	2.35	
3	2.98	3.14	3.02	2.89	3.16	
3	3.15	3.08	2.78	WNO	2.86	
4	WNO	WNO	WNO	WNO	WNO	
4	WNO	WNO	WNO	WNO	WNO	
5	—	—	—	—	—	Wafer Broke
5	WNO	WNO	WNO	WNO	WNO	
6	2.45	2.19	2.14	2.32	2.12	
6	WNO	WNO	WNO	WNO	WNO	
7	WNO	WNO	WNO	WNO	WNO	
7	WNO	WNO	WNO	WNO	WNO	
8	WNO	WNO	WNO	WNO	WNO	
8	2.89	2.97	3.13	3.25	3.19	
9	3.16	2.91	3.12	3.18	3.11	
9	2.43	2.35	2.14	2.40	2.28	
10	2.0	1.75	1.97	1.91	1.72	
10	WNO	2.7	WNO	2.61	2.73	
11	2.76	3.09	3.22	3.05	3.04	
11	3.12	3.21	WNO	2.71	2.27	
12	3.24	3.08	WNO	2.89	2.72	
12	3.5	3.71	3.52	3.53	3.71	
13	2.54	2.63	2.88	2.31	2.71	
13	WNO	WNO	WNO	WNO	WNO	
14	WNO	1.74	2.24	2.07	2.38	
14	WNO	WNO	WNO	WNO	WNO	
15	WNO	WNO	WNO	WNO	WNO	
15	—	—	—	—	—	No wafer
16	WNO	WNO	WNO	WNO	WNO	
16	WNO	WNO	WNO	WNO	WNO	
17	3.09	2.91	3.06	3.09	3.29	
17	3.39	2.5	2.57	2.62	2.35	
18	3.39	3.34	3.45	3.44	3.33	
18	—	—	—	—	—	No wafer

* WNO—Window not open.

In terms of the transformed variables, the optimization problem is to determine the optimum factor levels such that the s/n is maximum while keeping the mean on target. This problem can be solved in two stages:

(i) Determine which factors have a significant effect on the s/n . This is done through the analysis of variance (ANOVA) of the s/n . These factors are called the *control factors*, implying that they control the process variability. For each control factor we choose the level with the highest s/n as the optimum level. Thus the overall s/n is maximized.

(ii) Select a factor that has the smallest effect on the s/n among all factors that have a significant effect on the mean. Such a factor is called a *signal factor*. Ideally, the signal factor should have no effect

on the s/n . Choose the levels of the remaining factors (factors that are neither control factors nor signal factors) to be the nominal levels prior to the optimization experiment. Then set the level of the signal factor so that the mean response is on target.

In practice, the following two aspects should also be considered in selecting the signal factor: (i) If possible, the relationship between the mean response and the levels of the signal factor should be linear, and (ii) It should be convenient to change the signal factor during production. These aspects are important from the on-line quality control considerations. The signal factor can be used during manufacturing to adjust the mean response.¹⁻³

Why do we work in terms of the s/n ratio rather than the standard deviation? Frequently, as the mean decreases, the standard deviation also decreases and vice versa. In such cases, if we work in terms of the standard deviation, the optimization cannot be done in two steps; i.e., we cannot minimize the standard deviation first and then bring the mean on target.

Through many applications, Professor Taguchi has empirically found that the two-stage optimization procedure involving the s/n indeed gives the parameter-level combination where the standard deviation is minimum, while keeping the mean on target. This implies that the engineering systems behave in such a way that the manipulatable production factors can be divided into three categories:

(i) Control factors, which affect process variability as measured by the s/n

(ii) Signal factors, which do not influence (or have negligible effect on) the s/n but have a significant effect on the mean

(iii) Factors that do not affect the s/n or the process mean.

The two-stage procedure also has an advantage over a procedure that directly minimizes the mean square error from the target mean value. In practice, the target mean value may change during the process development. The advantage of the two-stage procedure is that for any target mean value (of course, within reasonable bounds) the new optimum factor-level combination is obtained by suitably adjusting the level of only the signal factor. This is so because in step (i) of the algorithm the coefficient of variation is minimized for every mean target value.

7.2 Multiple response variables

Now let us consider the case where there are two or more response variables. In such cases, engineering judgment may have to be used to resolve the conflict if different response variables suggest different levels for any one factor. The modified two-stage procedure is as follows:

(i) Separately determine control factors and their optimum levels corresponding to each response variable. If there is a conflict between the optimum levels suggested by the different response variables, use engineering judgment to resolve the conflict.

(ii) Select a factor that has the smallest effect (preferably no effect) on the signal-to-noise ratios for all the response variables but has a significant effect on the mean levels. This is the signal factor. Set the levels of the remaining factors, which affect neither the mean nor the s/n , at the nominal levels prior to the optimization experiment. Then set the level of the signal factor so that the mean responses are on target. Once again engineering judgment may have to be used to resolve any conflicts that arise.

The selection of the control factors, signal factor, and their optimum levels for the present application will be discussed in Section IX. The remaining portions of Sections VII and VIII contain the data analysis that forms the basis for selecting the optimum factor levels.

7.3 Pre-etch line width

Mean, standard deviation, and s/n were calculated for each of the eighteen experiments. For those experiments with two wafers, ten data points were used in these calculations. When there was only one wafer, five data points were used. These results are shown in Table IV. The presence of unequal sample sizes has been ignored in the subsequent analysis. Let \bar{x}_i and η_i denote the mean and the s/n for the i th experiment.

Table IV—Pre-etch line-width data

Experiment Number	Mean Line Width, \bar{x} (μm)	Standard Deviation of Line Width, s (μm)	s/n $\eta = \log(\bar{x}/s)$
1	2.500	0.0827	1.4803
2	2.684	0.1196	1.3512
3	2.660	0.1722	1.1889
4	1.962	0.1696	1.0632
5	1.870	0.1168	1.2043
6	2.584	0.1106	1.3686
7	2.032	0.0718	1.4520
8	3.267	0.2101	1.1917
9	2.829	0.1516	1.2709
10	2.660	0.1912	1.1434
11	3.166	0.0674	1.6721
12	3.323	0.1274	1.4165
13	2.576	0.0850	1.4815
14	2.308	0.0964	1.3788
15	2.464	0.0385	1.8065
16	2.667	0.0706	1.5775
17	3.156	0.1569	1.3036
18	3.494	0.0473	1.8692

By computing a single mean \bar{x}_i and a single variance s_i^2 (needed for computing η_i) from the two wafers of each experiment i , we pool together the between wafer and the within wafer variance. That is,

$$E(s_i^2) = (\text{Between wafer variance for experiment } i) \times \frac{5}{9} \\ + (\text{Within wafer variance for experiment } i).$$

Thus, when we maximize η , we minimize the sum of the between-wafer and the within-wafer variances of the line width, which is the response of interest to us. There can be situations when one wants to separately estimate the effects of the factor levels on the between-wafer and within-wafer variances. In those cases, one would compute the s/n and the mean line width for each individual wafer.

In the analysis of the pre-etch and the post-etch line widths, we compute the \bar{x}_i and the s_i^2 for each experiment by pooling the data from both wafers used in that experiment. A relative measure of the between-wafer and within-wafer variance is obtained in Section VIII, while the post-etch window-size data is being analyzed.

7.3.1 Analysis of s/n

The estimates of the average s/n for all factor levels are given in Table V. The average for the first level of factor A is the average of the nine experiments (experiments 1 through 9), which were conducted with level 1 of the factor A. Likewise, the average for the second level of factor A is the mean of experiments 10 through 18, which were conducted with level 2 of the factor A. Let us denote these average effects of A_1 and A_2 by m_{A_1} and m_{A_2} , respectively. Here $m_{A_1} = 1.2857$ and $m_{A_2} = 1.5166$. The other entries of Table V were calculated similarly.

The average signal-to-noise ratios for every level of the eight factors are graphically shown in Fig. 3. Qualitatively speaking, the mask dimension and the aperture cause a large variation in the s/n. The

Table V—Pre-etch line width for average s/n

Factor	Average s/n		
	Level 1	Level 2	Level 3
A Mask Dimension	1.2857	1.5166	
BD Viscosity Bake Temperature	(B_1D_1) 1.3754	(B_2D_1) 1.3838	(B_1D_2) 1.4442
B Viscosity	1.4098	1.3838	
D Bake Temperature	1.3796	1.4442	
C Spin Speed	1.3663	1.3503	1.4868
E Bake Time	1.4328	1.4625	1.3082
F Aperture	1.5368	1.4011	1.2654
G Exposure Time	1.3737	1.3461	1.4836
H Developing Time	1.3881	1.4042	1.4111

Overall average s/n = 1.4011.

developing time and the viscosity cause a small change in the s/n. The effect of the other factors is in between.

For three-level factors, Fig. 3 can also be used to judge the linearity of the effect of the factors. If the difference between levels 1 and 2, and levels 2 and 3 is equal and these levels appear in proper order (1, 2, 3, or 3, 2, 1), then the effect of that factor is linear. If either the differences are unequal or the order is mixed up, then the effect is not linear. For example, the aperture has approximately linear response while the bake time has a nonlinear response.

We shall perform a formal analysis of variance (ANOVA) to identify statistically significant factors. The analysis of variance of general linear models is widely known in literature, see e.g., Searle⁷ and Hicks.⁶ Simple ANOVA methods for orthogonal array experiments are described in Taguchi and Wu.³ The linear model used in analyzing this data is:

$$y_i = \mu + x_i + e_i, \quad (1)$$

where

$i = 1, \dots, 18$ is the experiment number.

μ is the overall mean.

x_i is the fixed effect of the factor-level combination used in experiment i . Here we consider only the main effect for each of the factors. Thus it represents the sum of the effects of the eight factors.

e_i is the random error for experiment i .

y_i is the s/n for experiment i .

To clarify the meaning of the term x_i , let us consider experiment 1, which was run at level 1 of each of the eight factors A through H. Note that the factor I is irrelevant for studying the pre-etch line width. So x_1 is the sum of the main effects associated with the first level of each of the factors A through H.

The sum of squares and the mean squares for the eight factors are tabulated in Table VIa. The computations are illustrated in Appendix A.

The expected mean squares are also shown in Table VIa. See Refs. 6 and 7 for the computation of expected mean squares, which are used in forming appropriate F-tests. The error variance, i.e., variance of e_i , is denoted by σ^2 . The variability due to the factors A through H is denoted by ϕ with an appropriate subscript.

In Table VIa we see that the mean sum of squares for factors BD, C, G, and H are smaller than the mean error sum of squares. So a new ANOVA table, Table VIb, was formed by pooling the sum of squares of these factors with the error sum of squares. The linear model underlying the ANOVA Table VIb is the same as Eq. (1), except that

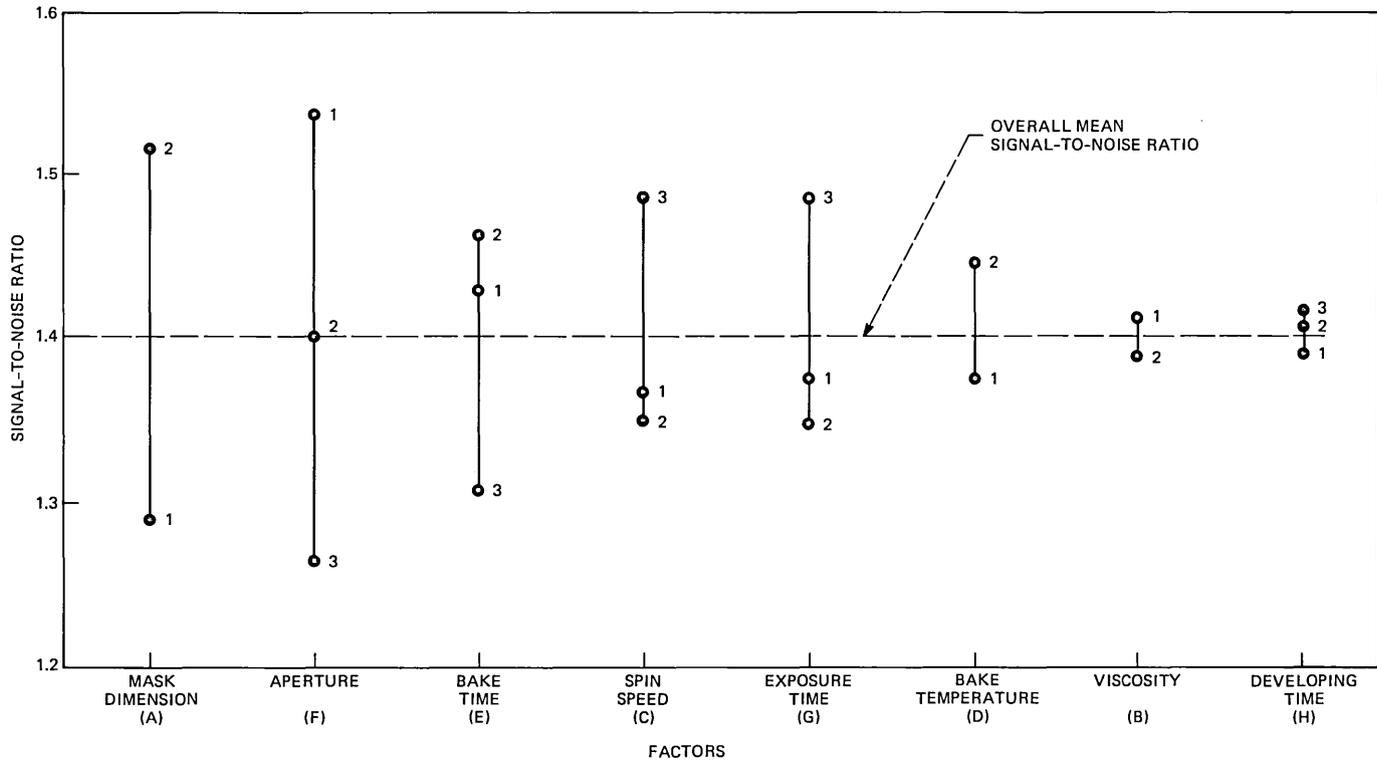


Fig. 3—Signal-to-noise ratios for pre-etch line width. The average s/n for each factor level is indicated by a dot. The number next to the dot indicates the factor level.

now x_i stands for the sum of the main effects of only A, E, and F. The F ratios, computed by dividing the factor mean square by the error mean square, are also shown in Table VIb. Factors A and F are significant using F-table values for the 5-percent significance level. So the mask dimension and the aperture are the control factors.

In performing the analysis of variance, we have tacitly assumed that the response for each experiment, here the s/n, has a normal distribution with constant variance. We are presently investigating the distributional properties of the s/n and their impact on the analysis of variance. In this paper we treat the significance levels as approximate.

The engineering significance of a statistically significant factor can be measured in terms of the percent contribution, a measure introduced by Taguchi.³ The percent contribution is equal to the percent of the total sum of squares explained by that factor after an appropriate estimate of the error sum of squares has been removed from it. The larger the percent contribution, the more can be expected to be achieved by changing the level of that factor. Computation of the percent contribution is illustrated in Appendix B, and the results are shown in Table VIb.

From Table VIb we see that both the factors A (mask dimension) and F (aperture) contribute in excess of 20 percent each to the total

Table VI—Pre-etch line width

(a) ANOVA for s/n					
Source	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square	
A Mask Dimension	1	0.2399	0.2399	$\sigma^2 + \phi_A$	
BD Viscosity Bake Temperature	2	0.0169	0.0085	$\sigma^2 + \phi_{BD}$	
C Spin Speed	2	0.0668	0.0334	$\sigma^2 + \phi_C$	
E Bake Time	2	0.0804	0.0402	$\sigma^2 + \phi_E$	
F Aperture	2	0.2210	0.1105	$\sigma^2 + \phi_F$	
G Exposure Time	2	0.0634	0.0317	$\sigma^2 + \phi_G$	
H Developing Time	2	0.0017	0.0009	$\sigma^2 + \phi_H$	
Error	4	0.1522	0.0381	σ^2	
Total	17	0.8423			

(b) Pooled ANOVA for s/n					
Source	Degrees of Freedom	Sum of Squares	Mean Square	F	Percent Contribution
A Mask Dimension	1	0.2399	0.2399	9.56*	25.5
E Bake Time	2	0.0804	0.0402	1.60	3.6
F Aperture	2	0.2210	0.1105	4.40*	20.3
Error	12	0.3010	0.0251		50.6
Total	17	0.8423			100.00

$$F_{1,12}(0.95) = 4.75.$$

$$F_{2,12}(0.95) = 3.89.$$

* Factors significant at 95-percent confidence level.

sum of squares. So the factors A and F are not only statistically significant, they have a sizable influence on the s/n. These results are consistent with Fig. 3. They will be used in Section IX for selecting the control factors.

7.3.2 Analysis of the means

Now we analyze the mean pre-etch line widths, \bar{x}_i values, to find a signal factor.

The estimates of the mean line widths for all factor levels are given in Table VII. These estimates are graphically shown in Fig. 4. It is apparent that the levels of viscosity, mask dimension, and spin speed cause a relatively large change in the mean line width. Developing time and aperture have a small effect on the line width. The remaining two factors have an intermediate effect.

The linear model used to analyze this data is the same as eq. (1), except that now y_i stands for the mean pre-etch line width rather than the s/n.

The original and the pooled ANOVA tables for the mean pre-etch line width are given in Tables VIIIa and b, respectively. Because the design is not orthogonal with respect to the factors B and D, we need a special method, described in Appendix C, to separate S_{BD} into S_B and S_D .

It is clear from Table VIIIb that the mask dimension (A), viscosity (B), and spin speed (C) have a statistically significant effect on the mean pre-etch line width. Also, these factors together contribute more than 70 percent to the total sum of squares. These results will be used in Section IX for selecting the signal factor.

7.4 Post-etch line width

The analysis of the post-etch line-width data is similar to the analysis of the pre-etch line-width data. The mean, the standard deviation, and the s/n for each experiment are shown in Table IX.

Table VII—Pre-etch line width for the mean line width

Factor	Mean Line Width (μm)		
	Level 1	Level 2	Level 3
A Mask Dimension	2.39	2.87	
BD Viscosity Bake Temperature	(B_1D_1) 2.83	(B_2D_1) 2.31	(B_1D_2) 2.74
B Viscosity	2.79	2.31	
D Bake Temperature	2.57	2.74	
C Spin Speed	2.40	2.59	2.89
E Bake Time	2.68	2.68	2.53
F Aperture	2.68	2.56	2.64
G Exposure Time	2.74	2.66	2.49
H Developing Time	2.60	2.60	2.69

Overall mean line width = 2.63 μm .

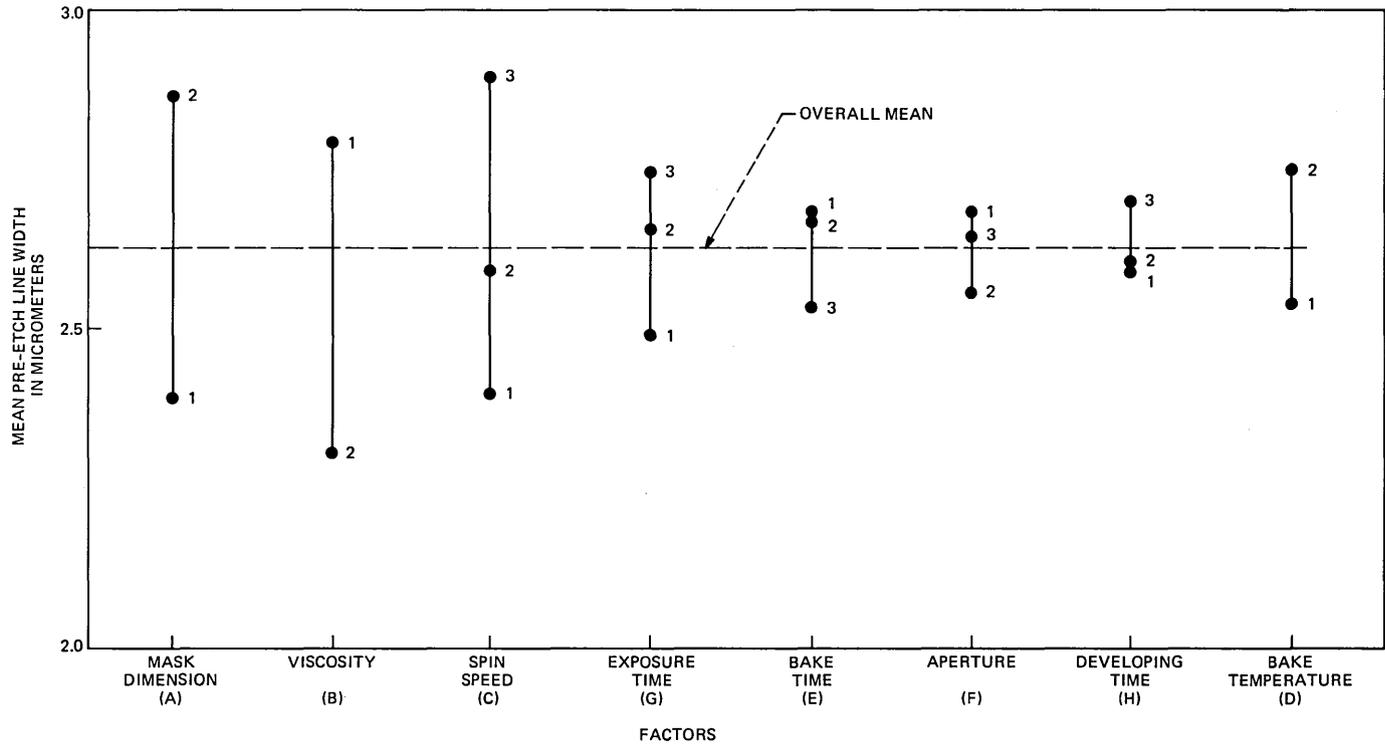


Fig. 4—Mean pre-etch line width. The mean line width for each factor level is indicated by a dot. The number next to each dot indicates the factor level.

Table VIII—Pre-etch line width

(a) ANOVA for mean line width					
Source		Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square
A	Mask Dimension	1	1.05	1.050	$\sigma^2 + \phi_A$
BD	Viscosity Bake Temperature	2	0.95	0.475	$\sigma^2 + \phi_{BD}$
C	Spin Speed	2	0.73	0.365	$\sigma^2 + \phi_C$
E	Bake Time	2	0.10	0.050	$\sigma^2 + \phi_E$
F	Aperture	2	0.05	0.025	$\sigma^2 + \phi_F$
G	Exposure Time	2	0.19	0.095	$\sigma^2 + \phi_G$
H	Developing Time	2	0.04	0.020	$\sigma^2 + \phi_H$
Error		4	0.26	0.065	σ^2
Total		17	3.37		

(b) Pooled ANOVA for mean line width						
Source		Degrees of Freedom	Sum of Squares	Mean Square	F	Contribution Percent
A	Mask Dimension	1	1.05	1.050	19.81*	29.6
B	Viscosity	1	0.83	0.834	15.74*	22.6
C	Spin Speed	2	0.73	0.365	6.89*	18.5
G	Exposure Time	2	0.19	0.095	1.79	2.5
Error		11	0.58	0.053		26.8
Total		17	3.37			100.0

$F_{1,11}(0.95) = 4.84.$

$F_{2,11}(0.95) = 3.98.$

* Factors significant at 95-percent confidence level.

Table IX—Post-etch line-width data

Experiment Number	Mean Line Width, \bar{x} (μm)	Standard Deviation of Line Width, s (μm)	s/n $\eta = \log(\bar{x}/s)$
1	2.84	0.11	1.42
2	3.14	0.063	1.70
3	3.65	0.15	1.40
4	2.57	0.20	1.11
5	1.72	0.40	0.63
6	3.12	0.27	1.07
7	2.62	0.19	1.14
8	4.10	0.18	1.37
9	3.55	0.26	1.13
10	3.31	0.35	0.98
11	3.60	0.38	0.98
12	4.17	0.27	1.18
13	3.10	0.16	1.29
14	3.14	0.16	1.29
15	2.55	0.21	1.09
16	2.81	0.37	0.88
17	3.78	0.22	1.23
18	4.34	0.078	1.75

The average s/n and the mean line width for each factor level are shown in Tables Xa and b, respectively.

The linear model (1) was again used to analyze the post-etch line-width data. The ANOVA for the signal-to-noise ratios, Table XIa, indicates that none of the nine process factors has a significant effect (approximately 5-percent level) on the s/n for the post-etch line width. The pooled ANOVA for the mean post-etch line widths is shown in Table XIb. It is obvious from the table that the viscosity, exposure, spin speed, mask dimension, and developing time have significant effects (5-percent level) on the mean line width. The contribution of these factors to the total sum of squares exceeds 90 percent. The mean line width for each factor level is shown graphically in Fig. 5.

VIII. ANALYSIS OF POST-ETCH WINDOW-SIZE DATA

Some windows are printed and open while the others are not. Thus the window-size data are mixed categorical-continuous in nature. Analysis of such data is done by converting all the data to the categorical type and then using the 'accumulation analysis' method, which is

Table X—Post-etch line width

(a) Average signal-to-noise ratios				
Factor	Average s/n			
	Level 1	Level 2	Level 3	
A Mask Dimension	1.22	1.19		
BD Viscosity Bake Temperature	(B_1D_1) 1.28	(B_2D_1) 1.08	(B_1D_2) 1.25	
B Viscosity	1.27	1.08		
D Bake Temperature	1.18	1.25		
C Spin Speed	1.14	1.20	1.27	
E Bake Time	1.16	1.28	1.17	
F Aperture	1.28	1.22	1.11	
G Exposure Time	1.26	1.33	1.02	
H Developing Time	1.09	1.20	1.32	
I Etch Time	1.21	1.18	1.23	
Overall average s/n = 1.205				
(b) Mean line width				
Factor	Mean Line Width (μm)			
	Level 1	Level 2	Level 3	
A Mask Dimension	3.03	3.42		
BD Viscosity Bake Temperature	(B_1D_1) 3.45	(B_2D_1) 2.70	(B_1D_2) 3.53	
B Viscosity	3.49	2.70		
D Bake Temperature	3.08	3.53		
C Spin Speed	2.88	3.25	3.56	
E Bake Time	3.15	3.18	3.35	
F Aperture	3.28	3.22	3.18	
G Exposure Time	3.52	3.34	2.83	
H Developing Time	3.04	3.09	3.56	
I Etch Time	3.14	3.22	3.32	
Overall mean line width = 3.23 μm .				

described by Taguchi in Refs. 3 and 8. Factors that are found significant in this analysis are control factors.

The window sizes were divided into the following five categories:

Category	Description (micrometers)
I	Window not open or not printed
II	(0, 2.25)
III	[2.25, 2.75)
IV	[2.75, 3.25]
V	(3.25, ∞)

Note that these categories are ordered with respect to window size. The target window size at the end of step (*vi*) was 3 μm . Thus category IV is the most desired category, while category I is the least desired category. Table XII summarizes the data for each of the experiments by categories. To simplify our analysis, we shall presume that a missing wafer has the same readings as the observed wafer for that experiment. This is reflected in Table XII, where we show the combined readings for the two wafers of each experiment.

Table XI—Post-etch line width

a) ANOVA for s/n						
Source		Degrees of Freedom	Sum of Squares	Mean Square	F	
A	Mask Dimension	1	0.005	0.005	0.02	
B	Viscosity	1	0.134	0.134	0.60	
D	Bake Temperature	1	0.003	0.003	0.01	
C	Spin Speed	2	0.053	0.027	0.12	
E	Bake Time	2	0.057	0.028	0.13	
F	Aperture	2	0.085	0.043	0.19	
G	Exposure Time	2	0.312	0.156	0.70	
H	Developing Time	2	0.156	0.078	0.35	
I	Etch Time	2	0.008	0.004	0.02	
Error		2	0.444	0.222		
Total		17	1.257			
b) Pooled ANOVA for mean line width						
Source		Degrees of Freedom	Sum of Squares	Mean Square	F	Percent Contribution
A	Mask Dimension	1	0.677	0.677	16.92*	8.5
B	Viscosity	1	2.512	2.512	63.51*	32.9
C	Spin Speed	2	1.424	0.712	17.80*	17.9
G	Exposure Time	2	1.558	0.779	19.48*	19.6
H	Developing Time	2	0.997	0.499	12.48*	12.2
Error		9	0.356	0.040		8.9
Total		17	7.524			100.0

$$F_{1,9}(0.95) = 5.12.$$

$$F_{2,9}(0.95) = 4.26.$$

* Factors significant at 95-percent confidence level.

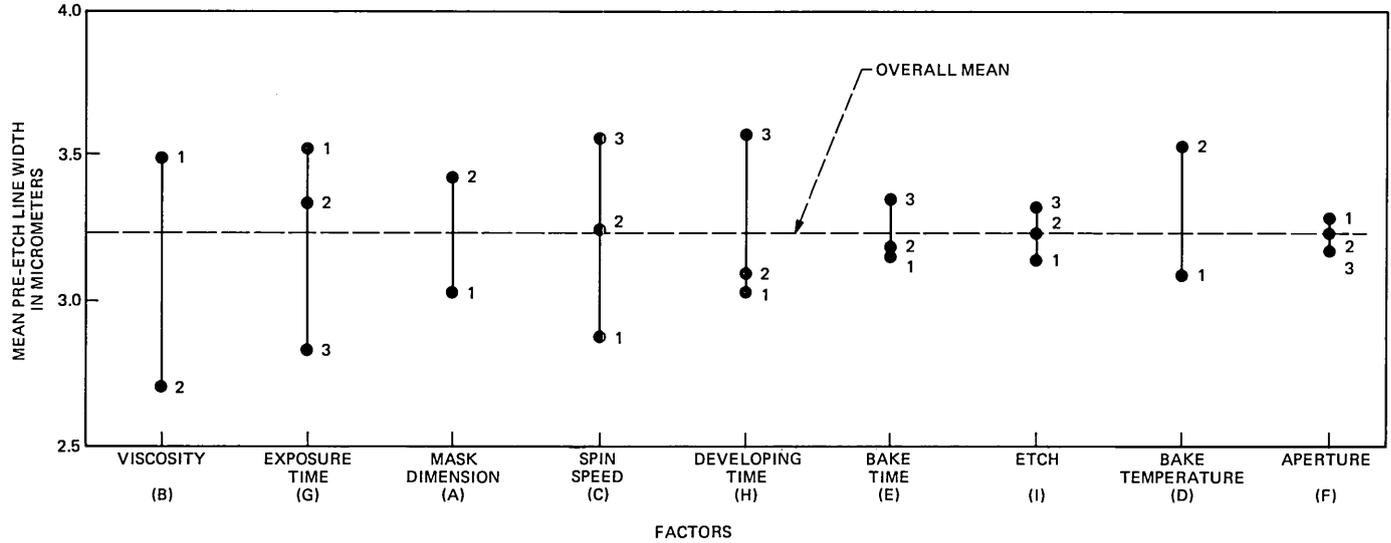


Fig. 5—Mean post-etch line width. The mean line width for each factor level is indicated by a dot. The number next to a dot indicates the factor level.

Table XII—Post-etch window-size data—frequencies by experiment

Experiment No.	Frequency Distribution for Wafer 1					Frequency Distribution for Wafer 2					Combined Frequency for the Two Wafers				
	I	II	III	IV	V	I	II	III	IV	V	I	II	III	IV	V
1	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
2	0	1	0	2	2	0	2	3	0	0	0	3	3	2	2
3	0	0	0	4	0	1	0	0	3	0	1	0	0	9	0
4	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
5	*	*	*	*	*	5	0	0	0	0	10	0	0	0	0
6	0	3	2	0	0	5	0	0	0	0	5	3	2	0	0
7	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
8	5	0	0	0	0	0	0	0	5	0	5	0	0	5	0
9	0	0	0	5	0	0	1	4	0	0	0	1	4	5	0
10	0	5	0	0	0	2	0	3	0	0	2	5	3	0	0
11	0	0	0	5	0	1	1	2	1	0	1	1	2	6	0
12	1	0	1	3	0	0	0	0	0	4	1	0	1	3	5
13	0	0	3	2	0	5	0	0	0	0	5	0	3	2	0
14	1	3	1	0	0	5	0	0	0	0	6	3	1	0	0
15	5	0	0	0	0	*	*	*	*	*	10	0	0	0	0
16	5	0	0	0	0	5	0	0	0	0	10	0	0	0	0
17	0	0	0	3	2	0	0	4	0	1	0	0	4	3	3
18	0	0	0	0	5	*	*	*	*	*	0	0	0	0	10

* Implies data missing.

Table XIII gives the frequency distribution corresponding to each level of each factor. To obtain the frequency distribution for a specific level of a specific factor, we summed the frequencies of all the experiments that were conducted with that particular level of that particular factor. For example, the frequency distribution for the first level of factor C (low spin speed) was obtained by summing the frequency distributions of experiments with serial numbers 1, 4, 7, 10, 13, and 16. These six experiments were conducted with level 1 of factor C.

The frequency distributions of Table XIII are graphically displayed by star plots in Fig. 6. From this figure and the table it is apparent that a change in the level of viscosity, spin speed, or mask dimension causes a noticeable change in the frequency distribution. A change in the level of etch time, bake time, or bake temperature seems to have only a small effect on the frequency distribution. The effects of the other factors are intermediate.

We now determine which factors have a significant effect on the frequency distribution of the window sizes. The standard chi-square test for multinomial distributions is not appropriate here because the categories are ordered. The accumulation analysis method has an intuitive appeal and has been empirically found by Professor Taguchi to be effective in analyzing ordered categorical data. The method consists of the following three steps:

(i) Compute the cumulative frequencies. Table XIII shows the cumulative frequencies for all factor levels. The cumulative categories are denoted with parentheses. Thus (III) means sum of categories I,

Table XIII—Post-etch window-size data—frequencies by factor level

Factor Levels	Frequencies					Cumulative Frequencies				
	I	II	III	IV	V	(I)	(II)	(III)	(IV)	(V)
Mask Dimension										
A_1	51	7	9	21	2	51	58	67	88	90
A_2	35	9	14	14	18	35	44	58	72	90
Viscosity, Bake Temperature										
B_1D_1	15	9	9	20	7	15	24	33	53	60
B_2D_1	46	6	6	2	0	46	52	58	60	60
B_1D_2	25	1	8	13	13	25	26	34	47	60
Spin Speed										
C_1	47	5	6	2	0	47	52	58	60	60
C_2	22	7	10	16	5	22	29	39	55	60
C_3	17	4	7	17	15	17	21	28	45	60
Bake Time										
E_1	31	2	10	14	3	31	33	43	57	60
E_2	26	3	7	7	17	26	29	36	43	60
E_3	29	11	6	14	0	29	40	46	60	60
Aperture										
F_1	32	7	5	6	10	32	39	44	50	60
F_2	36	3	4	10	7	36	39	43	53	60
F_3	18	6	14	19	3	18	24	38	57	60
Exposure Time										
G_1	26	3	10	13	8	26	29	39	52	60
G_2	18	12	11	7	12	18	30	41	48	60
G_3	42	1	2	15	0	42	43	45	60	60
Developing Time										
H_1	37	4	6	8	5	37	41	47	55	60
H_2	27	11	12	5	5	27	38	50	55	60
H_3	22	1	5	22	10	22	23	28	50	60
Etch Time										
I_1	37	5	3	5	10	37	42	45	50	60
I_2	21	8	14	15	2	21	29	43	58	60
I_3	28	3	6	15	8	28	31	37	52	60
Totals	86	16	23	35	20	86	102	125	160	180

II, and III. Note that the cumulative category (V) is the same as the total number of window readings for the particular factor level.

(ii) Perform “binary data” ANOVA⁷ on each cumulative category except the last category, viz. (V). Note that a certain approximation is involved in the significance levels suggested by this ANOVA because the observations are not normally distributed.

(iii) Assign weights to each cumulative category. These weights are inversely proportional to the Bernoulli trial variance. Let cum_c be the total number of windows in the cumulative category, c , as given in the bottom row of Table XIII. Then the weight for that category is:

$$W_c = \frac{1}{\frac{cum_c}{180} \times \left(1 - \frac{cum_c}{180}\right)} = \frac{180^2}{cum_c(180 - cum_c)}$$

These weights are shown in Appendix D for each category.

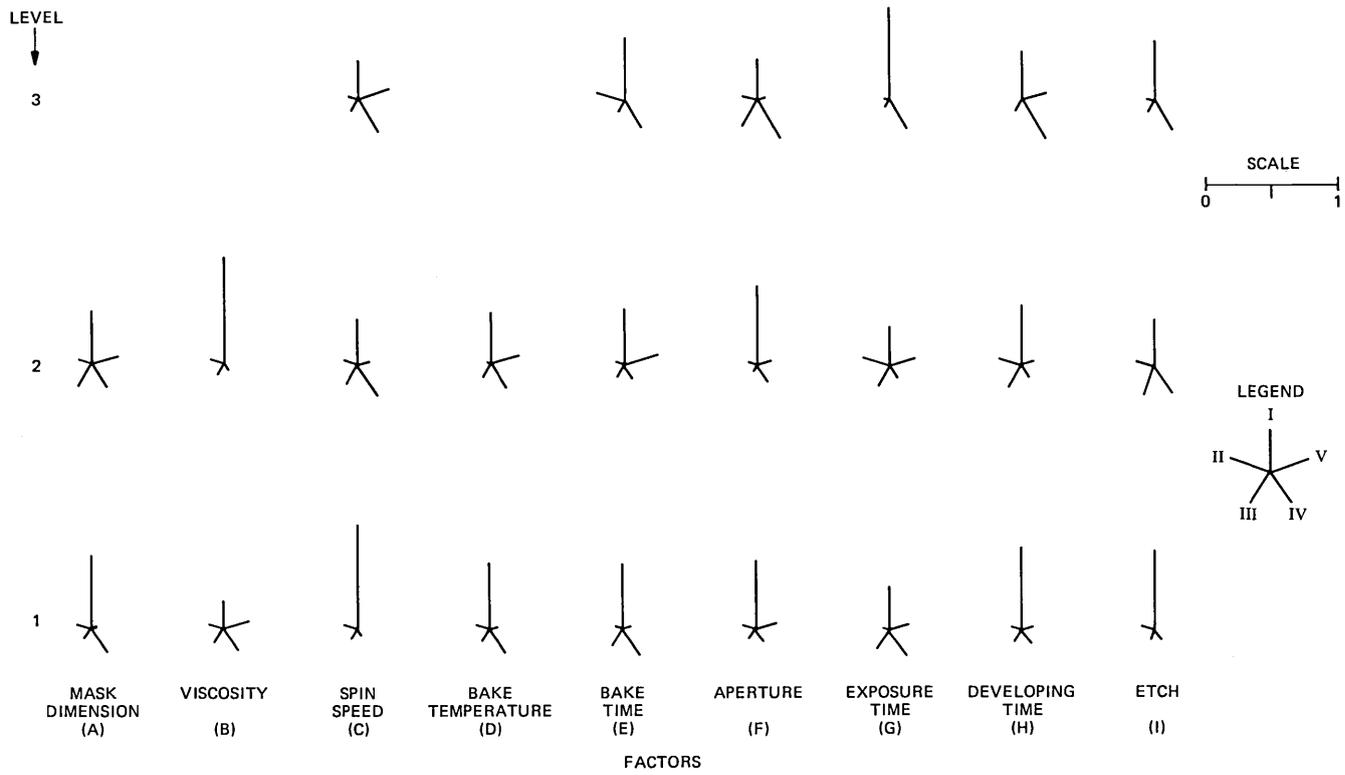


Fig. 6—Post-etch window size. The star gives the frequency distribution for each factor level.

Then for each factor and for each error term the accumulated sum of squares is taken to be equal to the weighted sum of the sum of squares for all cumulative categories.

The intuitive appeal for accumulation analysis is that by taking cumulative frequencies we preserve the order of the categories. By giving weights inversely proportional to the sampling errors in each cumulative category, we make the procedure more sensitive to a change in the variance. The difficulty is that the frequencies of the cumulative categories are correlated. So the true level of significance of the F-test may be somewhat different from that indicated by the F table. More work is needed to understand the statistical properties of the accumulation analysis.

Table XIV gives the final ANOVA with accumulated sum of squares. The computations are illustrated in Appendix D. For each cumulative category, the following nested, mixed linear model was used in performing the ANOVA:

$$y_{ijk} = \mu + x_i + e_{1ij} + e_{2ijk}, \quad (2)$$

where

- i = 1, \dots , 18 stands for the experiment.
- j = 1, 2 stands for wafer within the experiment.
- k = 1, \dots , 5 stands for replicate or position within wafer within the experiment.
- μ is the overall mean.
- x_i is the fixed effect of the factor-level combination used in experiment i . Here we consider only the main effect for each of the factors. See the discussion of model (1) in Section 7.3 for more details of the interpretation of x_i .
- e_{1ij} is the random effect for wafer j within experiment i .
- e_{2ijk} is the random error for replicate k within wafer j within experiment i .
- y_{ijk} is the observation for replicate k in wafer j in experiment i . y_{ijk} takes a value 1 if the window size belongs to the particular category. Otherwise, the value is zero.

The expected mean squares for this ANOVA model are also shown in Table XIV. The variances of e_1 and e_2 are denoted by σ_1^2 and σ_2^2 , respectively. The effects of the factors A through I are denoted by ϕ with an appropriate subscript. The effect of lack of fit is denoted by ϕ_ℓ . We assume that the random variables e_{1ij} and e_{2ijk} are independent for all values of i, j and k . The degrees of freedom shown in Table XIV have been adjusted for the fact that three experiments have only one wafer each.

For testing the significance of the effect of error between wafers within experiments, the relevant denominator sum of squares is the

Table XIV—Post-etch window size

ANOVA for accumulation analysis					
Source	Degrees of Freedom	Sum of Squares	Mean Square	F	(Expected Mean Square) ÷ \bar{W}
A Mask Dimension	4	26.64	6.66	2.67*	$\sigma_2^2 + 5\sigma_1^2 + \phi_A$
BD Viscosity-Bake Temperature	8	112.31	14.04	5.64*	$\sigma_2^2 + 5\sigma_1^2 + \phi_{BD}$
C Spin Speed	8	125.52	15.69	6.30*	$\sigma_2^2 + 5\sigma_1^2 + \phi_C$
E Bake Time	8	36.96	4.62	1.86	$\sigma_2^2 + 5\sigma_1^2 + \phi_E$
F Aperture	8	27.88	3.49	1.40	$\sigma_2^2 + 5\sigma_1^2 + \phi_F$
G Exposure Time	8	42.28	5.29	2.12*	$\sigma_2^2 + 5\sigma_1^2 + \phi_G$
H Developing Time	8	45.57	5.70	2.29*	$\sigma_2^2 + 5\sigma_1^2 + \phi_H$
I Etch Time	8	23.80	2.98	1.20	$\sigma_2^2 + 5\sigma_1^2 + \phi_I$
Lack of Fit	8	17.25	2.16	0.87	$\sigma_2^2 + 5\sigma_1^2 + \phi_I$
Error Between Wafers Within Experiment	60	149.33	2.49	11.69*	$\sigma_2^2 + 5\sigma_1^2$
Error Between Replicates Within Wafers Within Experiment	528	112.45	0.21		σ_2^2
Total	656	720.00			

$$\bar{W} = (W_{(I)} + W_{(II)} + W_{(III)} + W_{(IV)})/4.$$

$$F_{4,60}(0.95) = 2.53, F_{8,60}(0.95) = 2.10, F_{60,528}(0.95) = 1.32.$$

(b) Separation of S_{BD}

Source	Degrees of Freedom	Sum of Squares	Mean Square	F
B Viscosity	4	87.38	21.85	8.78*
D Bake Temperature	4	6.55	1.64	0.66

* Factors significant at 95-percent confidence level.

estimate of σ_2^2 . The corresponding F value is 11.69, which is significant far beyond the nominal 5-percent level. To test for the lack of fit of the main-effects-only model, the appropriate denominator is the estimate $\sigma_2^2 + 5\sigma_1^2$. The corresponding F ratio is 0.87. This indicates that the main-effects-only model adequately describes the observed data relative to the random errors between wafers. For testing the significance of the process factors, the denominator mean square is again the estimate of $\sigma_2^2 + 5\sigma_1^2$. We see that the mask dimension, viscosity, spin speed, exposure time, and developing time have a significant effect (approximately 5-percent level) on the window size. The effects of the other factors are not significant.

IX. SELECTION OF OPTIMUM FACTOR LEVELS

The following table summarizes the significant results of the analyses performed in Sections VII and VIII. In each category, the factors are arranged in descending order according to the F value.

Significant effect on s/n:

Pre-etch line width: A, F

Post-etch line width: None

Significant effect on mean:

Pre-etch line width: A, B, C

Post-etch line width: B, G, C, A, H

Significant factors identified by accumulation analysis:

Post-etch window size: B, C, A, H, G

Factors that have a significant effect on the s/n and the factors identified to be significant by the accumulation analysis are all control factors. Setting their levels equal to optimum levels minimizes the process variability. Here the control factors are A, F, B, C, H, and G.

To keep the process mean on target we use a signal factor. Ideally, the signal factor should have a significant effect on the mean, but should have no effect on the s/n. Then changing the level of the signal factor would affect only the mean. In practice, a small effect on the s/n may have to be tolerated.

Among the factors (A, B, C, G, and H) that have a significant effect on the mean, factors A, B, and C are relatively strong control factors as measured by the F statistics for the accumulation analysis and the ANOVA for pre-etch line-width s/n. Also, these factors are relatively difficult to change during production. So A, B, and C are not suitable as signal factor. Between the remaining two factors, G and H, G has greater effect on the mean and also shows as a less significant factor in accumulation analysis. So exposure time was assigned to be the signal factor.

The optimum levels for the control factors were selected as follows. The mask dimension (A) and the aperture (F) have a significant effect on the s/n for pre-etch line width. From Table V we see that the 2.5- μm mask (level 2) has a higher s/n than the 2.0- μm mask. Hence 2.5 μm was chosen to be the optimum mask dimension. Also, aperture 1 (level 1) has the highest s/n among the three apertures studied. However, because of the past experience, aperture 2 was chosen to be the preferred level.

The accumulation analysis of the post-etch window-size data indicated that the viscosity, spin speed, mask dimension, developing time, and exposure have statistically significant effects on the frequency distribution. The optimum levels of these factors can be determined from Table XIII and Fig. 6 to be those that have the smallest fraction of windows not open (category I) and the largest fraction of windows in the range $3.0 \pm 0.25 \mu\text{m}$ (category IV). Because it is more critical to have all the windows open, when there was a conflict we took the requirement on category I to be the dominant requirement. The optimum levels are: 2.5- μm mask dimension, viscosity 204, 4000-rpm spin speed, 60-second developing time, and normal exposure.

Table XV—Optimum factor levels

Label	Factors Name	Standard Levels	Optimum Levels
A	Mask Dimension (μm)	2.0	2.5
B	Viscosity	204	204
C	Spin Speed (rpm)	3000	4000
D	Bake Temperature ($^{\circ}\text{C}$)	105	105
E	Bake Time (min)	30	30
F	Aperture	2	2
G	Exposure (PEP setting)	Normal	Normal
H	Developing Time (s)	45	60
I	Plasma Etch Time (min)	13.2	13.2

Table XV shows side by side the optimum factor levels and the standard levels as of September 1980. Note that our experiment has indicated that the mask dimension be changed from 2.0 μm to 2.5 μm , spin speed from 3000 rpm to 4000 rpm, and developing time from 45 seconds to 60 seconds. The exposure time is to be adjusted to get the correct mean value of the line width and the window size. The levels of the other factors, which remain unchanged, have been confirmed to be optimum to start with.

In deriving the optimum conditions we have conducted a highly fractionated factorial experiment and have considered only the main effects of the factors. The interactions between the factors have been ignored. If the interactions are strong compared to the main effects, then there is a possibility that the optimum conditions thus derived would not improve the process. So experiments have to be conducted to verify the optimum conditions. The verification was done in conjunction with the implementation, which is described next.

X. IMPLEMENTATION AND THE BENEFITS OF THE OPTIMUM LEVELS

We started to use the optimum process conditions given in Table XV in the Integrated Circuits Design Capability Laboratory in January 1981. In the beginning the exposure was set at 90, which is the normal setting given in Table I. We observed that the final window at the end of step (xi) was much larger than the target size of 3.5 μm . Through successive experiments, we reduced the exposure time until the mean final window size came to about 3.5 μm . The corresponding exposure setting is 140. Since then the process has been run at these conditions. The benefits of running the process at these conditions are:

(i) The pre-etch line width is routinely used as a process quality indicator. Before September 1980 the standard deviation of this indicator was 0.29 μm on a base line chip (DSO chip). With the optimum process parameters, that standard deviation has come down to 0.14 μm . This is a two-fold reduction in standard deviation, or a four-fold reduction in variance. This is evidenced by Fig. 7, which shows a

typical photograph of the PLA area of a *BELLMAC-32* microprocessor chip fabricated by the new process. Note that windows in Fig. 7 are much more uniform in size compared to the windows in Fig. 2. Also, all windows are printed and open in Fig. 7.

(ii) After the final step of window forming, i.e., after step (xi), the windows are visually examined on a routine basis. Analysis of the quality control data on the DSO chip, which has an area of approximately 0.19 cm^2 , showed that prior to September 1980 about 0.12 window per chip was either not open or not printed (i.e., approximately one incidence of window not open or not printed was found in eight chips). With the new process only 0.04 window per chip is not open or printed (i.e., approximately one incidence of window not open or printed is found in twenty-five chips). This is a three-fold reduction in defect density due to unopened windows.

(iii) Observing these improvements over several weeks, the process engineers gained a confidence in the stability and robustness of the new process parameters. So they eliminated a number of in-process checks. As a result the overall time spent by the wafers in window photolithography has been reduced by a factor of two.

The optimum parameter levels were first used in the Integrated Circuit Device Capability Laboratory with only a few codes of ICs. Subsequently, these parameter levels were used with all codes of $3.5\text{-}\mu\text{m}$ technology chips, including *BELLMAC-4* microcomputer and *BELLMAC-32* microprocessor chips. The mask dimension change from 2.0 to $2.5 \mu\text{m}$ is now a standard for $3.5\text{-}\mu\text{m}$ CMOS technology.

XI. DISCUSSION AND FUTURE WORK

The off-line quality control method is an efficient method of improving the quality and the yield of a production process. The method has a great deal of similarity with the response surface method⁹ and the evolutionary operations method,¹⁰ which are commonly known in statistical literature in this country. Both the response surface and the evolutionary operations methods are used to maximize the yield of a production process and they both make use of the experimental design techniques. The main difference is that in the off-line quality control method the process variability that has a great impact on the product quality is the objective function. In the response surface and evolutionary operations methods, the process variability is generally not considered. Thus, intuitively, the optimum levels derived by using the off-line quality control method can be expected to be more robust, stable, and dependable.

In the response surface method one typically uses a relatively large fraction of the factorial experiment. However, in off-line quality control usually a very small fraction is chosen. Another difference is that in

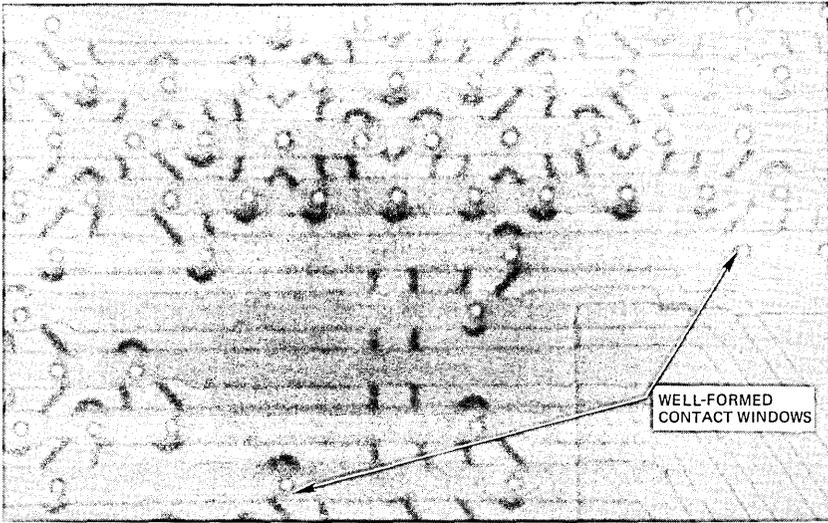


Fig. 7—Typical results for contact windows in PLA area (January through March, 1981) using parameters derived from this experiment. The contact windows are round shaped.

the response surface method the objective function is considered to be a continuous function approximated by a low-order polynomial. In off-line quality control, we can simultaneously study both the continuous and discrete factors.

Our application of the off-line quality control method to the window-cutting process in the Murray Hill 3.5- μm CMOS technology, as seen from the earlier sections, has resulted in improved control of window size, lower incidence of unopened windows, and reduced time for window photolithography. Presently, we have undertaken to optimize two more steps in IC fabrication. Those steps are polysilicon patterning and aluminum patterning. Both these processes, like the window-cutting process, involve photolithography and are among the more critical processes of IC fabrication. We think that the method has a great potential and would like to see applications in various parts of Bell Laboratories and Western Electric Company.

XII. ACKNOWLEDGMENTS

The authors acknowledge with gratitude the support and the encouragement given by R. L. Chadha, R. Edwards, J. V. Dalton, A. B. Hoadley, and G. R. Weber. R. G. Brandes, K. J. Orlowski, and members of the IC Process Operations Group provided valuable help in conducting the experiments. We also thank Ramon Leon for helpful discussions on the graphical representation of the data and I. J. Terpenning for helpful comments on an earlier draft of this paper. The

original manuscript was carefully prepared by Jill Cooper on the UNIX* operating system.

REFERENCES

1. G. Taguchi, "Off-line and On-line Quality Control Systems," International Conference on Quality Control, Tokyo, Japan, 1978, pp. B4-1-5.
2. R. N. Kacker and M. S. Phadke, unpublished work.
3. G. Taguchi and Y. Wu, *Introduction to Off-Line Quality Control*, Central Japan Quality Control Association, Nagoya, Japan, 1980.
4. A. B. Glaser and G. A. Subak-Sharp, *Integrated Circuits Engineering*, Reading, MA: Addison-Wesley, 1977.
5. W. Q. Meeker, G. H. Hahn, and P. I. Feder, "A Computer Program for Evaluating and Comparing Experimental Designs and Some Applications," *American Statistician*, 29, No. 1 (February 1975), pp. 60-4.
6. C. R. Hicks, *Fundamental Concepts in the Design of Experiments*, New York, NY: Holt, Rinehart and Winston, 1973.
7. S. R. Searle, *Linear Models*, New York, NY: John Wiley, 1971.
8. G. Taguchi, "A New Statistical Analysis Method for Clinical Data, the Accumulation Analysis, in Contrast with the Chi-square Test," Shinjuku Shobo, Tokyo, Japan, 1975.
9. R. H. Myers, *Response Surface Methodology*, Newton, MA: Allyn and Bacon, 1971.
10. G. E. P. Box and N. R. Draper, *Evolutionary Operations: A Statistical Method for Process Improvement*, New York, NY: John Wiley, 1969.
11. G. W. Snedecor and W. G. Cochran, *Statistical Methods*, Ames, Iowa: Iowa State University Press, 1980.

APPENDIX A

Computation of the Sum of Squares—Analysis of the Signal-to-Noise Ratio for Pre-Etch

The computations of the sum of squares tabulated in Table VIa are illustrated below.

S_m = Correction Factor

$$= \frac{\left(\sum_{i=1}^{18} \eta_i \right)^2}{18} = \frac{(25.2202)^2}{18} = 35.3366$$

S_A = Sum of squares for factor A

$$= \frac{(9m_{A_1})^2 + (9m_{A_2})^2}{9} - S_m$$

$$= \frac{(11.5711)^2 + (13.6491)^2}{9} - 35.3366$$

$$= 0.2399 \quad (d.f. = 1)$$

$$S_C = \frac{(6m_{C_1})^2 + (6m_{C_2})^2 + (6m_{C_3})^2}{6} - S_m$$

* Trademark of Bell Laboratories.

$$= \frac{(8.1979)^2 + (8.1017)^2 + (8.9206)^2}{6} - S_m$$

$$= 0.0668 \quad (d.f. = 2).$$

Sums of squares for the factors E, F, G, and H were calculated similarly. The combined sum of squares due to B and D is given by

$$S_{BD} = \text{Sum of squares for the column BD}$$

$$= \frac{(6m_{B_1D_1})^2 + (6m_{B_1D_2})^2 + (6m_{B_2D_1})^2}{6} - S_m$$

$$= \frac{(8.2524)^2 + (8.6649)^2 + (8.3029)^2}{6} - 35.3366$$

$$= 0.0169 \quad (d.f. = 2).$$

The total sum of squares is

$$S_T = \sum_{i=1}^{18} \eta_i^2 - S_m = 0.8423. \quad (d.f. = 17).$$

The error sum of squares is calculated by subtraction.

$$S_e = S_T - (S_A + S_{BD} + S_C + S_E + S_F + S_G + S_H)$$

$$= 0.1522 \quad (d.f. = 4).$$

Here we do not compute the sum of squares due to factor I (etch time), because it has no influence on the pre-etch line width.

APPENDIX B

Computation of the Percent Contribution—Analysis of the Signal-to-Noise Ratio for Pre-Etch Line Width

The computation of the percent contribution is explained below. The contribution of factor A to the total sum of squares

$$= S_A - (d.f. \text{ of A})(\text{error mean square}).$$

Hence, the percent contribution for factor A

$$= \frac{S_A - (d.f. \text{ of A})(\text{error mean square})}{\text{total sum of squares}} \times 100$$

$$= \frac{0.2399 - 0.0251}{0.8423} \times 100 = 25.5\%.$$

The percent contributions of E and F are determined similarly. Now consider, the contribution of error to the total sum of squares:

$$= S_e + (\text{total } d.f. \text{ for factors})(\text{error mean square}).$$

Hence, the percent contribution for error

$$\begin{aligned}
 &= \frac{S_e + (\text{total } d.f. \text{ for factors})(\text{error mean square})}{\text{total sum of squares}} \times 100 \\
 &= \frac{0.3010 + 5 \times 0.0251}{0.8423} \times 100 = 50.6\%.
 \end{aligned}$$

APPENDIX C

Separation of S_{BD} into S_B and S_D —Analysis of the Mean Pre-Etch Line Width

The sum of squares, S_{BD} , can be decomposed in the following two ways¹¹ to obtain the contributions of the factors B and D:

$$S_{BD} = S'_{B(D)} + S'_D$$

and

$$S_{BD} = S'_{D(B)} + S'_B.$$

Here $S'_{B(D)}$ is the sum of squares due to B, assuming D has no effect; S'_D is the sum of squares due to D after eliminating the effect of B. The terms $S'_{D(B)}$ and S'_B are interpreted similarly. We have

$$S'_{B(D)} = \frac{(6m_{B_1D_1} + 6m_{B_1D_2} - 12m_{B_2D_1})^2}{(1^2 + 1^2 + 2^2) \times 6} = 0.903 \quad (d.f. = 1)$$

$$S'_D = S_{BD} - S'_{B(D)} = 0.047 \quad (d.f. = 1).$$

Similarly,

$$S'_{D(B)} = \frac{(6m_{B_1D_1} + 6m_{B_2D_1} - 12m_{B_1D_2})^2}{(1^2 + 1^2 + 2^2) \times 6} = 0.116 \quad (d.f. = 1)$$

$$S'_B = S_{BD} - S'_{D(B)} = 0.834 \quad (d.f. = 1).$$

For testing the significance of the factors B and D we use S'_B and S'_D , respectively. Note that S'_B and S'_D do not add up to S_{BD} , which is to be expected because the design is not orthogonal with respect to the factors B and D.

APPENDIX D

Computation of the Sum of Squares for Accumulation Analysis—Analysis of Post-Etch Window Size

The weights for the cumulative categories (I), (II), (III), and (IV) are given below. The frequencies of the bottom line of Table XIII are used in computing these weights. Therefore:

$$W_{(I)} = \frac{1}{\frac{86}{180} \times \frac{180 - 86}{180}} = 4.008$$

$$W_{(II)} = \frac{180^2}{102 \times (180 - 102)} = 4.072$$

$$W_{(III)} = \frac{180^2}{125 \times (180 - 125)} = 4.713$$

$$W_{(IV)} = \frac{180^2}{160 \times (180 - 160)} = 10.125.$$

Computation of the sum of squares tabulated in Table XIV are illustrated below:

$$\begin{aligned} S_A &= W_{(I)} \times \left(\frac{51^2 + 35^2}{90} - \frac{86^2}{180} \right) + W_{(II)} \times \left(\frac{58^2 + 44^2}{90} - \frac{102^2}{180} \right) \\ &\quad + W_{(III)} \times \left(\frac{67^2 + 58^2}{90} - \frac{125^2}{180} \right) + W_{(IV)} \times \left(\frac{88^2 + 72^2}{90} - \frac{160^2}{180} \right) \\ &= 26.64, \end{aligned}$$

and

$$\begin{aligned} S_C &= W_{(I)} \times \left(\frac{47^2 + 22^2 + 17^2}{60} - \frac{86^2}{180} \right) \\ &\quad + W_{(II)} \times \left(\frac{52^2 + 29^2 + 21^2}{60} - \frac{102^2}{180} \right) \\ &\quad + W_{(III)} \times \left(\frac{58^2 + 39^2 + 28^2}{60} - \frac{125^2}{180} \right) \\ &\quad + W_{(IV)} \times \left(\frac{60^2 + 55^2 + 45^2}{60} - \frac{160^2}{180} \right) \\ &= 125.52. \end{aligned}$$

The Effects of Selected Signal Processing Techniques on the Performance of a Filter-Bank-Based Isolated Word Recognizer

By B. A. DAUTRICH, L. R. RABINER, and T. B. MARTIN

(Manuscript received December 3, 1982)

To implement an isolated word recognizer based on filter bank techniques, decisions must be made as to how to condition the speech signal prior to the filter bank analysis (preprocessing), how to condition the feature vector at the output of the filter bank analysis (postprocessing), and how to perform the time alignment in the pattern comparison between an unknown test pattern and previously stored reference patterns (registration and distance computation). In the past most designers of such word recognition systems made arbitrary choices about how the various signal processing operations were to be carried out. This paper presents results of a systematic study of the effects of selected signal processing techniques on the performance of a filter bank isolated word recognizer using telephone-quality speech. In particular, the filter bank analyzer was a 13-channel, critical-band-spaced filter bank with excellent time resolution (impulse response durations of from 3 to 30 ms) and poor frequency selectivity (highly overlapping filters with ratios of center frequency to 3-dB bandwidth of about 8 for each band). Among the signal processing techniques studied were: preemphasis of the speech signal; time and frequency smoothing of the filter bank outputs; thresholding, quantization, and normalization of the feature vector; principal components analysis of the feature vector; local and global distance computations for use in the time alignment procedure; and noise analysis in both training and testing. Each of the signal processing techniques was studied individually; hence no tests were run in which several of the techniques were used together. Results showed that some fairly simple signal processing operations provided the best overall performance in the noise-free case; in noisy conditions performance degraded significantly for signal-to-noise ratios less than about 24 dB.

I. INTRODUCTION

To implement an isolated word recognizer based on a filter analysis, decisions must be made as to how to preprocess the speech signal prior to the filter bank analysis, how to postprocess the feature vectors obtained at the output of the filter bank analysis, and how to perform the time alignment and distance computation in the pattern comparison between an unknown test pattern and previously stored reference patterns. Often such decisions are made arbitrarily based on experience, heuristic procedures, or sometimes a few brief tests with the system. To our knowledge no one has attempted to systematically examine the effects of various signal processing techniques on the performance (as measured in word error rate) of a filter-bank-isolated word recognizer. This paper provides such a comparison by examining several of the most popular signal processing techniques and showing how they affect the performance of a particular filter bank word recognizer using telephone-quality speech.¹

There are two inherent problems with any study that attempts to find the best signal processing techniques for a system via experimental means. The first is that the results presented are highly dependent on the signal processing techniques that were studied. Hence, the "optimal" way of processing the signal may not even have been investigated (due to lack of knowledge, etc). With our limited knowledge we know of no way to avoid this difficulty. The second problem is that, of necessity, each of the various signal processing techniques is studied independently of any other (thereby tacitly assuming independence of the various methods). Hence, any interactions between the techniques studied will go unnoticed. Again, we know of no practical way of studying the interactions between signal processing operations; the processing, assuming independence of operations, took about four full months on a modern minicomputer system!

The results to be presented in this paper are an extension of a previous study¹ that examined different filter bank structures and compared their performance to that of a conventional linear predictive coefficient (LPC) word recognizer.^{2,3} The key results of this earlier work were:

(i) The best performance in word recognition tests was achieved by both a 13-channel, critical-band-spaced filter bank, and a 15-channel, uniformly spaced filter bank. Both filter banks had composite frequency responses without gaps at the band edges. The 13-channel filter bank had highly overlapping filters; the 15-channel filter bank had filters with almost no overlap.

(ii) There were significant performance differences between talkers (especially female as opposed to male talkers).

(iii) Performance of the LPC and the best filter bank recognizers were comparable for a simple vocabulary of the 10 digits using telephone-quality speech (with no extra noise degradation) over a dialed-up telephone line using a local private branch exchange (PBX).

(iv) Performance of the LPC recognizer was superior to that of the best filter-bank recognizers for a complex vocabulary of the alphabet, digits, and three command words, again using telephone-quality speech.

A key question arising from these results was whether any of the proposed signal-processing techniques for the filter bank system could bring up the performance to that of the LPC system for the complex alpha-digits vocabulary. Unfortunately, we will see that none of the proposed methods was able to significantly improve filter-bank performance. (However, some were able to keep performance the same while reducing required storage.)

Two other implementational aspects of word recognizers were studied. The first involves the use of the normalize-and-warp procedure proposed by Myers et al.⁴ In this procedure a fixed-length pattern is created for both test and reference patterns prior to time alignment. In this manner the largest warping area is obtained, and the computational aspects of implementing the time-warping procedure are greatly simplified. Instead of considering just the word average length for warping, we studied the effects (for both the filter bank and LPC systems) of warping to prespecified lengths of various amounts. It was found that a large amount of compression could be made before system performance degraded by a significant amount.

The second implementational aspect studied was the effects of additive noise on the performance of both the LPC and filter bank recognizers. We considered cases in which both the training and testing occurred in the noisy background, and when only the testing occurred in the noisy background. It was found that far superior performance was obtained, at all signal-to-noise ratios, when both training and testing occurred in the noisy background. Furthermore, performance of both types of recognizers degraded for signal-to-noise ratios less than or equal to 24 dB. Also, for signal-to-noise ratios greater than 6 dB, the LPC recognizer outperformed the filter bank recognizer.

An overview of the work presented in this paper is as follows. In Section II we review the general implementation of the filter bank isolated word recognizer. In Section III we discuss the signal processing methods that were studied in conjunction with the filter bank. In Section IV we discuss the noise studies. In Section V we describe the experiments performed and give word error rates for the various tests. Finally, a discussion of the results is given in Section VI.

II. THE FILTER-BANK-ISOLATED WORD RECOGNIZER

Figure 1 shows a block diagram of the overall filter bank word recognizer. The input speech signal is recorded off a dialed-up telephone line, band-limited to 3200 Hz, and digitized at a 6.67-kHz rate. The digitized speech signal, $s(n)$, is first sent to a preprocessor to condition the signal for the filter bank analyzer. Preprocessing is basically a spectra-shaping operation (e.g., linear filtering) for increased immunity to finite word-length processing in the remainder of the system.⁵ The preprocessed signal, $\hat{s}(n)$, is then sent to a filter bank analyzer whose structure is shown in Fig. 2. The filter bank contains a set of Q parallel bandpass filters that cover the speech band of

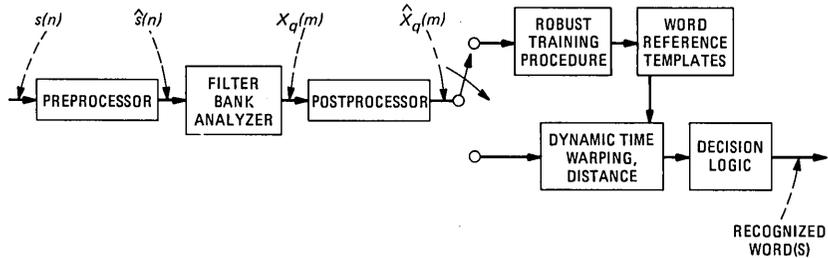


Fig. 1—Filter bank word recognizer with both preprocessing and postprocessing operations.

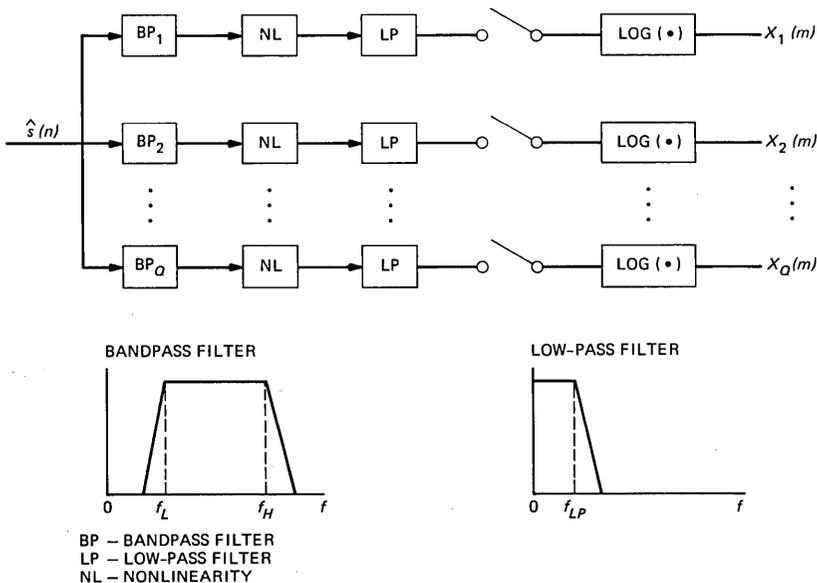


Fig. 2—Structure of filter bank analyzer.

interest (100 to 3200 Hz for telephone speech). Each bandpass filter is followed by a nonlinearity (NL), a low-pass (LP) filter, a sampler, and a logarithmic compressor. The output of the filter bank at time m is a vector

$$X(m) = [X_1(m), X_2(m), \dots, X_Q(m)], \quad (1)$$

whose components $X_i(m)$ represent the energy in the speech signal in channel i at time m .

In our previous work¹ we studied the effects of different types of filter banks on recognizer performance and found that the highest accuracy was obtained with two types of filter banks, namely:

(i) A 13-channel, critical-band-spaced filter bank with higher overlapping channels. This filter bank had excellent time resolution (on the order of 10 ms) but poor frequency resolution.

(ii) A 15-channel, uniformly spaced filter bank with essentially no overlap between channels. This filter had poor time resolution but excellent frequency resolution. The composite spectrum of this filter bank was flat to within fractions of a dB.

Both of these filter banks used a magnitude nonlinearity and a 3-pole, Bessel, low-pass filter with a 30-Hz cutoff frequency. The sampling rate of the output feature vector was 67 Hz—i.e., adjacent feature estimates were spaced 15 ms in time.

The output of the filter bank was sent to a postprocessor, which performed one or more of the following operations:

- (i) Time smoothing of feature vectors
- (ii) Frequency smoothing of channels within a feature vector
- (iii) Normalization of the feature vector
- (iv) Thresholding and/or quantization of the feature vector
- (v) Principal components analysis of the feature vector.

The output of the postprocessor was the input pattern to either the training mode (a robust training procedure⁶), or to the testing mode. In the training mode a set of word reference templates were created based on consistent matches of tokens of a word to previously analyzed tokens of the same word. In the testing mode the test pattern, T , consisting of the sequence of feature vectors

$$T = \{\hat{X}(1), \hat{X}(2), \dots, \hat{X}(M)\} \quad (2)$$

was compared to the reference pattern, R^i , for the i th vocabulary word using a dynamic time-warping alignment procedure.² For the i th reference pattern, a total average distance, D_i , between it and the test pattern was computed, and simple decision logic was used to make the word choice for recognition.

To implement the word recognizer of Fig. 1, one has to choose the types of processing to go into the preprocessor, the filter bank analyzer,

the postprocessor, and the dynamic time-warping algorithm. Based on the earlier study¹ we limited the filter bank to the 13-channel critical-band filter bank. However, for each of the remaining signal-processing blocks we tried to choose one or more possibilities and then did experiments to evaluate its usefulness to the overall word recognizer. We discuss our choices in detail in Section III. In addition we chose to study the effects of both noise addition and length quantization of both reference and test patterns on overall performance. These experiments are described in Section IV.

III. SIGNAL PROCESSING CHOICES IN THE RECOGNIZER

3.1 Preprocessor

The function of the preprocessor is to spectrally shape the speech signal to achieve some desired gross spectral shape. The most common form of preprocessing is simple preemphasis, which is used to compensate the inherent 6-dB per octave falloff in the speech spectrum. In such a case a simple first-order network of the form

$$H(z) = 1 - \alpha z^{-1} \quad (3)$$

has been found adequate³ for recognition purposes. Thus, the difference equation relating $\hat{s}(n)$ to $s(n)$ is of the form

$$\hat{s}(n) = s(n) - \alpha s(n - 1), \quad (4)$$

where a value of $\alpha = 0.95$ has been used previously.

3.2 Postprocessing

We denote the output of the q th channel of the filter bank at frame m as $X_q(m)$, $m = 1, 2, \dots, M$, $q = 1, 2, \dots, Q$. All of the postprocessing operations can be expressed in terms of signal processing on $X_q(m)$ to give the signal $\hat{X}_q(m)$. We have considered the operations in Sections 3.2.1 through 3.2.5.

3.2.1 Thresholding and energy normalization

The purpose of channel thresholding is to clamp low-level noise signals from channels at times when essentially no speech signal is present. This is done by applying a threshold so that channel signals below threshold are clamped at the threshold value. In this way much less sensitivity to background noise is achieved. In particular, this is achieved by determining X_q^{MAX} , the peak signal level for the q th channel, for each word as:

$$X_q^{\text{MAX}} = \max_{1 \leq m \leq M} [X_q(m)]. \quad (5)$$

Then the threshold for the q th channel is set at

$$T_q^* = X_q^{\text{MAX}} - T^*, \quad (6)$$

where T^* is a parameter of the recognition system. A typical range of T^* is from 30 to 50 (dB).¹ The thresholded channel signal is then given as

$$\hat{X}_q(m) = \max[X_q(m), T_q^*] \quad (7)$$

for all q and m .

The purpose of frame energy normalization is to compensate for variations in speech level from utterance to utterance. We have considered two distinct normalization methods, which we call average and peak normalization. For average normalization we calculate the frame average, $\bar{X}(m)$, as

$$\bar{X}(m) = \frac{1}{Q} \sum_{q=1}^Q X_q(m), \quad (8)$$

and for peak normalization we calculate the peak as

$$\bar{X}(m) = \max_{1 \leq q \leq Q} [X_q(m)]. \quad (9)$$

The energy-normalized feature vector is then given as[†]

$$\hat{X}_q(m) = X_q(m) - \bar{X}(m). \quad (10)$$

It can readily be shown that both peak and average normalization have the property that if a feature set T is derived from the speech signal $s(n)$, then the feature set T' derived from

$$s'(n) = \gamma s(n) \quad (11)$$

will be identical to T after the normalization of eq. (10) is carried out. Hence, gain variations are normalized out of the processing as desired.

3.2.2 Time smoothing of feature vectors

The purpose of time smoothing of feature vectors is to reduce the variability in channel outputs by averaging adjacent time frames. The cost of such smoothing is a decrease in time resolution achieved by the recognizer. If we assume that M^* adjacent frames are to be overlapped and smoothed, then time smoothing can be expressed as

$$\hat{X}_q(m) = \frac{1}{M^*} \sum_{\hat{m}=m-M^*+1}^m X_q(\hat{m}), \quad m = M^*, \dots, M. \quad (12)$$

[†] The reader is reminded that the channel signals are logarithmically encoded. Hence, normalization takes the form of subtraction.

It should be noted that the first $(M^* - 1)$ frames are eliminated and are used as initial conditions for the smoothing.

3.2.3 Frequency smoothing of channel outputs

As in time smoothing, the purpose of frequency smoothing is to reduce the variability in channel outputs by averaging adjacent channels for a given time frame. Again, the cost of this smoothing is a loss in frequency resolution. If we assume that Q^* adjacent channels are to be overlapped and smoothed, then frequency smoothing can be expressed as

$$\hat{X}_q(m) = \frac{1}{Q^*} \sum_{\hat{q}=q-Q^*+1}^q X_{\hat{q}}(m), \quad q = Q^*, \dots, Q. \quad (13)$$

It should again be noted that the first $(Q^* - 1)$ channels are eliminated and are used as initial conditions for the smoothing.

3.2.4 Quantization of channel outputs

The purpose of quantizing the channel outputs is to reduce the storage requirements of the recognizer both for reference patterns and for the test pattern. If we use a B -bit quantizer, and we assume the channel signals are in the range $[0, -T^*]$ because of the thresholding and energy normalization operations, then with a uniform quantizer we have a quantization width of

$$\Delta E = \frac{T^*}{2^B} \quad (14)$$

and we can express the quantized output signal as

$$\hat{X}_q(m) = \lfloor X_q(m) / \Delta E \rfloor \cdot \Delta E, \quad (15)$$

where $\lfloor x \rfloor$ is the greatest integer less than x , and it is assumed that $\hat{X}_q(m)$ is already thresholded and energy normalized.

3.2.5 Principal components analysis

The last form of postprocessing that we considered was a principal components analysis⁷ in which the Q -dimensional filter bank feature vector is transformed into a new P -dimensional feature vector, where $P < Q$, such that all the important information in the original vector is retained. The purpose of reducing the feature dimensionality is to reduce the required storage for reference and test patterns. This method has also been used by Pols⁷ to recognize vowels with good success.

The way in which the principal components analysis was performed was as follows. The first step is to collect a large number of filter bank

feature vectors and to compute the covariance matrix, Λ , between dimensions as

$$\Lambda_{ij} = \frac{\sum [X_i(\cdot) - \bar{X}_i(\cdot)][X_j(\cdot) - \bar{X}_j(\cdot)]}{\left\{ \sum [X_i(\cdot) - \bar{X}_i(\cdot)]^2 \sum [X_j(\cdot) - \bar{X}_j(\cdot)]^2 \right\}^{1/2}}, \quad (16)$$

where the summation is over the training set of feature vectors. The principal components analysis then determines a new dimension, which is a linear combination of the original Q dimensions, that contains as much of the total variance as possible. Then a second new dimension is determined such that it is orthogonal to the first new dimension and contains as much of the remaining variance as possible. This new dimension is again a linear combination of the Q old dimensions. This process is continued until we have P new orthogonal dimensions, all of which are linear combinations of the original Q dimensions. Hence, if we denote the transformed set as $\hat{X}_{\hat{q}}(m)$, the transformation to the new dimensions is of the form

$$\hat{X}_{\hat{q}}(m) = \sum_{q=1}^Q \beta_{\hat{q}}(q) X_q(m), \quad \hat{q} = 1, 2, \dots, P, \quad (17)$$

where $\beta_{\hat{q}}(q)$ is the coefficient vector for dimension \hat{q} .

The resulting P -dimensional space of the principal components analysis contains as much of the total variance of the original space as is possible in P dimensions. The new space is obtained formally by doing an eigenvector analysis of the original covariance matrix, Λ . The resulting eigenvectors are the coefficient vectors for the transformation of eq. (17).

3.3 Dynamic time-warping considerations

Once the feature vectors have been obtained, the recognizer must compare the unknown test pattern, T , to each word reference pattern, R_i , $i = 1, 2, \dots, V$, for a V -word vocabulary. For this comparison the technique of dynamic time warping (DTW) is used.^{2,8,9} If we denote the test pattern, T , as

$$T = \{T(1), T(2), \dots, T(M)\} \quad (18)$$

and the i th reference, R_i , as

$$R^i = \{R^i(1), R^i(2), \dots, R^i(N_i)\}, \quad (19)$$

then the DTW algorithm determines a warping path

$$n = w(m) \quad (20)$$

such that the total distance, $D(T, R^i)$, defined as

$$D(T, R^i) = \frac{1}{M} \sum_{m=1}^M d\{T(m), R^i[w(m)]\} \quad (21)$$

is minimized, where $d(T, R^i)$ is the local distance between test and reference frames.

We have considered several variations on the conventional DTW algorithm. First we have modified the global distance of eq. (21) to include a time weighting of the form

$$\hat{D}(T, R^i) = \frac{\sum_{m=1}^M W^T(m) d\{T(m), R^i[w(m)]\}}{\sum_{m=1}^M W^T(m)}, \quad (22)$$

where $W^T(m)$ is the weight applied to the local distance at frame m . It should be noted that in eq. (22) the weight is a function of only the test pattern, T .

We have also considered a variety of types of local distance calculations of the form

$$d(T, R) = \frac{\left\{ \sum_{q=1}^Q (W_q^T)^p [|T(q) - R(q)|^p] \right\}^{\frac{1}{p}}}{\left[\sum_{q=1}^Q (W_q^T)^p \right]^{\frac{1}{p}}}, \quad (23)$$

where W_q^T is a frequency weighting curve dependent only on the test pattern, T , and p is the distance power for emphasizing the frequency variations. Typical values for p are 1 (magnitude distance), 2 (squared distance), and 1/2 (square root distance). Again, it should be noted that the frequency weight of eq. (22) is only a function of the test frame.

An alternative form of distance weighting was suggested by Silverman and Dixon¹⁰ and is of the form

$$d(T, R) = \frac{1}{Q} \sum_{q=1}^Q |T(q) - R(q) - f(|\bar{T} - \bar{R}|)|, \quad (24)$$

where

$$f(y) = 1 - \frac{y}{y^{\text{MAX}}} \quad (25)$$

and y^{MAX} is the largest value that y can attain. The form of eq. (24) is similar to that of the average normalization [eq. (8)] discussed earlier in that the means of T and R (over channels) are essentially subtracted

- from each $T(q)$ and $R(q)$ component. However, this is only the case when $|\bar{T} - \bar{R}| \approx 0$, in which case $f(|\bar{T} - \bar{R}|) \approx 1$. For cases in which $|\bar{T} - \bar{R}|$ is large (i.e., close to the maximum difference of T^*), then $f(|\bar{T} - \bar{R}|) \approx 0$ and no mean correction is used. Thus, the weighting of eq. (24) places extra emphasis on regions of high average energy difference, and less emphasis on regions of low average energy difference. We denote the distance measure of eq. (24) as the Silverman-Dixon (SD) distance measure.

The last variation on the conventional DTW algorithm that we have investigated is the relaxation of the endpoint constraints on the warping path. Normally we use the simple constraints that

$$w(1) = 1 \quad \text{Initial Point} \quad (26a)$$

$$w(M) = N_i \quad \text{Final Point,} \quad (26b)$$

i.e., the first test frame is mapped to the first reference frame and the last test frame is mapped to the last reference frame. We have considered relaxation of both endpoint constraints of eq. (26) to the form

$$1 \leq w(M_B) \leq \delta_{\text{BEG}}, \quad 1 \leq M_B \leq \delta_{\text{BEG}} \quad (27a)$$

$$N_i - \delta_{\text{END}} \leq w(M_E) \leq N_i, \quad M - \delta_{\text{END}} \leq M_E \leq M. \quad (27b)$$

The new endpoint constraints say that the warping path can begin anywhere within a square of size $\delta_{\text{BEG}} \times \delta_{\text{BEG}}$ at the origin of the test-reference plane, and end anywhere with a square of size $\delta_{\text{END}} \times \delta_{\text{END}}$ at the upper right-hand corner of the test-reference plane. This situation is depicted in Fig. 3. By using local path constraints, which keep the slope of the warping path greater than 1/2 and less than 2, the warping path becomes constrained to lie within the shaded area of the test-reference plane.

3.4 The normalize-and-warp procedure

The conventional DTW algorithm works quite well for most cases of interest. However, in cases when the length of the test pattern, M , is significantly different from the length of a reference pattern, N , then the region in the test-reference plane in which the warping path can lie often becomes very small. To handle such cases the normalize-and-warp procedure was devised,⁴ and it basically consists of linearly prenormalizing both the test and reference patterns to a fixed length, \bar{N} , and then performing the DTW on these equal length patterns. In this manner the area in the test-reference plane in which the warping path can lie is maximized; hence we have the best chance of finding a good time-alignment path.

The normalize-and-warp procedure has been successfully used in a number of tests with an LPC recognizer^{4,11-13} with very good results.

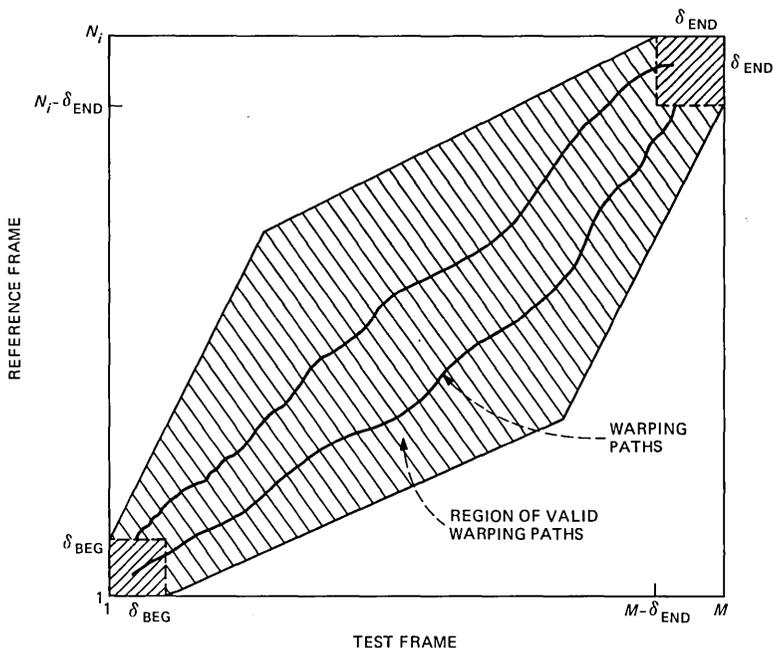


Fig. 3—The allowable warping region in the “test-reference” plane when the warping path has a beginning region of size $\delta_{\text{BEG}} \times \delta_{\text{BEG}}$ and an ending region of size $\delta_{\text{END}} \times \delta_{\text{END}}$.

For all these systems the fixed length to which all patterns were warped was the average duration of all words in the vocabulary. In this study we consider use of the normalize-and-warp procedure with the fixed length parameter a free variable.

3.5 Summary of signal processing choices

In this section we have enumerated a number of ways of implementing the signal processing of a filter bank isolated word recognizer. These factors include

- (i) Spectral preemphasis
- (ii) Thresholding of channel signals
- (iii) Energy normalization of channel signals
- (iv) Time smoothing of feature vectors
- (v) Frequency smoothing of channel outputs
- (vi) Quantization of channel signals
- (vii) Principal components analysis
- (viii) Time weighting of local DTW distances
- (ix) Frequency weighting of channel signals in DTW computation
- (x) Local distance metric for DTW computation
- (xi) Loosened endpoint constraints in DTW computation

(xii) Use of the normalize-and-warp procedure for DTW computation.

In Section V we give the actual choices that were studied for each of the above factors. First, however, we describe the tests of the noise immunity of both filter bank and LPC word recognizers.

IV. NOISE STUDIES WITH THE ISOLATED WORD RECOGNIZER

Almost all tests of isolated word recognizers are made in a laboratory environment with a high signal-to-noise ratio on the recordings (e.g., greater than 35 dB is typical). There are a wide variety of applications (namely those of the military) in which the word recognizer is required to operate in noisy environments [e.g., signal-to-noise ratios (s/n) around 0 to 20 dB]. Thus, an important consideration in the evaluation of an isolated word recognizer is how the performance degrades as the background goes from laboratory conditions to highly noisy conditions.

When a word recognizer must operate in high-background-noise environments, an important issue arises, namely, whether it is better to train the system in a noise-free environment (and test in the noisy background), or to train and test in the noisy background. We have attempted to study these questions by artificially adding uncorrelated, zero mean, white noise to the speech signals at a specified signal-to-noise ratio, and then performing the required word recognition tests on both the filter bank and LPC word recognizers. A discussion of the test conditions and the results is given in the next section.

When one is concerned with using a word recognizer in an environment with a poor signal-to-noise ratio, another important consideration is whether one would "cancel" any of the noise by using a noise spectral estimation technique and subtracting the noise spectrum out. These techniques have been investigated in the context of voice coding¹⁴⁻¹⁶ and have achieved various degrees of success. For sufficiently stationary noise backgrounds it seems reasonable to expect that a high degree of noise cancellation could be obtained. For such cases it would be of interest to understand how such noise cancellation algorithms work in the context of word recognition.

V. EXPERIMENTAL RESULTS ON ISOLATED WORD RECOGNITION

To evaluate the effects on performance (word error rate) of each of the recognition system factors of Sections III and IV, a series of tests were run with the following specifications:

Vocabulary	—	39 word alpha-digits
Number of talkers	—	2 male, 2 female
Training	—	7 replications for each word
Testing	—	10 replications for each word.

Table I—Word error rates for filter bank and LPC word recognizers

Talker	Candidate Position				
	1	2	3	4	5
(a) Rates for baseline filter bank recognizer as a function of talker and candidate position					
1 (Male)	9.0	4.1	0.5	0.5	0.0
2 (Male)	5.4	2.3	1.0	0.5	0.3
3 (Female)	13.1	2.8	0.5	0.3	0.3
4 (Female)	18.7	8.5	4.4	2.1	1.3
Average	11.6	4.4	1.6	0.9	0.5
(b) Rates for LPC recognizer as a function of talker and candidate position					
1 (Male)	5.1	0.5	0.0	0.0	0.0
2 (Male)	4.1	2.3	0.8	0.3	0.3
3 (Female)	10.3	2.3	1.3	1.0	0.5
4 (Female)	11.8	6.7	3.3	1.3	0.8
Average	7.8	3.0	1.4	0.7	0.4

All recordings were made over dialed-up telephone lines, and the test and training replications were obtained in different recording sessions. The speaker-dependent training used the robust training method⁶ to give a single reference pattern for each vocabulary word.

The filter bank used was the 13-channel, critical-band spacing system that gave essentially the best performance in earlier tests.¹ A “baseline” filter bank recognizer was defined that had the following signal processing options:

- (i) No preemphasis— $\alpha = 0$.
- (ii) Channel thresholding at $T^* = 50$ dB below the peak in each channel.
- (iii) Average energy normalization.
- (iv) No time smoothing— $M^* = 1$.
- (v) No frequency smoothing— $Q^* = 1$.
- (vi) No quantization of channel signals (i.e., floating point accuracy)— $B = \infty$.
- (vii) No principal components analysis.
- (viii) Uniform time weighting of local distances— $W^T(m) = 1$, all m .
- (ix) Uniform frequency weighting of local distances— $W_q^T = 1$, all q .
- (x) Magnitude local distance— $p = 1$.
- (xi) No opening up of DTW endpoint regions— $\delta_{\text{BEG}} = \delta_{\text{END}} = 1$.
- (xii) No length prenormalization prior to the DTW.
- (xiii) No additive noise— $s/n(\text{Test}) = s/n(\text{Train}) = \infty$.

The performance results of this baseline system are given in Table Ia and are shown plotted in Fig. 4a. Both the table and the figure show the word error rate as a function of candidate position for all four

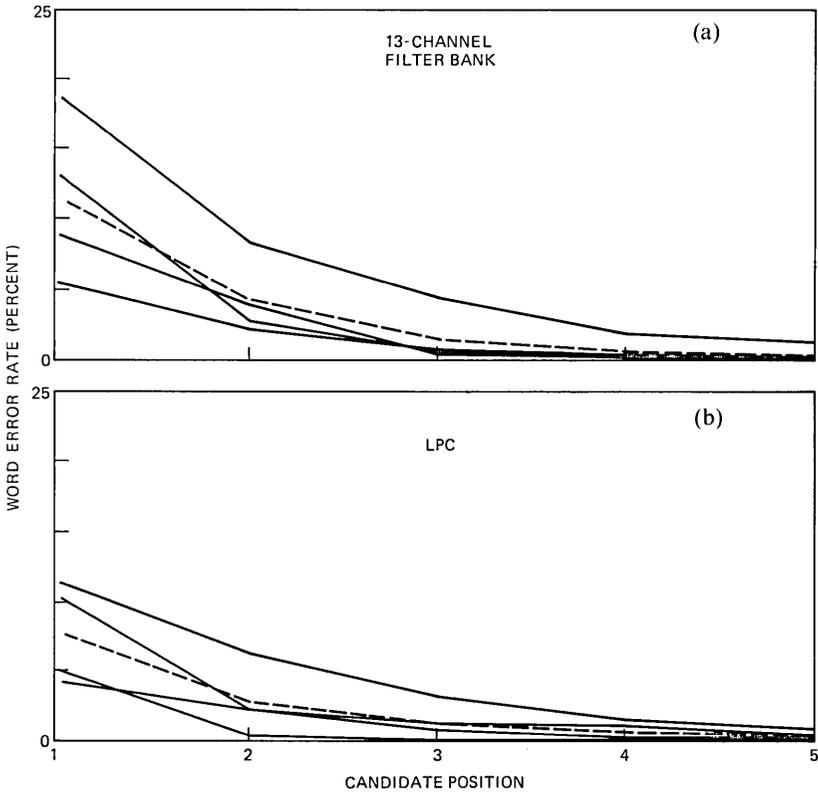


Fig. 4—Plots of word error rate versus candidate position for each of the four talkers and the average for (a) the baseline 13-channel filter bank system, and (b) the LPC word recognizer.

talkers (the solid curves in Fig. 4a) and the average (the dashed curve). These results show an average error rate of 11.6 percent in the top candidate position, with a high degree of variability in error rate across talkers. For comparison purposes, Table Ib and Fig. 4b show similar results on the LPC word recognizer. The average error rate for the top candidate position is about 4 percent lower for the LPC recognizer than for the 13-channel filter bank recognizer. Again we see a fair degree of variability in error rate scores across talkers for the LPC recognizer.

In the following sections we present results of tests designed to measure changes in performance of the filter bank recognizer as the factors noted above are varied. As discussed earlier we have been forced to use the expedient of only varying one parameter at a time; hence all information about interactions between two or more parameters is unavailable.

5.1 Effects of simple preemphasis

Only one value of the preemphasis constant, α , was studied, namely, $\alpha = 0.95$. This is the value used in previous work on the LPC recognizer.³ The results of recognition with the use of the preemphasis network are given in Table IIa (only results for top candidate position are included). It can be seen by comparing these results to those of the baseline system that a small improvement in average accuracy was obtained. This improvement was not statistically significant at the 0.9 confidence level.

5.2 Effect of clipping threshold

The values studied for the clipping threshold were $T^* = 40$ and $T^* = 30$ (dB). The resulting recognition scores are given in Table IIb for the top candidate position. The results for $T^* = 40$ are comparable to those for $T^* = 50$ in the baseline system, whereas for $T^* = 30$ a significant loss (2.4 percent) in word accuracy is obtained. Hence, a 30-dB range is deleterious to the channel signals in that useful recognition information is lost by clamping the signals at too high a level.

5.3 Effects of peak energy normalization

The results of using peak (rather than average) energy normalization of the channel signals are given in Table IIc.[†] The results show a large increase in word error rate for all talkers, thereby indicating a lack of stability of the peak in each frame and therefore its inappropriateness to be used as an energy normalization aid.

5.4 Effects of time smoothing

The value used for the smoothing duration, M^* , were 2 and 3 (frames). The recognition results for this condition are given in Table II d for the top candidate position. For $M^* = 2$ an insignificant increase in average error rate occurred, while for $M^* = 3$ there was a very significant increase in error rate. The results show that time smoothing produced far worse scores for female talkers (3 and 4) than for male talkers (1 and 2). This was undoubtedly due to the high variability in channel signals for the females (due to the high pitch frequency), which often led to smearing a "good" frame with a "bad" adjacent frame. The results indicate that time smoothing should not be done.

5.5 Effects of frequency smoothing

The results of smoothing across $Q^* = 2$ adjacent frequency channels are given in Table IIe. It can be seen that a uniform increase of about

[†] Recall that average normalization is one of the standard options used in the recognizer.

Table II—Word error rates for several signal processing techniques in the filter bank recognizer

Talker							
(a) Rate with preemphasis ($\alpha = 0.95$)							
α	1	2	3	4	Average		
0.95	7.7	7.7	11.8	17.4	11.2		
(b) Rates as a function of threshold parameter, T^* (dB)							
T^*	1	2	3	4	Average		
40	8.7	6.9	13.3	17.4	11.6		
30	9.5	8.5	17.2	21.0	14.0		
(c) Rates for peak energy normalization							
Method	1	2	3	4	Average		
Peak	23.1	19.5	14.4	40.5	24.4		
(d) Rates as a function of number of frames over which smoothing occurred, M^*							
M^*	1	2	3	4	Average		
2	7.9	5.6	13.1	21.3	12.0		
3	7.7	5.6	26.9	29.7	17.5		
(e) Rates as a function of number of frequency channels over which smoothing occurred, Q^*							
Q^*	1	2	3	4	Average		
2	12.6	7.4	15.4	20.5	14.0		
(f) Rates as a function of B , the number of bits used to quantize the channel signals							
B	1	2	3	4	Average		
6	6.7	5.6	14.4	17.4	11.0		
4	11.3	7.2	16.9	18.7	13.5		
(g) Rates as a function of P , the dimensionality of the principal components analysis							
P	1	2	3	4	Average		
12	11.3	9.9	25.1	29.5	18.7		
6	11.3						
4	11.5						
2	18.7						
(h) Rates using a nonuniform time weighting in the DTW algorithm							
	1	2	3	4	Average		
	9.2	6.5	14.5	19.5	12.4		
(i) Rates using a nonuniform frequency weighting in the DTW algorithm							
Weight	1	2	3	4	Average		
Fig. 5	8.7	8.7	16.2	21.3	13.7		
SD	8.2	7.4	16.4	17.4	12.4		
(j) Rates as a function of p , the power in the local distance computation							
p	1	2	3	4	Average		
2	12.3	7.1	21.4	23.1	16.0		
½	10.0	6.2	13.1	20.8	12.5		
(k) Rates as a function of the opening region parameters δ_{BEG} and δ_{END} of the DTW algorithm							
δ_{BEG}	δ_{END}	Region	1	2	3	4	Average
2	0	Square	8.7	5.6	13.1	22.3	12.4
0	4	Square	7.9	5.4	13.8	21.5	12.2
2	4	Square	8.5	5.6	14.4	22.8	12.8
2	4	Line	8.9	6.2	24.6	22.1	15.5

2.5 percent in word error rate is obtained for all four talkers. Thus we conclude that smoothing across channels leads to a loss in information for recognition and therefore should not be used.

5.6 Quantization of channel signals

The results on channel signal quantization are presented in Table II*f*. It can be seen that quantization of the channel signals to 6 bits actually decreases the average error rate by 0.6 percent; however, a further reduction to 4 bits leads to a 1.9-percent increase in error rate over the baseline system. Hence the results indicate that 6-bit quantization is adequate for the channel signals.

5.7 Results using principal components analysis

The results obtained using the principal components analysis for a single talker are presented in Table II*g*. It can be seen that for $P = 12$ a 7.1-percent increase in average error rate is obtained; however, small increases in error rate were attained for reductions in P down to 4. In fact, for talker 1 the word error rate increased by 0.2 percent in going from 12 to 4 principal components.

An explanation of why the $P = 12$ principal components analysis led to such large increases in error rate is as follows. The transformation of the feature vector used in the principal components analysis has the property that it is invariant to a quadratic distance measure. The distance measure used in the baseline system was an absolute value distance; hence a significant decrease in accuracy resulted. We will show in Section 5.10 that using a quadratic distance measure gave much worse recognition accuracy than the absolute distance. Thus it would appear that the principal components analysis is not a useful tool, at least for this particular filter bank word recognizer.

5.8 Results using time weighting in the global distance for DTW

The results using nonuniform time weighting in the DTW global distance calculation are given in Table II*h*. The actual weighting function, W^T , was a function only of the energy in the test pattern, of the type shown in Fig. 5. The test energy, E^T , was estimated as the sum of the individual channel energies. A high correlation (≈ 0.94) was measured between this estimate of test energy, and the actual test energy (as computed from the raw speech samples). For frames in which the test energy was within 20 dB of the peak energy (suitably normalized to 0 dB), the frame weight was 1.0; for frames with E_T less than 40 dB below the peak, the weight was set to 0.01; a linear interpolation of the weight was used for frames with $-40 \leq E_T \leq -20$ dB.

The results in Table II*h* show a small increase in word error rate for

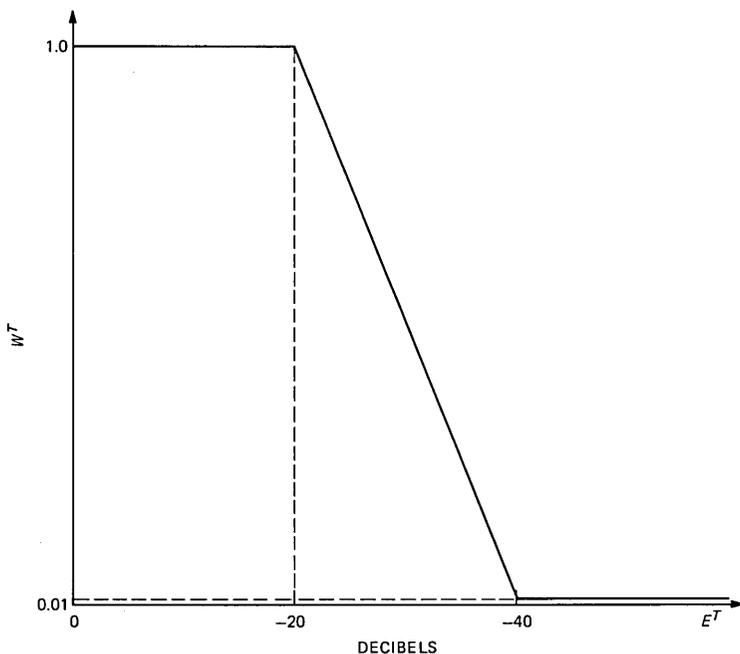


Fig. 5—The nonlinear weighting function, W^T , on the local distance as determined from the test energy, E^T , estimated from the sum of channel outputs.

each talker. Hence we conclude that the addition of time weighting (of the form of Fig. 5) in the DTW distance calculation is unnecessary.

5.9 Results using frequency weighting in the local distance for DTW

The results of using a nonuniform frequency weighting for local distances in the DTW algorithm are shown in Table III. The frequency-weighting characteristic was identical to the time-weighting characteristic of Fig. 5, except that the abscissa was the individual channel energy (relative to the peak channel energy for the word) and the ordinate was frequency weight W_q . It can be seen from Table III that a 2.1-percent increase in average word error rate is obtained using the nonuniform frequency weighting. Hence we again conclude that such weighting should not be used for this particular filter bank recognizer.

The results of using the SD weighting proposed by Silverman and Dixon (based on both reference and test frame energies) are also shown in Table III. Although the average word error rate is lower than for the nonuniform weighting of Fig. 5, it is still about 0.8 percent higher than obtained using simple uniform weights.

5.10 Effects of different local distance computations

The results of using different local distance computations in the DTW algorithm are given in Table IIj. Values for p of 2 (squared distance) and $1/2$ (square root distance) were considered. It can be seen that for $p = 2$ a 4.4-percent increase in average word error rate is obtained; for $p = 1/2$ the increase in average word error rate is 0.9 percent. These results indicate that the magnitude distance ($p = 1$) is the best compromise between giving extra weight to very different channel energies ($p = 2$) and giving a small weight to very different channel energies ($p = 1/2$).

5.11 Results of opening up the DTW starting and ending regions

The results of opening up the beginning and/or ending region of the DTW speech regions are given in Table IIk. Results are given for an initial or final square search region, as well as for an initial or final line search region (i.e., the path had to begin or end at the first or last frame of *either* the test or reference; it could not begin or end at a noninitial or nonfinal frame of *both*). It can be seen that all the cases studied led to a small (for square regions) or a large (for the line region) increase in average word error rate. This result is anticipated from previous results, which have shown that opening up the DTW search region consistently aids false matches (reference and test different) as much or more than true matches (reference and test the same).^{3,4,17}

5.12 Results using length normalization prior to DTW

The results of using fixed-length word normalization prior to the DTW (the normalize-and-warp procedure) are given in Table III and Fig. 6. Table IIIa and Fig. 6a show results for the 13-channel filter bank and Table IIIb and Fig. 6b show results for the LPC-based recognizer. The results show that for a broad range of warping lengths (from 20 to 45 frames) the average word accuracy does not change significantly. Significant degradation in performance is obtained only for the shortest warping lengths considered (i.e., 10 and 15 frames). Hence the results indicate that the normalize-and-warp procedure is suitable for a sizeable range of warping lengths so long as the length used does not become too small.

5.13 Results on noise studies

The results of the noise studies are given in Table IV and plotted in Fig. 7. Results are given for three cases:

(i) Signal-to-noise ratio (s/n)(Test) = s/n(Train), where s/n varied from ∞ down to 0 dB (Table IVa, Fig. 7a).

Table III—Word error rates for the normalize-and-warp procedure applied to both the filter bank and LPC word recognizers

		Length of Reference and Test							
(a) Rates as a function of warping length of reference and test prior to DTW for the 13-channel recognizer									
Talker	Variable	40	30	25	20	15	10		
1	9.0	8.7	9.5	8.7	8.7	10.3	12.6		
2	5.4	5.6	5.4	5.9	5.6	6.9	11.5		
3	13.1	13.1	14.6	14.1	14.6	15.4	19.5		
4	18.7	21.8	20.8	21.8	20.8	22.8	23.3		
Average	11.6	13.1	12.6	12.6	12.7	13.9	16.7		
(b) Rates as a function of warping length of reference and test prior to DTW for the LPC-based recognizer									
Talker	Variable	45	40	35	30	25	20	15	10
1	5.1	4.9	6.2	4.6	4.6	5.1	4.9	8.5	11.8
2	4.1	4.6	3.8	4.9	5.1	4.9	5.6	6.9	7.4
3	10.3	8.2	9.0	7.9	8.5	10.0	10.5	12.6	12.8
4	11.8	11.3	12.1	12.6	11.5	12.6	12.8	14.1	20.0
Average	7.8	7.3	7.8	7.5	7.4	8.2	8.5	10.5	13.0

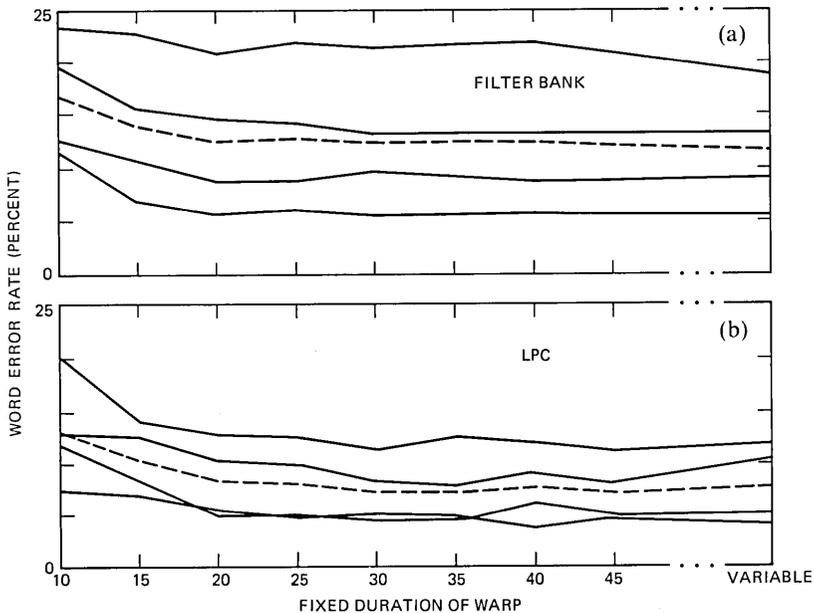


Fig. 6—Plots of word error rate versus fixed frame duration for linear prewarp prior to DTW alignment for (a) the filter bank system, and (b) the LPC system.

(ii) $s/n(\text{Train}) = \infty$, $s/n(\text{Test})$ variable from ∞ down to 0 dB (Table IVb, Fig. 7b).

(iii) $s/n(\text{Test}) = 18$ dB, $s/n(\text{Train})$ variable from ∞ down to 0 dB (Table IVc, Fig. 7c).

The first case represents the situation when both training and testing

Table IV—Word error rates for noise studies

	s/n (dB)							
(a) Rates for filter bank (FB) and LPC word recognizers as a function of s/n for case when noise added to both test and reference signals								
System	∞	30	24	18	12	6	0	
FB	9.0	9.5	11.0	13.1	13.6	16.7	21.0	
LPC	5.1	3.8	6.2	7.4	11.3	16.7	23.6	
(b) Rates for filter bank (FB) and LPC word recognizers as a function of s/n for the case when noise added to test only—i.e., s/n of reference for training was ∞								
System	∞	30	24	18	12	6	0	
FB	9.0	14.4	17.3	32.6	61.0	82.3	92.3	
LPC	5.1	10.0	17.4	37.2	65.1	76.9	90.8	
(c) Rates for filter bank (FB) and LPC word recognizers as a function of s/n for case when noise added to reference at variable s/n, and with test s/n set to 18 dB								
System	∞	30	24	18	12	6	0	
FB	32.6	12.3	11.8	13.1	14.1	18.5	89.5	
LPC	37.2	10.3	7.2	7.4	10.0	17.2	31.5	
(d) Rates for filter bank (FB) and filter bank with noise removal (FB/NR) for case when noise added to reference at variable s/n, and with test s/n set to 18 dB								
System	∞	30	24	18	12	6	0	
FB	32.6	12.3	11.8	13.1	14.1	18.5	89.5	
FB/NR	26.2	12.6	11.8	13.1	14.4	18.7	28.2	

of the word recognizer are done in the same noisy background; case (ii) represents the situation when there is “clean” training (no noise) but the test words are spoken in the noisy background; case (iii) represents the situation when there is noise in both training and testing; however there may be a mismatch in s/n.

The results in Table IV and Fig. 7 show that:

(i) For case (i), the LPC system performs as well as or better than the filter bank (FB) system for $s/n \geq 6$ dB. The filter bank (FB) system outperformed the LPC system only at a 0 dB s/n.

(ii) For case (i), there was little degradation in performance down to s/n's of close to 24 dB for either the FB or LPC recognizer.

(iii) For case (ii) the performance of both the FB and LPC recognizers was significantly worse at all s/n's than for case (i). Hence we see that using clean training data with noisy test data leads to badly degraded system performance for $s/n \leq 30$ dB.

(iv) For case (iii) the results indicate that when $s/n(\text{Test})$ and $s/n(\text{Train})$ differ by as little as 6 dB (or more) degraded performance results.

The results of Table IV and Fig. 7 indicate that it is mandatory that both the training (reference) and testing data be obtained in the same background noise conditions for best word recognition performance.

A test was also conducted on the filter bank recognizer to determine whether the effects of additive (background) noise could be lessened by subtracting out an (estimated) average noise spectrum prior to recognition. A one-second average was calculated for each channel

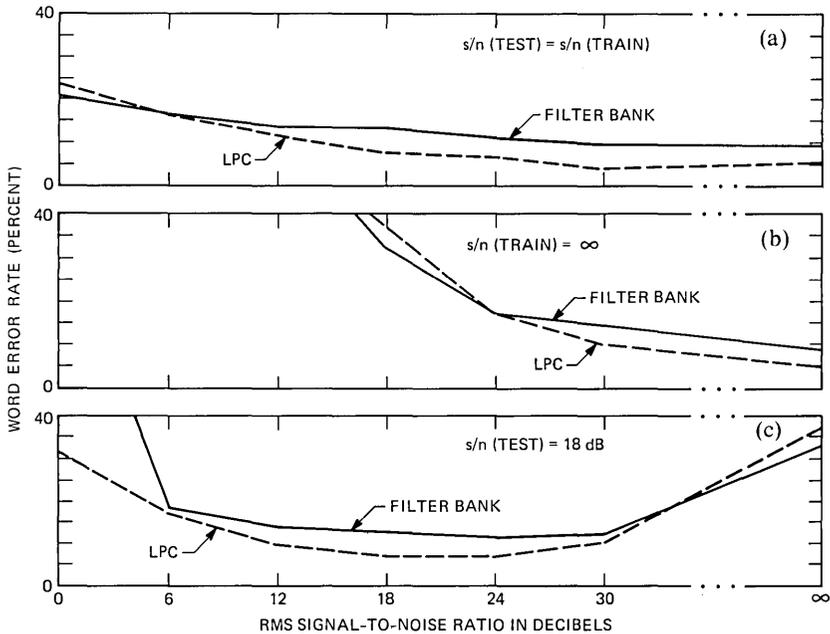


Fig. 7—Plots of word error rate (for a single talker) versus signal-to-noise ratio for both the LPC and filter bank recognizers for, (a) $s/n(\text{Test}) = s/n(\text{Train})$; (b) $s/n(\text{Train}) = \infty$, $s/n(\text{Test})$ variable; and (c) $s/n(\text{Test}) = 18 \text{ dB}$, $s/n(\text{Train})$ variable.

signal of the filter bank when only the additive white noise was present at the input. Each of these 13-channel average noise values was then subtracted from the corresponding channel signal to form a new channel signal, which was used in the recognition processing. For these tests the signal-to-noise ratio for the test data was held constant at 18 dB, and the noise level in the reference data was varied to give signal-to-noise ratios between 0 dB and ∞ .

The results of this experiment are given in Table IVd. These results show that the effects of this simple noise-cancelling arrangement are to broaden the range of signal-to-noise ratios over which the filter bank recognizer can operate. It can be seen that recognizer performance is not changed significantly between 30-dB and 6-dB signal-to-noise ratios from that obtained without the noise cancellation. However, for signal-to-noise ratios of ∞ and 0 dB, a considerable reduction of error rate is obtained with the noise cancellation method. The conclusion from this test is that noise cancelling is useful for reducing the effects of variations in noise level between reference and test data.

VI. DISCUSSION AND CONCLUSIONS

The results presented in Section V lead to the following conclusions:

- (i) Essentially none of the proposed signal processing techniques

for use in the filter bank word recognizer led to an improvement in performance of the system (i.e., reduced word error rate). At best any single technique led to a small (insignificant) increase or decrease in word error rate; at worst it led to a significant increase in word error rate.

(ii) The filter bank coefficients (for telephone inputs) needed only about 6 uniform bits for a representation with no increase in word error rate. Hence, the storage requirements on the $Q = 13$ channel recognizer were about 78 bits per frame using this 6-bit coding scheme.

(iii) The normalize-and-warp procedure was an effective method for reducing storage and processing requirements in the DTW computation in that fixed duration linear prewarps of size as small as 20 frames per word did not increase word error rate significantly for either the LPC or FB recognizers.

(iv) The best strategy for using a word recognizer in a noisy background was to both train and test the recognizer in the same noise background.

(v) The LPC word recognizer gave error rates the same or lower than the FB word recognizer for $s/n \geq 6$ dB.

Our initial goal was to find signal processing techniques to enhance the performance of the FB word recognizer so as to come closer to that of an existing LPC word recognizer. Our results indicate that we have not succeeded in attaining this goal. Hence our main question is whether we failed because we tried the wrong things, or because there is no way of doing consistently better with the FB limitations. There is no simple answer to this question. Perhaps our best response is that we tried a wide range of techniques that encompassed those methods previously proposed and studied in other FB recognizers. The lack of any significant improvement in performance for any of the proposed techniques indicates to us that perhaps the only way to improve accuracy is by some heuristic based on linguistic knowledge of the vocabulary words. We have meticulously avoided such techniques as they change the nature of the recognizer from a vocabulary-independent system to one that depends on the specific words to be recognized.

Another possible objection to the conclusions as drawn from the results given in Section V is that we studied each proposed signal processing technique independently. As such we avoided interactions between techniques that could have led to improved accuracy. Again we iterate our speculation that since no individual technique led to a real performance improvement, we are skeptical that combinations of techniques would lead to real improvements. Of course we have no concrete evidence that this is indeed the case.

Our noise analysis results dispel the common notion that LPC recognizers "fall apart" in noisy backgrounds while FB recognizers

degrade gracefully. Our results show that with proper training the LPC system outperforms the FB system at all reasonable signal-to-noise ratios.

Finally, the noise results show that training and testing should *always* be done in the same acoustic backgrounds. If there are gross differences in acoustic backgrounds, significant degradation in performance results.

VII. SUMMARY

We have presented results of a study to measure the effects of selected signal processing techniques on the performance of a filter bank word recognizer. We have shown that a fairly simple set of signal processing techniques led to the best overall performance of the word recognizer in the noise-free case. In noisy conditions the performance of the recognizer degraded significantly for signal-to-noise less than about 24 dB.

REFERENCES

1. B. A. Dautrich, L. R. Rabiner, and T. B. Martin, "On the Effects of Varying Filter Bank Parameters on Isolated Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-31*, No. 4 (August 1983).
2. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-23* (February 1976), pp. 67-72.
3. L. R. Rabiner and S. E. Levinson, "Isolated and Connected Word Recognition—Theory and Selected Applications," *IEEE Trans. on Communication, COM-29* (May 1981), pp. 621-59.
4. C. S. Myers, L. R. Rabiner, and A. E. Rosenberg, "Performance Trade-offs in Dynamic Time Warping Algorithms for Isolated Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-28* (December 1980), pp. 623-35.
5. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, New York, NY: Springer-Verlag, 1976, Chapter 9, pp. 212-26.
6. L. R. Rabiner and J. G. Wilpon, "A Simplified Robust Training Procedure for Speaker Trained, Isolated Word Recognition Systems," *J. Acoust. Soc. Amer.*, 68 (November 1980), pp. 1271-6.
7. L. C. W. Pols, "Real-Time Recognition of Spoken Words," *IEEE Trans. Comput., C-20* (September 1971), pp. 972-8.
8. H. Sakoe and S. Chiba, "Dynamic Programming Optimization for Spoken Word Recognition," *IEEE Trans. Acoust., Speech, and Signal Processing, ASSP-26* (February 1978), pp. 43-9.
9. L. R. Rabiner, A. E. Rosenberg, and S. E. Levinson, "Considerations in Dynamic Time Warping Algorithms for Discrete Word Recognition," *IEEE Trans. Acoust., Speech, and Signal Processing, ASSP-26*, No. 6 (December 1978), pp. 575-82.
10. H. F. Silverman and N. R. Dixon, "A Comparison of Several Speech-Spectra Classification Methods," *IEEE Trans. Acoust., Speech, and Signal Processing, ASSP-24*, No. 4 (August 1976), pp. 289-95.
11. J. G. Wilpon, L. R. Rabiner, and A. Bergh, "Speaker Independent Isolated Word Recognition Using a 129 Word Airline Vocabulary," *J. Acoust. Soc. Am.*, 72 (August 1982), pp. 390-6.
12. M. K. Brown and L. R. Rabiner, "An Adaptive Ordered Graph Search Technique for Dynamic Time Warping for Isolated Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-30*, No. 4 (August 1982), pp. 535-44.
13. M. K. Brown and L. R. Rabiner, "On the Use of Energy in LPC-Based Recognition of Isolated Words," *B.S.T.J.*, 61, No. 10, Part 1 (December 1982), pp. 2971-87.
14. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction,"

- IEEE Trans. Acoustics, Speech, and Signal Processing, *ASSP-29*, No. 2 (April 1979), pp. 113-20.
15. J. S. Lim and A. V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc. IEEE*, *67*, No. 12 (December 1979), pp. 1586-604.
 16. M. M. Sondhi, C. E. Schmidt, and L. R. Rabiner, "Improving the Quality of a Noisy Speech Signal," *B.S.T.J.*, *60*, No. 8 (October 1981), pp. 1847-59.
 17. L. R. Rabiner, "Note on Some Factors Affecting Performance of Dynamic Time Warping Algorithms for Isolated Word Recognition," *B.S.T.J.*, *61*, No. 3 (March 1982), pp. 363-73.

CONTRIBUTORS TO THIS ISSUE

Harry J. Boll, B.S., M.S., Ph.D. (Electrical Engineering), University of Minnesota, Minneapolis, in 1956, 1958, and 1962, respectively; Bell Laboratories, 1962—. After completing graduate work in secondary electron emission, he joined Bell Laboratories, Murray Hill, NJ, where he has worked in the general area of semiconductor devices and integrated circuits. He is presently a Group Supervisor in the Advanced Large-Scale Integration Development Laboratory, where his duties include very large-scale integrated logic design. Fellow, IEEE; member, Eta Kappa Nu, Tau Beta Pi. He received a Bell Laboratories Fellowship award for his work in custom integrated circuits.

Ta-Shing Chu, B.S., 1955, National Taiwan University; M.S., 1957, and Ph.D., 1960, Ohio State University; Research Associate, Courant Institute of Mathematical Sciences, New York University, 1961-1963; Bell Laboratories, 1963—. Mr. Chu has been engaged in research on microwave antennas and wave propagation for satellite communication and terrestrial radio systems. His work included surface wave diffraction, precision gain standard, propagation through precipitation, dual-polarization radio transmission, and the Crawford Hill 7-meter offset millimeter wave antenna. Fellow, IEEE; Member, International Scientific Radio Union, Sigma Xi, Pi Mu Epsilon.

Bruce A. Dautrich, B.S. (Electrical Engineering), 1980, Pennsylvania State University; M.S. (Electrical Engineering), 1981, Purdue University; Bell Laboratories, 1980—. Mr. Dautrich has been engaged in speech communication research and development and is presently concentrating on problems of speech recognition. Member Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi.

Bharat T. Doshi, B. Tech (Mechanical Engineering), 1970, I.I.T., Bombay; Ph.D. (Operations Research), 1974, Cornell University; Bell Laboratories, 1979—. Before joining Bell Laboratories Mr. Doshi was an Assistant Professor at Rutgers University. At Bell Laboratories his technical work has included modeling and analysis of processor schedules and overload controls. His research interests include queueing theory applied to performance analysis of computer and communication systems. Member, IEEE, ORSA; Associate Editor, OR Letters.

Robert W. England, B.S., 1973, Capitol Institute of Technology; Bell Laboratories, 1973-1981; AT&T Long Lines, 1981—. At Bell Laboratories, Mr. England worked on microwave antennas for satellite

communication. He was involved in the measurement of the Crawford Hill 7-meter antenna and the development of an earth station receiver for the ATS-6 satellite. Since coming to Long Lines, Mr. England has been Senior Engineer on Long Lines' new coaxial cable transmission system.

Michael J. Grieco, B.S. (Chemistry), 1966, Rutgers University, New Brunswick, NJ; Bell Laboratories, 1961-68, 1972—. Mr. Grieco has worked in many aspects of semiconductor processing such as photolithography, etching, C.V.D., and wafer cleaning. He has been involved in photolithographic process engineering for the *BELL-MAC*-8*, -4, and most recently the *BELLMAC-32* microprocessor. This process engineering work is being continued in 1- μ m MOS device technology.

Raghu N. Kackar, B.Sc. (Mathematics, Physics, Chemistry), 1970, University of Delhi; M. Stat., 1972, Agra University; M.S. (Statistics), 1974, University of Guelph; Ph.D. (Statistics), 1979, Iowa State University; Bell Laboratories, 1980—. Mr. Kackar is applying new methods for total quality control to enhance quality and productivity of several production lines. At present his primary emphasis is on development and application of the technique of designed experiments. Member, ASA, ASQC.

Thomas B. Martin, B.S.E.E., 1957, University of Notre Dame; M.S.E.E., Ph.D., University of Pennsylvania, 1960 and 1970, respectively. In 1970 he co-founded Threshold Technology Inc., which developed the first practical automatic speech recognition systems to be used in industry. In March 1982, he joined Bell Laboratories and has continued to work in speech recognition.

Stephen C. Mettler, B.S., 1962, U.S.A.F. Academy; M.S. (Physics), 1972, and Ph.D. (M.E.), 1976, Purdue University; Bell Laboratories, 1976—. Mr. Mettler is presently engaged in research on optical fiber components and splicing. Member, OSA.

Kinichiro Ogawa, B.S.E.E., 1966, M.S.E.E., 1968, University of Tokyo, D.Sc. (Electrical Engineering), Washington University; Nippon Telegraph & Telephone Public Corporation, 1968-1976; Bell Laboratories, 1976—. At N.T.T., Mr. Ogawa worked on long-haul analog coaxial cable systems and on video transmission. He also worked on

* Trademark of Bell Laboratories

the development of pulse-code modulated/frequency division multiplex digital systems. At Bell Laboratories, he has been involved in research on optical data links and digital lightwave systems. Member, IEEE, Optical Society of America, IECE of Japan.

Brian Owen, B.Sc., 1963; Ph.D., 1967 (Electrical Engineering), Birmingham University, England; Bell Laboratories, 1968—. Mr. Owen is a Member of the Technical Staff in the Lightwave Subsystems Department at Allentown. From 1968 to 1977 he was involved in the design of microwave and millimeter-wave components. Since 1977 he has been involved in photonics working primarily on optical receivers.

Madhav Phadke, B. Tech, 1969 (Mechanical Engineering), Indian Institute of Technology, Bombay; M.S., 1972 (Statistics), and Ph.D., 1973 (Mechanical Engineering), University of Wisconsin, Madison; Bell Laboratories, 1977—. Prior to joining Bell Laboratories, Mr. Phadke worked as a Research Associate in the Statistics Department of the University of Wisconsin and as a Visiting Scientist at the IBM Watson Research Center. He has published articles in professional journals covering topics in quality assurance, design of experiments, process control, system identification, time series analysis, mechanical design, and air pollution data analysis. His current interests are in the applications of Japanese methods of quality control and productivity improvement in design, development, and manufacturing of products. Member, Phi Kappa Phi, Sigma Xi, American Statistical Association, the American Society for Quality Control.

Lawrence R. Rabiner, S.B. and S.M., 1964, Ph.D. (Electrical Engineering), The Massachusetts Institute of Technology; Bell Laboratories, 1962—. From 1962 through 1964, Mr. Rabiner participated in the cooperative plan in electrical engineering at Bell Laboratories. He worked on digital circuitry, military communications problems, and problems in binaural hearing. Presently, he is engaged in research on speech communications and digital signal processing techniques. He is coauthor of *Theory and Application of Digital Signal Processing* (Prentice-Hall, 1975), *Digital Processing of Speech Signals* (Prentice-Hall, 1978), and *Multirate Digital Signal Processing* (Prentice Hall, 1983). Former President, IEEE, ASSP Society; former Associate Editor, ASSP Transactions; former member, Technical Committee on Speech Communication of the Acoustical Society, ASSP Technical Committee on Speech Communication; Member, IEEE Proceedings Editorial Board, Eta Kappa Nu, Sigma Xi, Tau Beta Pi. Fellow, Acoustical Society of America, IEEE.

Donald V. Speeney, Assoc. in Technology (Electronics), 1960, Temple University; Bell Laboratories, 1960—. Since joining Bell Laboratories, Mr. Speeney has worked in the area of semiconductor device technology development and device physics. This includes work on the discrete components of germanium, silicon, and gallium arsenide, as well as CMOS microcomputer and microprocessor integrated circuits. More recently, he has been working on processes and technology for producing silicon integrated circuits.

York Y. Wang, B.S. (Physics), 1970, Chung-Yuan Christian College, Taiwan; M.S. (Physics), 1972, and Ph.D. (Physics), 1976, University of Missouri at Columbia; Jet Propulsion Laboratory, 1975-1977, Bell Laboratories, 1977—. At Jet Propulsion Laboratory, Mr. Wang worked on Viking and Voyager Projects as a mission planner. At Bell Laboratories, he has been involved in microwave radio propagation studies, space diversity signal combining techniques, digital microwave radio systems performance and applications, digitization of analog microwave radio systems, and advanced network services planning. Member, IEEE, Sigma Xi.

Ian A. White, B.Sc. (Hon.) (Physics), 1974, University of Queensland, Australia; Ph.D. (Applied Mathematics), 1977, Australian National University; Bell Laboratories, 1979—. From 1977 to 1979 Mr. White did post-doctoral research in the theory of fiber and integrated optics at the Optical Sciences Center, University of Arizona at Tucson. Since 1979 he has been with Bell Laboratories, Norcross, Georgia, where he is a member of the Exploratory Components and Media Studies Group.

PAPERS BY BELL LABORATORIES AUTHORS

COMPUTING/MATHEMATICS

- Abramovici M., Breuer M. A., **Fault-Diagnosis in Synchronous Sequential-Circuits Based on an Effect Cause Analysis.** IEEE Comput 31(12):1165-1172, 1982.
- Brown D. J., Baker B. S., Katseff H. P., **Lower Bounds for Online Two-Dimensional Packing Algorithms.** ACT Inform 18(2):207-225, 1982.
- Ding Z. D., Hwang F. K., **Minimizing a Combinatorial Function.** Siam J Alg 3(4):523-528, 1982.
- Finkel R. A., Fishburn J. P., **Improved Speedup Bounds for Parallel Alpha-Beta Search (Letter).** IEEE Patt. A 5(1):89-92, 1983.
- Flatto L., **A Geometric Problem in Approximation-Theory.** J Approx Th 36(4):321-327, 1982.
- Golumbic M. D., Rotem D., Urrutia J., **Comparability-Graphs and Intersection Graphs.** Discr Math 43(1):37-46, 1983.
- Irani K. B., Chen K. W., **Minimization of Interprocessor Communication for Parallel Computation.** IEEE Comput 31(11):1067-1075, 1982.
- Kasami T., Lin S., Wei V. K., Yamamura S., **Graph Theoretic Approaches to the Code Construction for the 2-User Multiple-Access Binary Adder Channel.** IEEE Info T 29(1):114-130, 1983.
- Lagarias J. C., **Best Simultaneous Diophantine Approximations. 2. Behavior of Consecutive Best Approximations.** Pac J Math 102(1):61-88, 1982.
- Melici J. A., **The BX-25 Certification Facility.** Comput Netw 6(5):319-329, 1982.
- Nair V. N., **Q-Q Plots with Confidence Bands for Comparing Several Populations.** Sc J Stat 9(4):193-200, 1982.

ENGINEERING

- Anthony P. J., **Alteration of Diffusion Profiles in Semiconductors Due to p-n Junctions.** Sol ST Elec 25(12):1171-1177, 1982.
- Bohrer M. P., **Diffusional Boundary-Layer Resistance for Membrane-Transport.** Ind Eng F 22(1):72-78, 1983.
- Chen P. C., Pavlidis T., **Segmentation by Texture Using Correlation (Letter).** IEEE Patt A 5(1):64-69, 1983.
- Cheng J., Boyd G. D., Westerwick E. H., **Reduction of Threshold Voltage for Bistable Liquid-Crystal Switching Using a Priming Voltage.** IEEE Device 29(12):1853-1856, 1982.
- Damen T. C., Duguay M. A., Miller B. I., Lin C., **Single-Mode Semiconductor-Laser Optically Pumped by an Injection-Laser.** Electr Lett 19(1):3-4, 1983.
- Emkey W. L., Bogert G. A., **Multiport Polarization Monitor.** Optics Lett 8(2):94-95, 1983.
- Engelmaier W., Frisch D. C., **Injection-Molding Shapes New Dimensions for Boards.** Electronics 55(25):155-158, 1982.
- Falcone R. W., Bokor J., **Dichroic Beam Splitter for Extreme-Ultraviolet and Visible Radiation.** Optics Lett 8(1):21-23, 1983.
- Fichtner W., Cooper J. A., Tretola A. R., Kahng D., **A Novel Buried-Drain DMOSFET Structure.** IEEE Device 29(11):1785-1791, 1982.
- Fork R. L., Shank C. V., Hirlimann C., Yen R., Tomlinson W. J., **Femtosecond White-Light Continuum Pulses.** Optics Lett 8(1):1-3, 1983.
- Forrest S. R., **Photo-Diodes for Long-Wavelength Communication-Systems.** Laser Focus 18(12):81+, 1982.
- Gilbert J. A., et al., **The Application of Fiber Optics to Remote Speckle Metrology Using Incoherent-Light.** Opt Laser E 3(3):183-195, 1982.
- Glass A. M., Liao P. F., Olson D. H., Humphrey L. M., **Optical Metal-Oxide Tunnel Detectors with Microstructured Electrodes.** Optics Lett 7(12):575-577, 1982.

- Gloge D. C., **Citation Classic—Multimode Theory of Graded-Core Fibers.** *CC/Eng Tech* 1983(4):22, 1983.
- Herbst D., Bosch M. A., Lemons R. A., Tewksbury S. K., Harrison T. R., **PMOS Transistors Fabricated in Large-Area Laser-Crystallized Si on Silica.** *Electr Lett* 19(1):12-13, 1983.
- Hu E. L., Tennant D. M., Howard R. E., Jackel L. D., Grabbe P., **Vertical Silicon Membrane Arrays Patterned With Tri-Level E-Beam Resist.** *J Elec Mat* 11(5):883-888, 1982.
- Iannuzzi M., Kozakiewicz R. P., **Bias Humidity Performance and Failure Mechanisms of Nonhermetic Aluminum SICs in an Environment Contaminated with SO₂.** *IEEE Compon* 5(4):345-354, 1982.
- Kaminow I. P., Eisenstein G., Stulz L. W., Dentai A. G., **Lateral Confinement InGaAsP Superluminescent Diode at 1.3 μm.** *IEEE J Q El* 19(1):78-82, 1983.
- Kavehrad M., **Optimum QPSK Transmission Over Bandlimited Channels With Cascaded Nonlinearities.** *Proc. Thirteenth Ann. Pittsburg Conf. 13, Pt 1:153-161, 1982.*
- Kavehrad M., **Phase-Noise Effects on QPSK Carriers in Burst Transmission.** *IEE J Comm Radars Sig Proc* 129 Pt F No 5:366-372, 1982.
- Keller H. N., **Significant Features of Solder Connections to Gold-Plated Thin-Films.** *IEEE Compon* 5(4):408-419, 1982.
- Kuo J. M., Thiel F. A., **Study of Deep Level in Bulk P-InP By Admittance Spectroscopy.** *Electr Lett* 19(2):41-43, 1983.
- Macklin J. J., Wood O. R., Silfvast W. T., **New Recombination Lasers in Li, Al, Ca, and Cu in a Segmented Plasma-Device Employing Foil Electrodes.** *IEEE J Q El* 18(11):1832-1835, 1982.
- Marcuse D., **Computer-Model of an Injection-Laser Amplifier.** *IEEE J Q El* 19(1):63-73, 1983.
- McFee J. H., Swartz R. G., Archer V. D., Finegan S. N., Feldman L. C., **A Quantitative Study of the Relationship Between Interfacial Carbon and Line Dislocation Density in Silicon Molecular-Beam Epitaxy.** *J Elchem So* 130(1):214-216, 1983.
- Pearsall T. P., Hendel R., Oconnor P., Alavi K., Cho A. Y., **Selectively-Doped Al_{0.48}In_{0.52}As/Ga_{0.47}In_{0.53}As Heterostructure Field-Effect Transistor.** *Elec Dev L* 4(1):5-8, 1983.
- Peterson G. E., Paek U. C., **Electrical Transmission-Lines as Models For Soliton Propagation in Infrared Fibers.** *P Soc Photo* 320:62-68, 1982.
- Sanders J. V., **A Three-Dimensional Dynamic Analysis of a Towed System.** *Ocean Eng* 9(5):483-499, 1982.
- Sargent M., et al., **Quantum-Theory of Laser and Optical-Bistability Instabilities.** *Optics Lett* 8(2):76-78, 1983.
- Schwarz S. A., Thornber K. K., **Mobility Degradation due to the Gate Field in the Inversion Layer of MOSFETS—Comment.** *Elec Dev L* 4(1):11-12, 1983.
- Sessler G. M., et al., **Non-Destructive Laser Method for Measuring Charge Profiles in Irradiated Polymer-Films.** *IEEE Nucl S* 29(6):1644-1649, 1982.
- Wiesenfeld J. M., Chraplyvy A. R., Stone J., Burrus C. A., **Measurement of Very High-Speed Photodetectors with Picosecond InGaAsP Film Lasers.** *Electr Lett* 19(1):22-24, 1983.

MANAGEMENT/ECONOMICS

- Fishburn P. C., **Time Preference.** *Int Econ R* 23(3):677-694, 1982.
- Hakansson N. H., Kunkel J. G., Ohlson J. A., **Sufficient and Necessary Conditions for Information to have Social Value in Pure Exchange.** *J Finance* 37(5):1169-1181, 1982.
- Ordover J. A., Panzar J. C., **On the Non-Linear Pricing of Inputs.** *Int Econ R* 23(3):659-675, 1982.
- Rassenti S. J., Smith V. L., Bulfin R. L., **A Combinatorial Auction Mechanism for Airport Time Slot Allocation.** *Bell J Econ* 13(2):402-417, 1982.
- Vinod H. D., **Maximum-Entropy Measurement Error-Estimates of Singular Covariance Matrices in Undersized Samples.** *J Economet* 20(2):163-174, 1982.

PHYSICAL SCIENCES

- Allen S. J., Allyn C. L., Cox H. M., Derosa F., Mahoney G. E., **Dispersion of the Saturated Current in GaAs from DC to 1200 GHz.** *Appl Phys L* 42(1):96-98, 1983.
- Baumgart H., Leamy H. J., Celler G. K., Trimble L. E., **Grain-Boundary Diffusion in Polycrystalline Silicon Films on SiO₂.** *J Physique* 43(NC1):363-368, 1982.
- Becker R. S., Golovchenko, J. A., Patel J. R., **X-Ray Evanescent-Wave Absorption and Emission.** *Phys Rev L* 50(3):153-156, 1983.
- Behringer R. P., et al., **Heat-Transport and Temporal Evolution of Fluid-Flow Near the Rayleigh-Benard Instability in Cylindrical Containers.** *J Fluid Mec* 125(DEC):219-258, 1982.
- Bondybey V. E., Schwartz G. P., English J. H., **Laser Vaporization of Alloys—Laser-Induced Fluorescence of Heteronuclear Metal Clusters.** *J Chem Phys* 78(1):11-15, 1983.
- Boochand P., Grothaus J., Phillips J. C., **Broken Chemical Order and Phase-Separation in Ge_xSe_{1-x} Glasses.** *Sol St Comm* 45(2):183-185, 1983.
- Botineau J., Stolen R. H., **Effect of Polarization on Spectral Broadening in Optical Fibers.** *J Opt Soc* 72(12):1592-1596, 1982.
- Brand H., Cross M. C., **Explanation of Flow Dissipation in ³He-B (Letter).** *Phys Rev L* 49(26):1959, 1982.
- Bridges T. J., Bergman J. G., Chraplyvy, A. R., **Broad-Band Infrared Generation by Stimulated Raman-Scattering in Liquid Filled Fibers.** *P Soc Photo* 320:60-61, 1982.
- Bruck M. A., et al., **Extended X-Ray Absorption Fine-Structure (EXAFS) Spectroscopic Analysis of 6-Mercaptopurine Riboside Complexes of Platinum(II) and Palladium(II) (Review or Bibliog.).** *ACS Symp S* 1983(209):245-262, 1983.
- Capasso F., Tsang W. T., Bethea C. G., Hutchinson A. L., Levine B. F. **New Graded Band-Gap Picosecond Phototransistor.** *Appl Phys L* 42(1):93-95, 1983.
- Capasso F., Tsang W. T., Williams G. F., **New Very Low-Noise Multilayer and Graded-Gap Avalanche Photodiodes for the 0.8 to 1.8 μm Wavelength Region.** *P Soc Photo* 340:50-55, 1982.
- Celler G. K., Trimble L. E., **Microelectronics Applications of Deposited Si Films Recrystallized from the Melt.** *J Physique* 43(NC1):353-362, 1982.
- Celler G. K., Robinson M., Lischner D. J., **Seeded Recrystallization of Thick Polysilicon Films on Oxidized 3-in Wafers.** *Appl Phys L* 42(1):99-101, 1983.
- Cheng J., Lemons R. A., **Topography Induced Preferential Crystallization of Thermally Grown Silicon Dioxide Films.** *J Elec Mat* 12(1):61-69, 1983.
- Cowens J. W., Stevie F. A., Alderfer J. L., Hansen G. E., Pendyala L., Creaven P. J., **Synthesis and Identification of Derivatives of a Platinum Containing Complex.** *Int J Mass* 48(FEB):177-180, 1983.
- Chu S. N. G., **Elastic Interaction Between a Screw Dislocation and Surface Crack.** *J Appl Phys* 53(12):8678-8685, 1982.
- Craighead H. G., Howard R. E., Jackel L. D., Mankiewich P. M., **10-nm Linewidth Electron-Beam Lithography on GaAs.** *Appl Phys L* 42(1):38-40, 1983.
- Cross M. C., **Phase Dynamics of Convective Rolls.** *Phys Rev A* 27(1):490-498, 1983.
- Dahlberg S. C., **Photo-Voltage Studies of Organic Semiconductor Interfaces.** *Appl Surf S* 14(1):47-55, 1982.
- Ebeling K. J., Coldren L. A., Miller B. I., Rentschler J. A., **Single-Mode Operation of Coupled-Cavity GaInAsP-InP Semiconductor-Lasers.** *Appl Phys L* 42(1):6-8, 1983.
- English A. T., Kinsbron E., **Electromigration-Induced Failure by Edge Displacement in Fine-Line Aluminum-0.5-Percent Copper Thin-Film Conductors.** *J Appl Phys* 54(1):268-274, 1983.
- English A. T., Kinsbron E., **Electromigration Transport Mobility Associated with Pulsed Direct-Current in Fine-Grained Evaporated Al-0.5-Percent-Cu Thin-Films.** *J Appl Phys* 54(1):275-280, 1983.
- Flamm D. L., Wang D. N. K., Maydan D., **Multiple-Etchant Loading Effect and Silicon Etching in ClF₃ and Related Mixtures.** *J Elchem So* 129(12):2755-2760, 1982.

- Franey J. P., **A Novel System for Atmospheric Corrosion Experiments.** *Corros Sci* 23(1):1-8, 1983.
- Freeman R. R., Jopson R. M., Bokor J., Bucksbaum, P., **Laser Generation of Light in the Extreme Ultraviolet and Soft-X-Ray Regime.** *AIP Conf Pr*1982(94): 491-494, 1982.
- Gibbs H. M. et al., **Room Temperature Optical Bistability and Self-Defocusing in Semiconductor Etalons.** *Appl Phys B* 29(3):171-172, 1982.
- Guélin M., Cernicharo J., Linke R. A., **Detection of HC^{17}O^+ in Sagittarius 82.** *Astrophys J* 263(2):L89-L93, 1982.
- Hebard A. F., Nakahara S., **Structural Phase-Transitions of Indium Indium Oxide Thin-Film Composites.** *Appl Phys L* 41(12):1130-1132, 1982.
- Heffner W. R., Berreman D. W., **Switching Characteristics of a Bistable Cholesteric Twist Cell.** *J Appl Phys* 53(12):8599-8606, 1982.
- Heller A., Aharon-Shalom E., Bonner W. A., Miller B., **Hydrogen-Evolving Semiconductor Photo-Cathodes—Nature of the Junction and Function of the Platinum Group Metal Catalyst.** *J. Am Chem S* 104(25):6942-6948, 1982.
- Hellman M. Y., Bowmer T., Taylor G. N., **Determination of Gel Content and Percent Gel in Radiation-Cured Polyvinyl Chloride—Cross-linking Monomer Coatings by Combined GPC-LC Techniques.** *Macromolec* 16(1):34-38, 1983.
- Hutton R. S., Roth H. D., **Spin-Density Distribution and Geometric Isomerism in Triplet Carbenes.** *J Am Chem S* 104(26):7395-7399, 1982.
- Hwang J. C. M., Temkin H., Brennan T. M., Frahm R. E., **Growth of High-Purity GaAs-Layers by Molecular-Beam Epitaxy.** *Appl Phys L* 42(1):66-68, 1983.
- Jayaraman A., **Diamond Anvil Cell and High-Pressure Physical Investigations (Review or Bibliog.)** *Rev M Phys* 55(1):65-108, 1983.
- Jelinski L. W., Dumais J. J., Watnick P. I., Bass S. V., Shepherd L., **C-13-NMR of ABS in the Solid-State—Characterization and Relaxation Studies.** *J Pol Sc PC* 20(12):3285-3296, 1982.
- Johari G. P., **Glass-Transition and Molecular Motions in the Glassy State of a Nematic Liquid.** *Phil Mag B* 46(6):549-564, 1982.
- Kazarinov R. F., Nordland W. A., Wagner W. R., Temkin H., Manchon D. D., **Near-Equilibrium LPE Growth of $\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y}$ Lattice Matched to InP.** *J Cryst Gr* 60(2):235-238, 1982.
- Kelso S. M., Aspnes D. E., Pollack M. A., Nahory R. E., **Optical-Properties of $\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y}$ From 1.5 to 6.0 eV Determined by Spectroscopic Ellipsometry.** *Phys Rev B* 26(12):6669-6681, 1982.
- Kolodner P., Tyson J. A., **Remote Thermal Imaging with 0.7- μm Spatial-Resolution Using Temperature-Dependent Fluorescent Thin-Films.** *Appl Phys L* 42(1):117-119, 1983.
- Kramer L., Benjacob E., Brand H., Cross M. C., **Wavelength Selection in Systems Far From Equilibrium.** *Phys Rev L* 49(26):1891-1894, 1982.
- Kuo C. Y., Vieira M. M. F., Kerl R. J., Patel C. K. N., **Observation of Delta-V-4 Vibrational Overtone Transitions in Solid HCP Para-Hydrogen.** *Phys Rev L* 50(4):256-259, 1983.
- Lake G., **The Rotation of Elliptical Galaxies—An Application of the Theory of Tidal Torques.** *Astrophys J* 264(2):408-412, 1983.
- Lamola A. A., **Fluorescence Studies of Protoporphyrin Transport and Clearance.** *Act der-Ven*1982(S100):57-66, 1982.
- Lee T. P., **Recent Development in Light-Emitting-Diodes (LEDs) for Optical Fiber Communications-Systems.** *P Soc Photo* 340:22-31, 1982.
- Leung S. Y., Schumaker N. E., **Simulation of Slider-Induced Convection in Horizontal LPE Slider System.** *J Cryst Gr* 60(2):421-433, 1982.
- Lin C., Lee T. P., Burrus C. A., **Picosecond Frequency Chirping and Dynamic Line Broadening in InGaAsP Injection-Lasers Under Fast Excitation.** *Appl Phys L* 42(2):141-143, 1983.
- Lovinger A. J., **Microstructure and Unit-Cell Orientation in Alpha-Polypropylene.** *J Pol Sc PP* 21(1):97-110, 1983.
- Mayer T. M., Fisanick G. J., Eichelberger, T. S., **Deposition of Chromium Films by Multi-Photon Dissociation of Chromium Hexacarbonyl.** *J Appl Phys* 53(12): 8462-8469, 1982.

- Meiboom S., Sammon M., Brinkman W. F., **Lattice of Disclinations—The Structure of the Blue Phases of Cholesteric Liquid-Crystals.** *Phys Rev A* 27(1):438-454, 1983.
- Mills A. P., **Line-Shape Effects in the Measurement of the Positronium Hyperfine Interval.** *Phys Rev A* 27(1):262-267, 1983.
- Myers P. C., Linke R. A., Benson P. J., **Dense Cores in Dark Clouds. 1. Co Observations and Column Densities of High-Extinction Regions.** *Astrophys J* 264(2):517-537, 1983.
- Nassau K., **Material Dispersion Considerations For Infrared Fibers.** *P Soc Photo* 320:43-48, 1982.
- Obryan H. M., Yan M. F., **Second-Phase Development in Ba-Doped Rutile.** *J Am Ceram* 65(12):615-619, 1982.
- Odagaki T., Lax M., **Some Exact Properties of Hopping Conduction.** *Phys Rev B* 26(12):6480-6489, 1982.
- Oron M., Tamari N., **High-Power Single-Mode InGaAsP Lasers Fabricated by Single Step Liquid-Phase Epitaxy.** *Appl Phys L* 42(2):139-141, 1983.
- Patterson G. D., **Photon-Correlation Spectroscopy of Bulk Polymers (Review or Bibliog.).** *Adv Polym S* 48:125-159, 1983.
- Phillips J. C., **Reversible and Sensitizable Photostructural Transformations in Chalcogenide Alloy Glasses.** *J Non-Cryst* 51(3):301-304, 1982.
- Phillips J. C., **Spectroscopic and Morphological Structure of Tetrahedral Oxide Glasses (Review or Bibliog.).** *Sol St Phys* 37:93-171, 1982.
- Pleiner H., Brand H., **Light-Scattering in Nematic Liquid-Crystals in the Presence of Shear-Flow.** *J Phys Lett* 44(1):L23-L31, 1983.
- Rana V. V. S., Nelson T. J., Lecraw R. C., Blank S. L., **Preparation and Properties of V-Substituted Garnet-Films for Ion-Implanted 1.0-Micron Bubble-Devices with Improved High-Temperature Propagation.** *J Appl Phys* 53(12):9093-9097, 1982.
- Reagor B. T., Kelley D. F., Huchital D. H., Rentzepis P. M., **Kinetics of Intramolecular Electron-Transfer in Binuclear (NC)₅Fe^{II}-CN-Co^{III}(chelate) Complexes by Picosecond Absorption-Spectroscopy.** *J Am Chem S* 104(26):7400-7403, 1982.
- Reynolds C. L., Jr. and Tamargo M. C., **Liquid Phase Epitaxial Growth of (Al, Ga)As Double-Heterostructure Laser Material in a Sapphire Boat.** *J Appl Phys* 53(12):9217-9219, 1982.
- Roth H. D., Schilling M. L., Hutton R. S., Truesdale E. A., **Photoinduced Electron-Transfer Reactions—The Radical Cations of Methano-Bridged Paracyclophanes.** *J Am Chem S* 105(2):153-157, 1983.
- Raine G. P., Schaefer H. F., Haddon R. C., **The Dimers of Carbon-Monoxide and Carbon Monosulfide—Chemically Bound Triplet Electronic Ground-States.** *J Am Chem S* 105(2):194-198, 1983.
- Schmidt P. F., **Contamination-Free High-Temperature Treatment of Silicon or Other Materials.** *J Elchem So* 130(1):196-199, 1983.
- Schubert R., Tompkins H. G., **Effects of Exposure to Vapors of Saturated Salt Aqueous-Solutions.** *IEEE Compon* 5(4):425-430, 1982.
- Shah J., Pinczuk A., Stormer H. L., Gossard A. C., Wiegmann W., **Electric-Field Induced Heating of High Mobility Electrons in Modulation-Doped GaAs-Al-GaAs Heterostructures.** *Appl Phys L* 42(1):55-57, 1983.
- Shank C. V., Yen R., Hirlimann C., **Time-Resolved Reflectivity Measurements of Femtosecond-Optical-Pulse Induced Phase-Transitions in Silicon.** *Phys Rev L* 50(6):454-457, 1983.
- Snyder L. C. et al., **Computations to Model 3-Center Bonds in Hydrogenated Amorphous-Silicon.** *Phys Rev B* 26(12):6727-6733, 1982.
- Stavola M., Patel J. R., Kimerling, L. C., Freeland P. G., **Diffusivity of Oxygen in Silicon at the Donor Formation Temperature.** *Appl Phys L* 42(1):73-75, 1983.
- Stocker H. J., Aspnes D. E., **Surface Chemical-Reactions on In_{0.53}Ga_{0.47}As.** *Appl Phys L* 42(1):85-87, 1983.
- Stone J., Whalen M. S., **Index of Refraction Dispersion of Normal-Type and Paratype InP Between 0.95 and 2.0 eV.** *Appl Phys L* 41(12):1140-1142, 1982.
- Temkin H., Hwang J. C. M., **Undoped, Semi-Insulating GaAs-Layers Grown by Molecular-Beam Epitaxy.** *Appl Phys L* 42(2):178-180, 1983.

- Tumpkins H. G., Sharma S. P., **The Interaction of Imidazole, Senzimidazole and Related Azoles with a Copper Surface.** *Surf Int An* 4(6):261-266, 1982.
- Tsang W. T., Dixon M., Dean B. A., **The Characterization and Functional Reliability of 45 Mbit/s Optical Transmitters Containing MBE-Grown Lasers.** *IEEE J Q El* 19(1):59-62, 1983.
- Tsui D. C., Gossard A. C., Dolan G. J., **Radiation Effects on Modulation-Doped GaAs-Al_xGa_{1-x}As Heterostructures.** *Appl Phys L* 42(2):180-182, 1983.
- Tung R. T., Gibson J. M., Poate J. M., **Formation of Ultrathin Single-Crystal Silicide Films on Si—Surface and Interfacial Stabilization of Si-NiSi₂ Epitaxial Structures.** *Phys Rev L* 50(6):429-432, 1983.
- Vandenberg J. M., Hamm R. A., **An Insitu X-Ray Study of Phase Formation in Cu-Al Thin-Film Couples.** *Thin Sol Fi* 97(4):313-323, 1982.
- Varma G. M., Dynes R. C., Banavar J. R., **Thermally Created Tunnelling States in Glasses (Letter).** *J Phys C* 15(35):1221-1225, 1982.
- Walker L. R., **Interpretation of the Fe₅₇ Isomer-Shift (Citation Classics).** *CC/Phy Chem*1983(7):20, 1983.
- Wertheim G. K., Phillips J. C., Bridenbaugh P. M., **A Limit on the Site Inequivalence in G-As₂Se₃.** *Sol St Comm* 45(2):179-181, 1983.
- White J. C., Henderson D., **Tuning and Saturation Behavior of the Anti-Stokes Raman Laser.** *Optics Lett* 8(1):15-17, 1983.
- Wilson T., Sheppard C. J., **Imaging with Finite Values of Fresnel Number.** *J Opt Soc* 72(12):1639-1641, 1982.
- Wolfe A., **ULF Geomagnetic Power Near L = 4. 7. A Conjugate Area Study of Power Controlled by Solar-Wind Quantities.** *J Geo R-S P* 87(NA12):425-431, 1982.
- Wrobel J., Takahashi K., Lannoy G., Cook J. M., Bertz S. H., Honkan V., **Stereocontrolled Synthesis of (+/-)-Modhephene via the Weiss Reaction (Letter).** *J Org Chem* 48(1):139-141, 1983.
- Yan M. F., Rhodes W. W., **Effects of Cation Contaminants in Conductive TiO₂ Ceramics.** *J Appl Phys* 53(12):8809-8818, 1982.
- Zussman A., Eger D., Oron M., Szapiro S., Shachna A., Hans B., **Lung-Wavelength PbSnTe-PbTeSe Lattice-Matched Single-Heterostructure Lasers Grown by LPE.** *IEE Proc-I* 129(6):203-208, 1982.

SOCIAL AND LIFE SCIENCES

- Clark W. C., Carroll J. D., Yang J. C., Pruzansky S., **Perceptual Dimensions of Thermal Pain—A Multidimensional-Scaling Approach (Review or Bibliog.).** *Adv Pain R* 5:871-876, 1982.
- Desarbo W. S., Hausman R. E., Lin S., Thompson W., **Constrained Canonical Correlation.** *Psychometri* 47(4):489-516, 1982.
- Fishburn P. C., **Foundations of Risk Measurement. 2. Effects of Gains on Risk.** *J Math Psych* 25(3):226-242, 1982.
- Goddard C. T., **Debunking the Learning-Curve.** *IEEE Compon* 5(4):328-335, 1982.
- Julesz B., Kropfl W., **Binocular Neurons and Cyclopean Visually Evoked-Potentials in Monkey and Man.** *Ann Ny Acad* 388(Jun):37-44, 1982.
- Rosenbaum D. A., Kornblum S., **A Priming Method for Investigation the Selection of Motor-Responses.** *Act Psychol* 51(3):223-243, 1982.
- Rosenthal N. E., Lewy A. J., Wehr T. A., Kern H. E., Goodwin F. K., **Seasonal Cycling in a Bipolar Patient.** *Psychiat R* 6(1):25-31, 1983.
- Starr S. J., Thompson C. R., Shute S. J., **Effects of Video Display Terminals on Telephone Operators.** *Human Fact* 24(6):699-711, 1982.

SPEECH/ACOUSTICS

- Hurd H. L., **Clipping Degradation in a System Used to Detect Time-Periodic Variance Fluctuations.** *J Acoust So* 72(6):1827-1830, 1982.

Karim M. R., An Investigation of Delta Modulation of Speech for Mobile Radio. IEEE Veh T 31(4):158-165, 1982.

Umeda N., Kahn D., Frequency of Occurrence of Two-Word and Three-Word Sequences in English (Letter). J Acoust So 72(6):2031-2033, 1982.

CONTENTS, JULY-AUGUST 1983

Part 1

On the Start-up Problem in Digital Echo Cancelers

J. Salz

Sample Reduction and Subsequent Adaptive Interpolation of Speech Signals

R. Steele and F. Benjamin

Practical Design Considerations for Coupled-Single-Amplifier-Biquad Active Bandpass Filters

J. Tow

Modal Structure of an MCVD Optical Waveguide Fiber

A. Carnevale and U. C. Paek

Asymptotic Analysis of a Queueing Model With Bursty Traffic

D. Y. Burman and D. R. Smith

MODERNIZATION OF THE SURBURBAN ESS

Overview: Evolution of the Suburban ESS

T. E. Grassman and J. E. Yates

Adding Data Links to an Existing ESS

C. E. Ishman, R. B. Sanderson, L. M. Taff, D. P. Truax, and
C. T. Tulloss

Hosting the No. 10A Remote Switching System

D. W. Brown, J. J. Driscoll, F. M. Lax, M. W. Saad, and
J. G. Whitemyer

Billing and Measurements Modernization

J. P. Ludwig and D. A. Ward

Part 3

HUMAN FACTORS AND BEHAVIORAL SCIENCE

Introduction

E. E. Sumner

History and Methods

A Brief History of Applied Behavioral Science at Bell Laboratories

B. L. Hanson

Methods for Field Testing New Telephone Services

D. J. Eigen

Characteristics of Human Performance

Textons, The Fundamental Elements in Preattentive Vision and Perception of Textures

B. Julesz and J. R. Bergen

Central Control of Movement Timing

D. A. Rosenbaum

Experiments on Quantitative Judgments of Graphs and Maps

W. S. Cleveland, C. S. Harris, and R. McGill

Retrospective Reports Reveal Differences in People's Reasoning

D. E. Egan

New Technological Demands

Human Factors Engineering for the Loop Plant

J. Donegan and D. N. Koppes

Effects of Shape and Size of Knobs on Maximal Hand-Turning Forces Applied by Females

G. Kohl

Human Factors Comparison of Two Fiber-Optic Continuous-Groove Field-Repair Splicing Techniques

L. M. Paul

Performance in Locating Terminals on a High-Density Connector

L. E. Flamm

Membrane Keyboards and Human Performance

K. M. Cohen Loeb

Interface Design

Statistical Semantics: Analysis of the Potential Performance of Keyword Information Systems

G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais

On Abbreviating Command Names

L. A. Streeter, J. M. Ackroff, and G. A. Taylor

Designing and Evaluating Standard Instructions for Public Telephones

C. J. Karhan, C. A. Riley, and M. S. Schoeffler

A Study of the Match Between the Stylistic Difficulty of Technical Documents and the Reading Skills of Technical Personnel

E. U. Coke and M. E. Koether

Toward Bell System Applications of Automatic Speech Recognition

J. E. Holmgren

New Functions for Technology

The *Unix*TM Writer's Workbench Software: Philosophy

L. T. Frase

The *Unix*TM Writer's Workbench Software: Rationale and Design

N. H. Macdonald

The *Unix*TM Writer's Workbench Software: Results of a Field Study

P. S. Gingrich

THE BELL SYSTEM TECHNICAL JOURNAL is abstracted or indexed by *Abstract Journal in Earthquake Engineering, Applied Mechanics Review, Applied Science & Technology Index, Chemical Abstracts, Computer Abstracts, Current Contents/Engineering, Technology & Applied Sciences, Current Index to Statistics, Current Papers in Electrical & Electronic Engineering, Current Papers on Computers & Control, Electronics & Communications Abstracts Journal, The Engineering Index, International Aerospace Abstracts, Journal of Current Laser Abstracts, Language and Language Behavior Abstracts, Mathematical Reviews, Science Abstracts (Series A, Physics Abstracts; Series B, Electrical and Electronic Abstracts; and Series C, Computer & Control Abstracts), Science Citation Index, Sociological Abstracts, Social Welfare, Social Planning and Social Development, and Solid State Abstracts Journal*. Reproductions of the Journal by years are available in microform from University Microfilms, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.



Bell System