

THE JULY-AUG. 1979 VOL. 58, No. 6, PART 2

BELL SYSTEM

TECHNICAL JOURNAL



ISSN0005-8580

General Articles

C. Scagliola	Evaluation of Adaptive Speech Coders Under Noisy Channel Conditions	1369
N. F. Maxemchuk and J. A. Stuller	An Adaptive Intraframe DPCM Codec Based Upon Nonstationary Image Model	1395
N. F. Maxemchuk and J. A. Stuller	Reduction of Transmission Error Propagation in Adaptively Predicted, DPCM Encoded Pictures	1413
D. H. Smithgall	Application of Optimization Theory to the Control of the Optical Fiber Drawing Process	1425
H. S. Witsenhausen	On the Structure of Real-Time Source Coders	1437
A. H. Bobeck, S. L. Blank, A. D. Butherus, F. J. Ciak, and W. Strauss	Current-Access Magnetic Bubble Circuits	1453
	Contributors to This Issue	1541
	Papers by Bell Laboratories Authors	1545
	Errata	1549
R. A. Semplak	B.S.T.J. Brief: A 30-GHz Scale-Model, Pyramidal, Horn Reflector Antenna	1551

THE BELL SYSTEM TECHNICAL JOURNAL

ADVISORY BOARD

D. E. PROCKNOW, *President*, *Western Electric Company, Incorporated*

I. M. ROSS, *President*, *Bell Telephone Laboratories, Incorporated*

W. M. ELLINGHAUS, *President*, *American Telephone and Telegraph Company*

EDITORIAL COMMITTEE

D. GILLETTE, *Chairman*

S. J. BARBERA

A. A. PENZIAS

A. G. CHYNOWETH

W. B. SMITH

T. H. CROWLEY

B. E. STRASSER

I. DORROS

I. WELBER

H. B. HEILIG

M. P. WILSON

EDITORIAL STAFF

G. E. SCHINDLER, JR., *Editor*

J. B. FRY, *Associate Editor*

H. M. PURVIANCE, *Art Editor*

B. G. GRUBER, *Circulation*

THE BELL SYSTEM TECHNICAL JOURNAL is published monthly, except for the May-June and July-August combined issues, by the American Telephone and Telegraph Company, C. L. Brown, Chairman and Chief Executive Officer; W. M. Ellinghaus, President; W. G. Burns, Vice President and Treasurer; F. A. Hutson, Jr., Secretary. Editorial enquiries should be addressed to the Editor, The Bell System Technical Journal, Bell Laboratories, 600 Mountain Ave., Murray Hill, N.J. 07974. Checks for subscriptions should be made payable to The Bell System Technical Journal and should be addressed to Bell Laboratories, Circulation Group, Whippany Road, Whippany, N.J. 07981. Subscriptions \$20.00 per year; single copies \$2.00 each. Foreign postage \$1.00 per year; 15 cents per copy. Printed in U.S.A. Second-class postage paid at New Providence, New Jersey 07974 and additional mailing offices.

© 1979 American Telephone and Telegraph Company

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.

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 58

July-August 1979

Number 6, Part 2

Copyright © 1979 American Telephone and Telegraph Company. Printed in U.S.A.

Evaluation of Adaptive Speech Coders Under Noisy Channel Conditions

By C. SCAGLIOLA

(Manuscript received June 20, 1978)

An experiment has been performed in which the digital transmission of speech coded by adaptive differential PCM was simulated under noisy channel conditions. The experiment was done with two aims: (i) to get information on the subjective effect of channel errors and the influence of various design parameters on the speech quality under various conditions and (ii) to find objective measures for predicting the overall quality of the processed speech over a wide range of circuit conditions. The subjective results show that, for a speech transmission through a channel with bit error probability up to 1/256, best results can be obtained with a slow error recovery, associated with fast quantizer adaptation. The use of slow error recovery and slow quantizer adaptation is preferable for channels with very high bit error rates, like 1/32. Overall subjective quality is well predicted by the sum of two terms: (i) an objective performance measure of the noise present on the output signal, disregarding any effect of level mismatching due to the sensitivity of the adaptation algorithms to channel errors and (ii) a measure of the level mismatching which takes into account both the average gain on the output signal and its fluctuation in time. The best prediction scores are achieved by three newly defined objective performance measures, two-level compensated segmental SNR's, and a spectral signal-to-distortion ratio.

I. INTRODUCTION

The design of digital waveform coders for speech communications must face the inevitable presence of channel errors. Adaptive coders, like ADPCM (adaptive differential PCM), in which the adaptation of the

quantizer step-size is derived from the transmitted binary stream and no error-protected side information is sent to the receiver, may be particularly sensitive to this problem. In fact, a single channel error may cause a multiplicative offset between the signal level at the receiver and that at the transmitter. This offset may persist indefinitely if no error dissipation mechanism is provided.

Recently, some algorithms of quantizer adaptation have been developed that make the effect of a single transmission error die out over time, so the transmitter and receiver can resynchronize their step-size estimates.^{1,2} The possibility of obtaining such results is physically due to the fact that these algorithms have an imperfect adaptation: the step size increases more quickly and decreases more slowly for low input levels than for high ones. In this way, the step size is overestimated for low input levels and underestimated for high ones, thereby reducing the dynamic range of the coder. Dynamic range and error dissipation rate vary inversely, and the designer has to balance between them.

In the case of speech transmission, the choice of the appropriate design parameters must be based on a precise evaluation of the subjective quality of the coded speech. Use of the conventional long-term signal-to-noise ratio as an estimator of the subjective quality would be, in this instance, completely misleading, at least because an offset in the signal amplitude between input and output, due to an offset in step sizes caused by an error, will be reflected in a noticeable squared difference between the two waveforms, while it may not be subjectively disturbing.

To study the subjective performance of ADPCM coders operating under both error-free and noisy channel conditions, an experiment has been conducted, as summarized in Section II. The following three sections provide a brief description of the coding method, the definition of several objective performance measures, and the description of the experimental design and testing procedure. Sections VI and VII provide analyses of subjective and objective measurement data. In Section VIII, the results of the previous two sections are discussed and a physical interpretation is given of the principal findings on quality prediction.

II. OVERVIEW OF THE EXPERIMENT

The experiment included 12 different ADPCM coding schemes, which comprised all combinations of two bit rates, two adaptation time constants, and three error dissipation rates. These systems processed a total of 288 speech samples from four talkers (two male and two female), at two different power levels (24 dB apart) and with three different probabilities of independent errors on the channel.

Twenty listeners rated the quality of the processed speech samples

on a scale from 1 to 9. The odd values were associated with the adjectives: unsatisfactory, poor, fair, good, excellent. In addition to the subjective data, a fairly large number of objective performance measures were also taken on the processed speech samples.

The aims of the experiment included the study of:

- (i) The influence on speech quality of the above design parameters.
- (ii) The optimum combination of parameters for a given error probability.
- (iii) The objective measures or combinations of objective measures which are good predictors of speech quality even under noisy channel conditions.

The principal conclusions drawn from the analyses of the subjective and objective data are:

- (i) Contrary to a common feeling, a very slow error dissipation is sufficient to ensure good robustness of the ADPCM coder to error rates even much higher than those encountered in a normal telephone connection. When no recovery mechanism is provided, a fairly slow adaptation makes the system not very sensitive to channel errors in the range of error rates typical of a telephone connection.
- (ii) For speech transmission through a channel with bit error rate up to 1/256, best results can be obtained with a very slow error dissipation, associated with fast quantizer adaptation; when the slow error dissipation is associated instead with a slow adaptation, the system becomes fairly robust to very high error rates, like 1/32, at the expense of a slight quality deterioration at low error rates.
- (iii) Good predictors of subjective quality were found to be two-segmental SNR measures in which a compensation of the level mismatching between input and output was performed on a frame-by-frame basis. The combination of any of these measures with two separate measures of level mismatching further improved the prediction accuracy.

III. ROBUST ADPCM SYSTEM: A BRIEF DESCRIPTION

Figure 1 is a block diagram of the ADPCM coder-decoder used in the experiment. The predictor is a second-order transversal filter, with tap coefficients 1 and -0.5 . The step size $\Delta(k)$ is adapted according to the robust algorithm described in Ref. 1, which permits synchronizing the step-size estimates at transmitter and receiver, after a transmission error occurs, during a period of error-free transmission

$$\Delta(k + 1) = \Delta^\beta(k) \cdot M(I(k)), \quad (1)$$

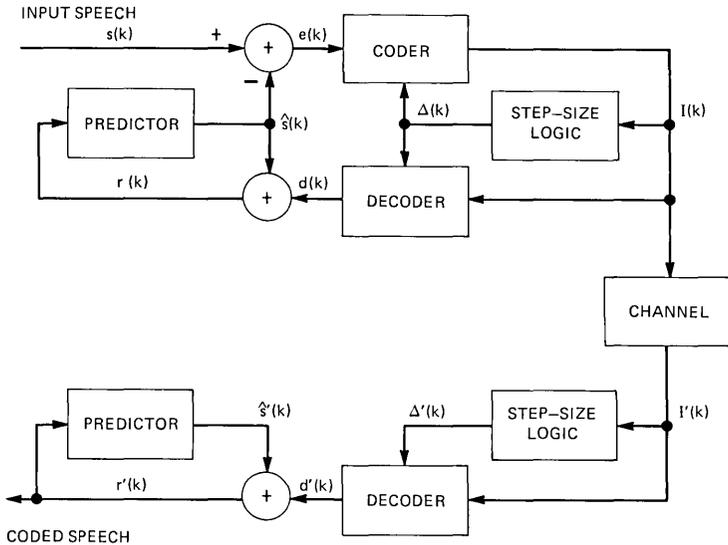


Fig. 1—Block diagram of ADPCM coder-decoder.

where the exponent β , $0 < \beta < 1$, is the decay constant and $M(I(k))$ is the step-size multiplier at time k . The multiplier $M(I(k))$ depends only on the actual code word $I(k)$ and assumes $N = 2^{B-1}$ distinct values $\{M_1, M_2, \dots, M_N\}$, where B is the number of bits used to encode the prediction error $e(k)$.

The decay speed is shown to be independent of the actual code word transmitted and also of the values of the multipliers, being only a function of the decay constant β .¹ When $\beta = 1$, the decay speed is zero and the decay time infinite. With $\beta < 1$, the decay speed increases, but at the same time the dynamic range decreases.

It has been shown that the loading factor (ratio between range of the quantizer and rms quantizer input) is a constant as a function of the input level if $\beta = 1$, but it is a decreasing function of the input level if $\beta < 1$.¹ The form of their relationship is almost linear, with slope approximately inversely proportional to $\log(M_N/M_1)/(1 - \beta)$, as indicated by Fig. 6 of Ref. 1 and as recently proved theoretically by D. Mitra.³ Therefore, with a small β and a small ratio between the maximum and minimum multipliers, the coder will produce more granular noise for low input level and more overload distortion for high levels.

The values of the multipliers play another important role in the overall performance of the coders under noisy channel conditions. In fact, they determine the magnitude of the initial offset after a single

transmission error. In the worst case, the offset is given by

$$\left| \frac{\Delta'(k)}{\Delta(k)} \right|_{\max} = \frac{M_N}{M_1}. \quad (2)$$

In the present experiment, the multipliers M_i were related by the linear relationship:

$$M_i = [\alpha + C(1 - \alpha)(i - 0.5)]\hat{\Delta}^{(1-\beta)}, \quad (3)$$

where $\hat{\Delta}$ is the step size that gives optimum performance at the desired nominal input level. This was fixed at -21 dBm, i.e., 27 dB under the saturation threshold for the signal that in the internal 16-bit computer representation is 32767. This relationship was chosen because, for $\beta = 1$, the adaptation algorithm coincides with the magnitude estimation algorithm described by Castellino et al.,⁴ and for this algorithm more information about the subjective effects of its parameters are available.⁵ In eq. (3), C essentially determines the mixture of granular noise and clipping distortion in the decoded prediction error at the nominal level. The parameter α controls mainly the speed of adaptation and hence the ratio between maximum and minimum multipliers.

IV. OBJECTIVE PERFORMANCE MEASURES

Several objective performance indices were measured for each utterance in the experiment. The speech samples used as input to the coders were low-pass filtered at 3.4 kHz before being sampled and converted into digital form by a 16-bit A/D converter operating at 8-kHz sampling rate. Figure 2 is a block diagram of the simulated circuit arrangement for performing coding and measurements. A filter after the ADPCM decoding limits the bandwidth of the output speech as in a real situation. A secondary path provides the reference signal $s_0(k)$ with which the filtered output $r_0(k)$ is compared to compute the objective performance measures. With two identical filters in the main

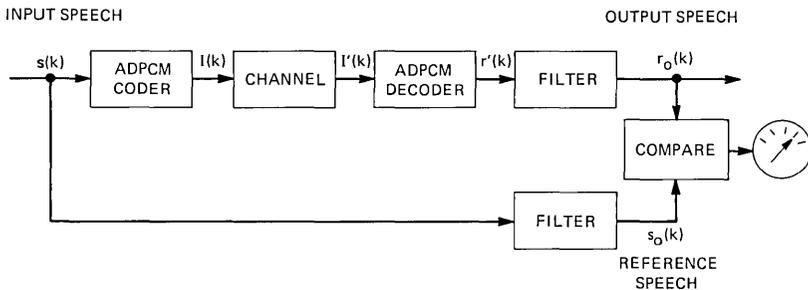


Fig. 2—Block diagram of simulated circuit arrangement for coding and measurements.

and reference paths, only the distortion introduced by the coder is measured. The filters are 5th-order elliptic low-pass with 3.4-kHz cutoff frequency, 0.25-dB in-band ripple, and at least 40 dB stopband attenuation.

The measures are classified in the two categories of time domain measures and frequency domain measures. Frequency-weighted signal-to-noise ratios are included in the first category because they rely strongly on the exact time synchronization of the two waveforms and on the absence of phase distortion.

4.1 Time domain measures

4.1.1 Long-term signal-to-noise ratio (SNR)

$$\text{SNR} = 10 \log \frac{\sum_k s_o^2(k)}{\sum_k [s_o(k) - r_o(k)]^2}, \quad (4)$$

where k ranges over all the samples of the utterance. SNR is the ratio between the long-term signal energy and the long-term noise energy, the noise being defined as the difference between reference signal $s_o(k)$ and output signal $r_o(k)$.

4.1.2 Segmental signal-to-noise ratio (SNR_{seg})

Here the utterance is divided into adjacent segments of J samples each, and the signal-to-noise ratio in each segment is measured in decibels. The noise is still defined as the difference between corresponding samples of reference and output speech. The segmental SNR is the average of these measures over the M segments of the utterance.

$$\text{SNR}_{\text{seg}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\sum_{j=1}^J s_o^2(j + mJ)}{\sum_{j=1}^J [s_o(j + mJ) - r_o(j + mJ)]^2}. \quad (5)$$

In this experiment, $J = 128$, corresponding to 16 ms segments. This measure, proposed by Noll,⁶ was recently found to correlate very nicely with subjective ratings of ADPCM-coded speech.⁵ A very important feature added to this basic formula consists in discarding from the computation those segments in which the signal power is below -54 dBm. This threshold, whose value was found to be appropriate for high quality speech,⁷ was introduced to avoid a slight idle channel noise having an unduly great negative weight in the overall performance measure. This is done also in all the following time domain measures.

4.1.3 Compensated signal-to-noise ratio (SNR_{com})

This measure was specifically formulated to compensate for the level variations that may occur under channel error conditions when the coder has a slow error dissipation. The difference between reference and output signals due to level offset should not be measured in fact as noise.

To compensate for these level variations, let us formulate the coding process in the m th segment as composed of an amplification of the input signal, the addition of an uncorrelated random noise, and a possible dc component. Therefore, the output process in the m th segment can be written as

$$r_o(k) = g(m)s_o(k) + q(k) + Q. \quad (6)$$

This coincides with the simple linear regression model of $r_o(k)$ on $s_o(k)$. Therefore, the gain factor $g(m)$ is the slope of the regression line of the output on the reference signal in the m th segment, and the noise term $q(k)$ is the minimum error made in predicting $r_o(k)$ from $s_o(k)$.⁸

Let us define the signal-to-noise ratio in the m th segment as the ratio between the variances of $g(m) \cdot s_o(k)$ and $q(k)$. This can be shown to be only a function of the correlation coefficient $\rho(m)$ between the reference and output signals in the m th segment:⁸

$$\rho(m) = \frac{JS_{sr}(m) - S_s(m)S_r(m)}{\sqrt{[JS_{ss}(m) - S_s^2(m)][JS_{rr}(m) - S_r^2(m)]}}, \quad (7)$$

where $S_x(m)$ indicates the summation of $x(j)$ and $S_{xy}(m)$ the summation of $x(j) \cdot y(j)$ over the J samples of the m th segment.

Averaging in decibels across the M segments in the utterance, the compensated SNR turns out to be:

$$\text{SNR}_{\text{com}} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{\rho^2(m)}{1 - \rho^2(m)}. \quad (8)$$

4.1.4 Average gain (G)

This is the average in decibels of the gain factor $g(m)$ defined before, across the M segments. From simple linear regression analysis, the gain $g(m)$ is:⁸

$$g(m) = \frac{JS_{sr}(m) - S_s(m)S_r(m)}{JS_{ss}(m) - S_s^2(m)}. \quad (9)$$

The average gain, which is an indication of how much the output level was increased or decreased on the average, with respect to the input level, is:

$$G = \frac{1}{M} \sum_{m=0}^{M-1} 20 \log g(m). \quad (10)$$

4.1.5 Gain fluctuation (σ_g)

This is simply the standard deviation of the gain $g(m)$, measured in decibels, across the M segments. It is a measure of how much the output level fluctuates owing to transmission errors.

$$\sigma_g = \left[\frac{1}{M} \sum_{m=0}^{M-1} (20 \log g(m))^2 - G^2 \right]^{1/2}. \quad (11)$$

4.1.6 Maximum signal-to-noise ratio (SNR_{\max})

An alternate way of compensating for the level mismatching between reference and output speech signals was found by defining the noise in the m th segments as

$$\epsilon(k) = s_o(k) - \hat{s}_o(k), \quad (12)$$

where

$$\hat{s}_o(k) = a_1(m) \cdot r_o(k) + a_o \quad (13)$$

is the least-square estimate of $s_o(k)$ based on $r_o(k)$.

The ratio between the variances of the reference signal and the minimum estimation error $\epsilon(k)$ (hence the name "maximum SNR") is again only a function of the correlation coefficient $\rho(m)$, defined by eq. (7).

Averaging again in decibels across the M segments in the utterance, the maximum SNR is

$$\text{SNR}_{\max} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{1}{1 - \rho^2(m)}. \quad (14)$$

It is readily seen that SNR_{\max} is always greater than SNR_{com} and that the two measures give essentially different results only for low-quality coding conditions.

4.1.7 Frequency-weighted, segmental, signal-to-noise ratios

This term indicates a fairly large class of measures. In these measures, the frequency axis is partitioned into many bands, usually nonuniform, the reference and output spectra are compared, some performance measure is then computed over each band, and these measures are averaged across the bands. In the measures described below, the spectra are computed over 256 points (32 ms). The segmental measures are obtained by averaging the measures taken every 128 samples (16 ms).

Three measures are reported here. They are described more in Refs. 7 and 9. The partitioning of the frequency axis is effected in those three cases according to the 16 classical articulation bands.¹⁰

$$\text{SNRF}_1 = \frac{1}{M} \sum_{m=0}^{M-1} \left[\frac{1}{16} \sum_{j=1}^{16} 10 \log \frac{S_j(m)}{N_j(m)} \right] \quad (15)$$

$$\text{SNRF}_2 = \frac{1}{M} \sum_{m=0}^{M-1} \frac{\sum_{j=1}^{16} L_j(m) 10 \log \frac{S_j(m)}{N_j(m)}}{\sum_{j=1}^{16} L_j(m)} \quad (16)$$

$$\text{SNRF}_5 = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{1}{1 + \sum_{j=1}^{16} \frac{N_j(m)}{S_j(m)}}, \quad (17)$$

where $S_j(m)$ is the energy of the reference signal $s_0(k)$ in the j th frequency band in the frame m and $N_j(m)$ is the corresponding noise energy. The noise is again defined as the difference between reference and output signals, the latter being preventively divided by the gain $g(m)$ previously defined to compensate for level fluctuations. However $g(m)$ is computed on the 256-point analysis window. In eq. (16), a “loudness weighting” has been introduced. The term $L_j(m)$ corresponds approximately to the subjective loudness in band j , and is computed as

$$L_j(m) = \int_{\text{band } j} |S_r(f)|^{1/2} df, \quad (18)$$

where $S_r(f)$ is the spectrum of the gain-compensated output speech $r_0(k)/g(m)$.

4.2 Frequency domain measures

All the spectral measures here presented and used in the experiment are based on the concept of linear prediction or inverse filtering.¹¹ The speech signal is represented by the p th order autoregressive model:

$$s(k) = \sum_{i=1}^p a_i s(k-i) + u(k), \quad (19)$$

where $u(k)$ is the white spectrum excitation function and the a_i 's are the coefficients of the inverse filter

$$A_s(z) = 1 - \sum_{i=1}^p a_i z^{-i}. \quad (20)$$

The coefficients a_i 's are computed to minimize the residual power of the signal at the inverse filter output.

In this paper, the dissimilarity between the spectra of reference and output speech in a given frame is computed essentially by comparing the residual powers of the signals $s_0(k)$ and $r_0(k)$ filtered by the inverse filters $A_s(z)$ and $A_r(z)$, derived from the same two signals. Four residual powers can be computed in the m th signal frame:

- (i) $P_e(m)$ obtained passing $s_0(k)$ through $A_s(z)$.
- (ii) $P_d(m)$ obtained passing $r_0(k)$ through $A_s(z)$.
- (iii) $P_e'(m)$ obtained passing $r_0(k)$ through $A_r(z)$.
- (iv) $P_d'(m)$ obtained passing $s_0(k)$ through $A_r(z)$.

Four objective measures based on these concepts are presented in the following paragraphs.

4.2.1 LPC distance measure (D_1)

This measure, proposed by Itakura,¹² is also called log likelihood ratio. The distance between output and reference speech in the m th frame is defined as

$$D_1 = \ln \frac{P_d(m)}{P_e'(m)}. \quad (21)$$

It can be shown that D_1 can be expressed in terms of spectral differences between the LPC models of the two frames of speech.¹³ Moreover, it results that the spectral difference is most heavily weighted in the peaks of the input speech smoothed spectrum, i.e., in the speech formants.

Interchanging the roles of reference and output speech, a different log likelihood ratio is obtained:

$$D_2 = \ln \frac{P_d'(m)}{P_e(m)}, \quad (22)$$

which has the same basic properties as D_1 .

The measure used here is actually the arithmetic mean of D_1 and D_2 , averaged across the M segments of the utterance:

$$D_I = \frac{1}{2M} \sum_{m=0}^{M-1} \left[\ln \frac{P_d(m)}{P_e'(m)} + \ln \frac{P_d'(m)}{P_e(m)} \right]. \quad (23)$$

4.2.2 Bharucha index (D_B)

This index is again a distance measure, similar to the log likelihood ratio. It has been formulated by Bharucha¹⁴ and its definition is also

reported in Ref. 7. The basic idea is that of measuring the noise introduced by the coder by “notching out” the speech spectrum by means of a time-varying linear filter, whose transfer function is matched to the inverse of the short-term spectral envelope. The quality index proposed by Bharucha is essentially the average increase in the residual power at the output of the “notch filter,” due to coding:

$$D_B = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \left[\frac{P_s(m)}{P_r(m)} \cdot \frac{P_d(m)}{P_e(m)} \right], \quad (24)$$

where $P_s(m)$ and $P_r(m)$ are the powers of the reference and output speech in the m th segment and provide the appropriate scaling in the measure. It can be shown that an uncorrelated noise component in the output speech is inversely weighted, frame by frame, by the smoothed LPC spectrum of the input speech signal.^{7,14} Therefore, the noise has more weight in those frequency bands where the signal energy is low; this is probably in conformity with subjective noise evaluation.

4.2.3 Spectral signal-to-distortion ratios

Following the same basic idea of the Bharucha index, two other measures were derived in the form of signal-to-distortion ratios. In fact, they are measured in decibels, and they increase with increasing quality, like the time domain SNRS. With the first SDR, a distortion power is defined as the difference between the prediction error powers $P_d(m)$ and $P_e(m)$ defined above.

Before taking the difference, however, the term $P_d(m)$ is multiplied by $P_s(m)/P_r(m)$ to compensate for level differences between input and output.

$$\text{SDR}_1 = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{P_s(m)}{P_d(m) \cdot [P_s(m)/P_r(m)] - P_e(m)}. \quad (25)$$

In the second SDR measure, the difference between the signal-to-prediction error ratios in decibels that is averaged to give D_B is instead computed relative to the input signal-to-prediction error ratio, and this new ratio is again averaged in decibels across the segments of the utterance.

$$\text{SDR}_2 = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log \frac{10 \log[P_s(m)/P_e(m)]}{10 \log[P_s(m)/P_e(m)] \cdot [P_d(m)/P_r(m)]}. \quad (26)$$

In this experiment, D_B , SDR_1 , and SDR_2 are computed with an analysis window of 160 samples (20 ms) that was shifted by 128 samples (16 ms) every frame. The inverse filters for computing $P_e(m)$ and $P_d(m)$ were always of the 20th order.

V. EXPERIMENTAL DESIGN AND PROCEDURE

5.1 Circuit conditions

The choice of the experimental design variables was dictated by the criterion of a broad quality range and of each value the variables assumed having caused an effect that could be perceived for at least some combination of the other variables.

Two bit/sample values were chosen, $B = 3$ and $B = 4$, and three values of the decay constant, $\beta = 1$, $\beta = 255/256$, and $\beta = 63/64$.

The condition $\beta = 1$ was included because it should give the broadest dynamic range and because it makes the ADPCM system identical to the earlier schemes.^{4,15}

For the step-size multipliers, given by formula (3), C was kept constant for each bit rate and equal to the values 0.65 and 0.41 that were found to be subjectively optimum for $B = 3$ and $B = 4$, respectively.⁵ The parameter α was given two values, 0.75 and 0.96875. The first value produces a rather fast adaptation, with a corresponding time constant of 0.5 ms. With $\alpha = 0.75$, the ratio M_N/M_1 is about 4 dB and 5.5 dB for $B = 3$ and $B = 4$, respectively. The second value of α produces a rather slow adaptation, the corresponding time constant being 4 ms. The ratio M_N/M_1 is much smaller, about 0.5 dB and 0.7 dB for 3- and 4-bit/sample, respectively.

Two different input levels, 24 dB apart, were used for every combination of the other conditions. They were $L = -33$ dBm and $L = -9$ dBm, i.e., symmetrical around the nominal input level -21 dBm for which the coder was designed to have optimum performance. The level -21 dBm was not included to keep the dimension of the experiment within reasonable limits of feasibility.

Finally the channel was characterized by three different probabilities of independent errors, $P(e) = 0$, $P(e) = 1/256$, $P(e) = 1/32$.

Summarizing, the experiment included 72 conditions which comprised all the combinations of $B = 3, 4$; $\alpha = 0.75, 0.96875$; $\beta = 1, 255/256, 63/64$; $L = -33, -9$; $P(e) = 0, 1/256, 1/32$.

5.2 Preparation of stimuli

Each experimental condition was simulated four times, using as input signals four sentences spoken by two male and two female talkers. Each talker spoke into a high-quality dynamic microphone, while seated in a sound-proof booth. The digital recordings had been generated by low-pass filtering the amplified microphone signal at 3.4 kHz, and then sampling and converting it into digital form by a 16-bit A/D converter operating at 8-kHz sampling rate.

For simulating the different input levels, the sentences, all previously adjusted to the same mean power level of -21 dBm, were multiplied by a constant factor at the coder input, and then divided by the same

factor at the output. In this way, each processed sentence was listened at the same level, unless channel errors and/or slope overload of the coder had caused output level variations.

From the 288 simulations, two analog test tapes were generated, each containing in a different random order two simulations of each experimental condition, one with a male and one with a female talker. For each talker, 18 different sentences read from a different phonetically balanced list were used, so that in each tape the same sentence appeared only twice.

5.3 Testing procedure

Twenty paid subjects (10 for each tape), all students from junior and senior classes of local high schools, judged the 288 stimuli. They listened to the processed speech binaurally over Pioneer SE 700 earphones at a nominal level of 80 dB_{SPL}, while seated in a double-walled sound booth. As pointed out before, the level of individual sentences varied according to the particular experimental conditions. The total listening time for each group of subjects was about 30 minutes, with a short break after the 80th sentence. After each stimulus, the subjects had 4 seconds to record their judgments. They were asked to rate the quality of the stimuli according to the adjectives: excellent, good, fair, poor, unsatisfactory. Their answer sheet contained 144 rows of short lines divided into nine columns, with the odd ones labeled with the adjectives. In this way, the subjects were allowed to check intermediate ratings, if they chose to do so.

The categorical judgments expressed by the listeners were subsequently converted into numerical scores, assigning value 1 to the category "unsatisfactory," value 9 to the category "excellent," and intermediate integer values to intermediate categories.

Before the actual test sessions took place, the subjects listened to 12 practice sentences different from those used in the experiment, spoken by the same four talkers, and representative of the range of quality they expected in the test.

VI. ANALYSIS OF SUBJECTIVE RESULTS

6.1 Control variables

The purpose of the subjective test was to assess the different behavior of coders in the presence of different operating conditions, namely, channel error rate and input level. Other sources of variability in listener responses are expected to cancel out in the average data for each experimental condition. Before averaging the data for each experimental condition, it was necessary to assess the importance of such extraneous sources of variability, as differences in the way the listeners judged the stimuli and differences due to talker voices.

6.1.1 Listeners

To assess the variability due to listener differences, their responses were analyzed according to MDPREF.^{16,17} This is a factor analytic procedure that derives a geometrical multidimensional space representation, in which the stimuli are represented by points and the subjects by vectors. The projections of the points on a vector are the best fit with the scores given to the stimuli by that subject. Basically, MDPREF reveals whether the subjects attended to different psychological attributes in the stimuli or if they attached different weights to each of the various attributes. In the solution for the 72 experimental conditions and the 20 subjects, the first principal component accounted for only 55 percent of the variance, while the remaining 45 percent was distributed over all the other components: 4.2 percent for the 2nd, 3.8 percent for the 3rd, 3.4 percent for the 4th, 3 percent for the 5th, etc.

The fact that 45 percent of the total variance was accounted for by so many axes in an almost uniform fashion indicates that these axes do not represent different perceptual attributes of the stimuli, but that they account only for the "noise" in the subjective data. In other words, in spite of the low variance accounted for by the first axis, it is evident that the listeners attended essentially to the same attributes with the same weights and then only a unidimensional solution exists. Therefore, the mean of the listeners' ratings for each condition were used for the subsequent analyses.

6.1.2 Talkers

An analysis of variance was computed to study the variability of the scores obtained by the talkers of different sex and to assess the validity of averaging the ratings across talkers to perform an analysis of the effect of the design variables. The analysis showed that the difference due to the sex of the talker was highly significant. The average score was 4.40 for female talkers and 5.07 for male talkers. On the other side, however, all the interactions between the sex of the talker and the design variables were not significant. This indicates that the sex of the talker influenced the average value of the ratings, but not the relative ranking of the various experimental conditions. Therefore, the mean ratings across listeners and talkers were used for further analyses, reducing the variability of the data to that due to the physical variables of the coders and the circuits.

6.2 Design variables

In Figs. 3 and 4, the mean ratings across listeners and talkers are shown for each bit rate B and decay constant β as a function of the probability of error $P(e)$, with α and the level L as parameters.

When no error dissipation mechanism is provided (Figs. 3a and 4a), a slower adaptation, i.e., $\alpha = 0.96875$, makes the system less sensitive to the errors. With a slow dissipation, i.e., $\beta = 255/256$ (Figs. 3b and 4b), the slow adaptation is advantageous only at the higher bit error rate, while with no errors the performance appears to be worse than with fast adaptation. With faster error dissipation, i.e., $\beta = 63/64$ (Figs. 3c and 4c) and slow adaptation, the unbalancing of the load factor between low and high input level is very high and the performance at the low level is very degraded even with no errors. With fast adaptation, i.e., $\alpha = 0.75$, the dynamic range is instead very high; besides, even if the performance under error-free conditions is lower than with slower error dissipation, the system is very insensitive to channel errors.

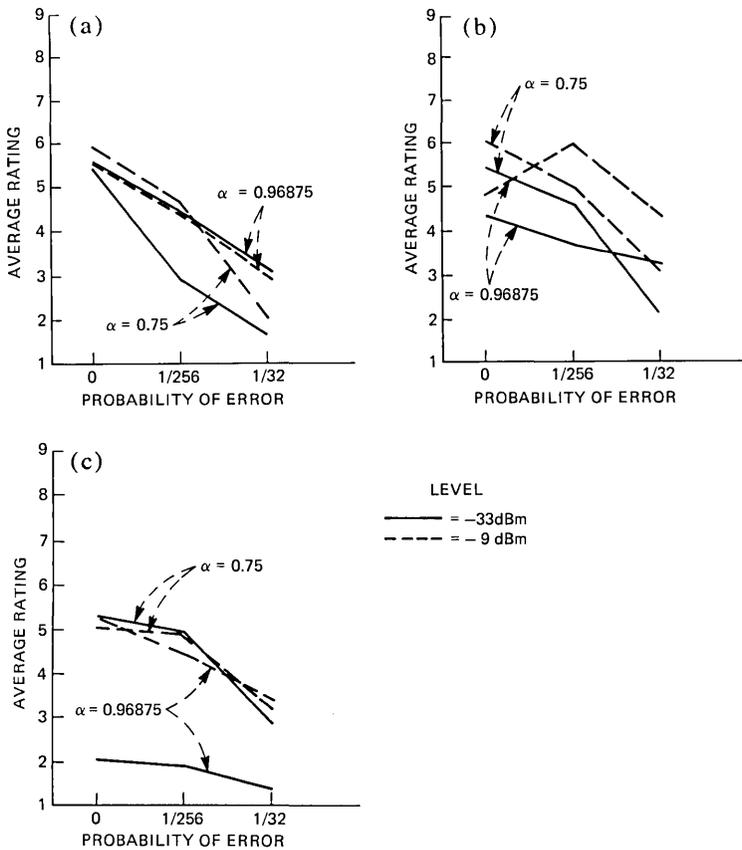


Fig. 3—Mean ratings as a function of probability of error, for $B = 3$: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

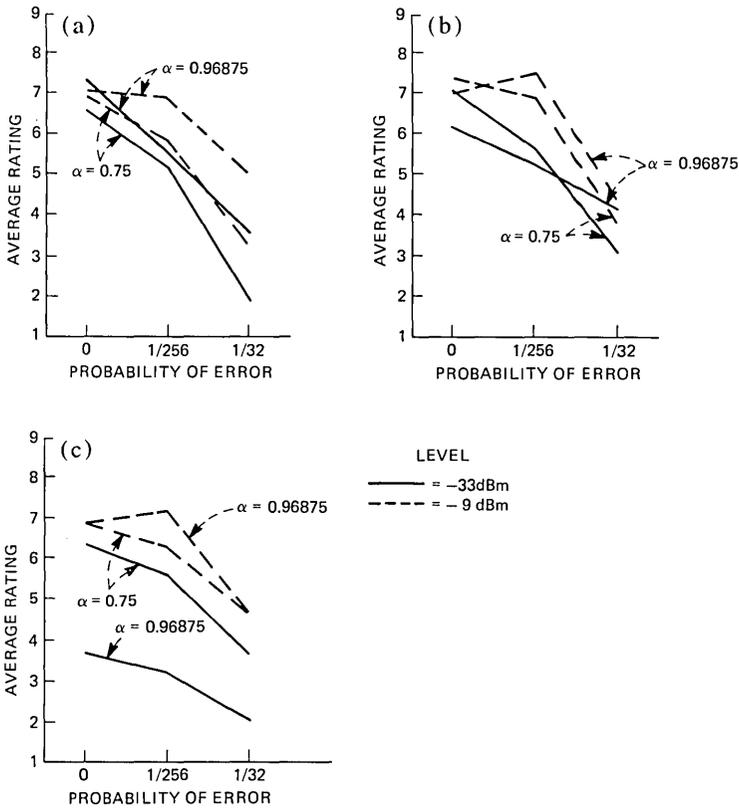


Fig. 4—Mean ratings as a function of probability of error, for $B = 4$: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

6.2.1 Analysis of variance

A five-way analysis of variance was performed to evaluate the effect of the experimental variables. The results of the analysis are reported in Table I. The last column reports the P value, that is, the probability that the test statistics assume a value greater than or equal to the computed F ratio, under the null hypothesis, compared with the two significance levels 0.05 and 0.01.

The analysis showed that all the main effects except that due to α are highly significant. The fact that α has no significant effect means that it has a positive effect for certain combinations of parameters and a negative one for others, as shown by the significant interactions.

The interactions between B and α and between B and β were found to be not significant, indicating that α and β have the same effects on the quality of coded speech whether it is a three-bit or a four-bit one. All the other two-way interactions are highly significant. Only the

Table I—Analysis of variance of the mean scores across listeners and talkers

Source	Degrees of Freedom	Sum of Squares	Mean Squares	F Ratio	Significance
<i>B</i>	1	31.8928	31.8928	204.7	$P < 0.01$
α	1	0.1733	0.1733	1.11	> 0.05
β	2	4.9167	2.4583	15.77	< 0.01
<i>L</i>	1	19.1570	19.1570	122.96	< 0.01
<i>P</i> (<i>e</i>)	2	85.5395	43.7697	274.52	< 0.01
<i>B</i> x α	1	0.2854	0.2854	1.84	> 0.05
<i>B</i> x β	2	0.0034	0.0017	0.01	NS
<i>B</i> x <i>L</i>	1	0.7642	0.7642	4.93	< 0.01
<i>B</i> x <i>P</i> (<i>e</i>)	2	1.8088	0.9044	5.84	< 0.01
α x β	2	11.0713	5.5356	35.78	< 0.01
α x <i>L</i>	1	3.6933	3.6933	23.87	< 0.01
α x <i>P</i> (<i>e</i>)	2	5.0383	2.5191	16.28	< 0.01
β x <i>L</i>	2	3.8310	1.9155	12.38	< 0.01
β x <i>P</i> (<i>e</i>)	4	4.7084	1.1772	7.60	< 0.01
<i>L</i> x <i>P</i> (<i>e</i>)	2	1.5342	0.7671	4.95	< 0.05
α x β x <i>L</i>	2	6.2810	3.1405	20.16	< 0.01
Residual	43	6.9980	0.1558		

interaction between *L* and *P*(*e*) is significant at $P < 0.05$ but not at $P < 0.01$. This indicates that the level has an effect almost independent of the probability of error.

The significant interaction between *B* and *L* is due to the fact that the difference between the ratings at the two levels is greater on the average for $B = 4$ than for $B = 3$. The significant interaction between *B* and *P*(*e*) is instead due to the fact that the coder with higher bit rate has a greater loss in quality in passing from $P(e) = 0$ to $P(e) = 1/32$.

Of the three-way interactions, only that among α , β , and *L* is significant. All the other three- and four-way interactions were not significant, and they were pooled in the residual.

VII. QUALITY PREDICTION BY OBJECTIVE MEASURES

To find an objective predictor of the speech quality, linear regression procedures were used. A linear model was chosen not only for its simplicity, but also because in many cases it proved to be adequate to represent the relationship between objective measures and subjective quality. A linear relationship exists, for instance, between the simple SNR and the quality of speech degraded only by the addition of stationary random noise or of speech dependent noise.¹⁸⁻²⁰ A linear relationship exists also between signal-to-granular noise ratio and probability of overload, and the quality of speech processed by ADPCM coders when no transmission errors are present.⁵

To perform regression analyses, the subjective ratings were averaged across listeners and talkers, and the objective measures were also

averaged, taking the arithmetic mean of the values obtained for each processed sentence. The gain fluctuation σ_g was instead averaged quadratically, taking the square root of the arithmetic mean of the squared values σ_g^2 .

Different sets of regression formulas were computed, in which the objective performance measures, like signal-to-noise ratios or spectral distance measures, were used either singly or in combination with the two measures of level mismatching. Figures 5 and 6 show the gain fluctuation and the average gain, both averaged across bit rate, as a function of the probability of error. A few remarks should be made on these figures. Although the two sets of measures have a fairly low correlation of 0.55, the patterns are much alike for low input level and fast adaptation. For the 18 conditions with $L = -33$ dBm and $\alpha = 0.75$, the correlation between G and σ_g is, in fact, 0.96. Therefore, even if in

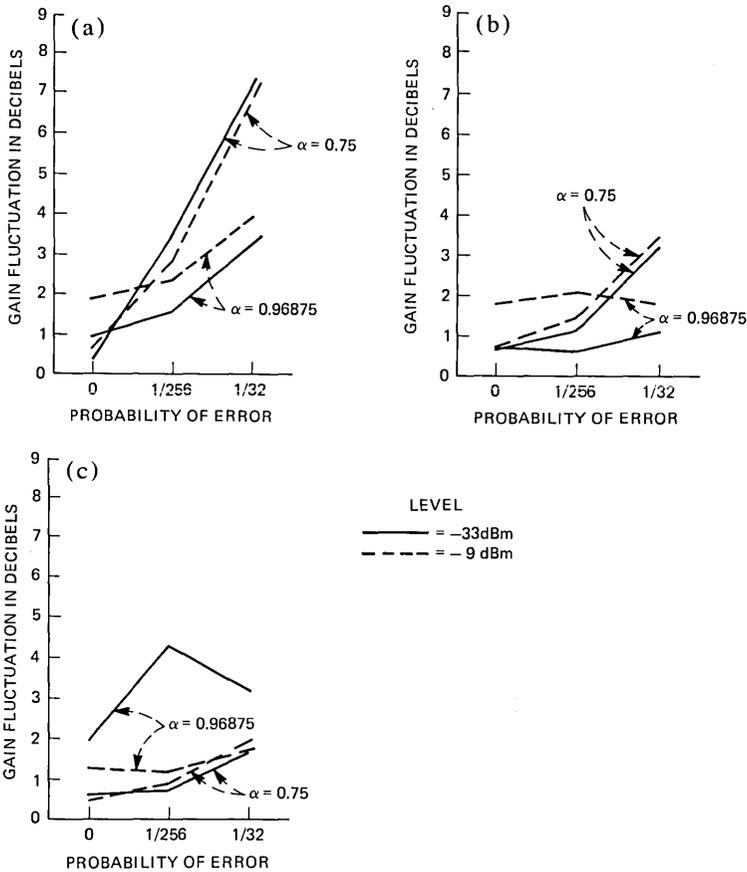


Fig. 5—Gain fluctuation, averaged across bit rate, as a function of probability of error: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

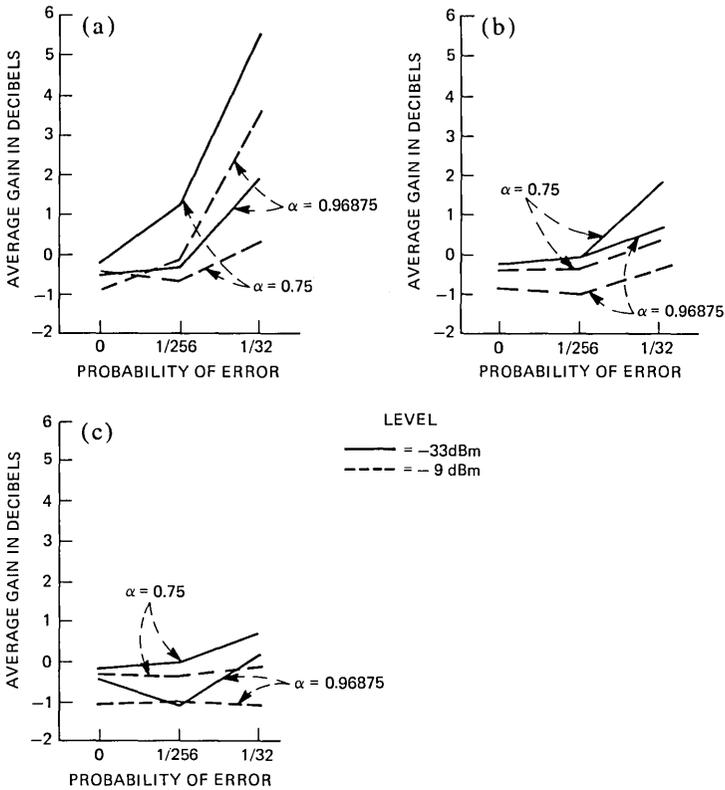


Fig. 6—Average gain, averaged across bit rate, as a function of probability of error: (a) $\beta = 1$. (b) $\beta = 255/256$. (c) $\beta = 63/64$.

a general case the two measures are independent each of the other, in this particular experiment a high value of σ_g is normally associated with a high average gain. This is particularly true if the input level is low, because, owing to the channel errors, the output level tends to be increased. If the input level is high and the quantizer step size is close to its maximum value, the output level is likely to increase only to a smaller extent.

The form of the relationship between the two level mismatching measures and the loss in quality due to the sensitivity of the adaptation algorithms to channel errors was not clear *a priori*, and therefore various nonlinear transformations were tried on those measurement data. No transformation on gain fluctuation proved useful in regression equations, while a compression of the average gain, given by

$$\tilde{G} = G/\sqrt{|G|}, \quad (27)$$

gave better predictions than G , when associated with the other performance measures.

Table II reports the results of the regression analysis. The prediction accuracy is indicated by both the correlation coefficient between the true and predicted subjective scores and the rms prediction error. After each regression analysis, however, a goodness-of-fit test was performed to test normality of prediction errors. The Kolmogorov-Smirnov test⁸ was used and in each case the hypothesis of normal distribution was accepted at the 0.20 significance level.

Table II—Formulas for predicting ratings using objective measures

	Formula for Predicting Rating	Correlation	rms error
1	$\hat{R} = 0.156 \text{ SNR} + 2.702$	0.667	1.202
2	$\hat{R} = 0.247 \text{ SNR}_{\text{seg}} + 1.369$	0.873	0.787
3	$\hat{R} = 0.271 \text{ SNR}_{\text{seg}} + 0.260\tilde{G} + 1.071$	0.881	0.765
4	$\hat{R} = 0.274 \text{ SNR}_{\text{seg}} + 0.110\sigma_R + 0.767$	0.875	0.780
5	$\hat{R} = 0.302 \text{ SNR}_{\text{seg}} + 0.270\tilde{G} + 0.122\sigma_R + 0.393$	0.883	0.756
6	$\hat{R} = 0.336 \text{ SNR}_{\text{com}} - 0.486$	0.911	0.665
7	$\hat{R} = 0.316 \text{ SNR}_{\text{com}} - 0.394\tilde{G} - 0.228$	0.935	0.571
8	$\hat{R} = 0.292 \text{ SNR}_{\text{com}} - 0.179\sigma_R + 0.569$	0.923	0.621
9	$\hat{R} = 0.297 \text{ SNR}_{\text{com}} - 0.337\tilde{G} - 0.088\sigma_R + 0.254$	0.938	0.561
10	$\hat{R} = 0.389 \text{ SNR}_{\text{max}} - 1.529$	0.913	0.656
11	$\hat{R} = 0.367 \text{ SNR}_{\text{max}} - 0.335\tilde{G} - 1.212$	0.931	0.590
12	$\hat{R} = 0.343 \text{ SNR}_{\text{max}} - 0.160\sigma_R - 0.449$	0.923	0.622
13	$\hat{R} = 0.345 \text{ SNR}_{\text{max}} - 0.283\tilde{G} - 0.088\sigma_R - 0.668$	0.933	0.581
14	$\hat{R} = 0.267 \text{ SNRF}_1 + 3.906$	0.878	0.773
15	$\hat{R} = 0.251 \text{ SNRF}_1 - 0.295\tilde{G} + 3.920$	0.891	0.731
16	$\hat{R} = 0.219 \text{ SNRF}_1 - 0.268\sigma_R + 4.606$	0.909	0.673
17	$\hat{R} = 0.216 \text{ SNRF}_1 - 0.150\tilde{G} - 0.235\sigma_R + 4.527$	0.912	0.662
18	$\hat{R} = 0.238 \text{ SNRF}_2 + 3.290$	0.887	0.744
19	$\hat{R} = 0.224 \text{ SNRF}_2 - 0.272\tilde{G} + 3.339$	0.899	0.708
20	$\hat{R} = 0.197 \text{ SNRF}_2 - 0.249\sigma_R + 4.050$	0.914	0.656
21	$\hat{R} = 0.195 \text{ SNRF}_2 - 0.140\tilde{G} - 0.219\sigma_R + 3.983$	0.916	0.647
22	$\hat{R} = 0.307 \text{ SNRF}_5 + 9.836$	0.855	0.838
23	$\hat{R} = 0.286 \text{ SNRF}_5 - 0.386\tilde{G} + 9.499$	0.879	0.770
24	$\hat{R} = 0.244 \text{ SNRF}_5 - 0.323\sigma_R + 9.461$	0.905	0.688
25	$\hat{R} = 0.243 \text{ SNRF}_5 - 0.194\tilde{G} - 0.278\sigma_R + 9.318$	0.909	0.671
26	$\hat{R} = -6.514 D_I + 7.613$	0.797	0.975
27	$\hat{R} = -6.035 D_I - 0.474\tilde{G} + 7.345$	0.837	0.883
28	$\hat{R} = -4.994 D_I - 0.388\sigma_R + 7.737$	0.878	0.772
29	$\hat{R} = -4.977 D_I - 0.225\tilde{G} - 0.334\sigma_R + 7.592$	0.885	0.751
30	$\hat{R} = -0.519 D_B + 9.154$	0.826	0.910
31	$\hat{R} = -0.482 D_B - 0.423\tilde{G} + 8.788$	0.856	0.834
32	$\hat{R} = -0.404 D_B - 0.306\sigma_R + 8.797$	0.868	0.800
33	$\hat{R} = -0.405 D_B - 0.262\tilde{G} - 0.241\sigma_R + 8.647$	0.878	0.773
34	$\hat{R} = 0.436 \text{ SDR}_1 - 1.268$	0.850	0.849
35	$\hat{R} = 0.407 \text{ SDR}_1 - 0.442\tilde{G} - 0.923$	0.883	0.757
36	$\hat{R} = 0.348 \text{ SDR}_1 - 0.281\sigma_R + 0.518$	0.885	0.751
37	$\hat{R} = 0.353 \text{ SDR}_1 - 0.304\tilde{G} - 0.203\sigma_R + 0.254$	0.897	0.712
38	$\hat{R} = 0.947 \text{ SDR}_2 - 0.249$	0.874	0.784
39	$\hat{R} = 0.893 \text{ SDR}_2 - 0.513\tilde{G} - 0.028$	0.917	0.642
40	$\hat{R} = 0.766 \text{ SDR}_2 - 0.306\sigma_R + 1.331$	0.918	0.641
41	$\hat{R} = 0.785 \text{ SDR}_2 - 0.353\tilde{G} - 0.212\sigma_R + 0.998$	0.934	0.578

Among the objective performance measures taken singly, the best one turns out to be SNR_{\max} , with a correlation coefficient of 0.913 and an rms error of 0.656 [formula 10 in Table II]. The compensated signal-to-noise ratio SNR_{com} gives almost the same results, while all the other measures achieve a correlation lower than 0.9. In particular, the conventional, long-term, signal-to-noise ratio has a correlation of only 0.667 and an rms prediction error almost double that of SNR_{\max} . The log likelihood ratio D_I is the second-worst predictor when used singly, with a correlation of only 0.797.

When the two measures of level mismatching, i.e., the average gain and the gain fluctuation, are included in the quality prediction formulas, the prediction accuracy is significantly improved, the rms prediction error having a 16-percent decrease on the average. The smallest improvement is displayed by SNR_{seg} . Among all the other measures, SNR_{com} gives the best prediction when combined with \bar{G} and σ_g (formula 9), with a correlation of 0.938 and an rms error of 0.561, about one-quarter of a category. Formulas 13 and 41, which use SNR_{\max} and SDR_2 , are almost as good as formula 9. The frequency-weighted SNRs also give a fairly good prediction, with correlations over 0.9 and the remaining frequency domain measures, D_I , D_B , and SDR_1 give a slightly poorer prediction.

VIII. DISCUSSION

8.1 Effects of coder design parameters

The subjective data have displayed complicated interactions among all the experimental design variables, the strongest interaction being the one between the adaptation constant α and the decay constant β . In fact, each of these two parameters affects different phenomena:

- (i) The *dynamic range* is reduced when β decreases from unity, but this reduction does not seem to be perceptible for any β if $\alpha = 0.75$. If $\alpha = 0.96875$ and $\beta = 255/256$, a certain reduction in the dynamic range begins to be perceived, producing a loss in quality at the low level of about 1.5 points with respect to the high level. When $\alpha = 0.96875$ and $\beta = 63/64$, the dynamic range is reduced still further, and the loss in quality of the low level with respect to the high one is very large, the average score dropping down from 5.3 to 2.4.
- (ii) The *effect of the errors* is smaller when α increases or β decreases. For instance, with $\alpha = 0.75$, the loss in quality passing from $P(e) = 0$ to $P(e) = 1/256$ averages 0.49 when $\beta = 63/64$, while it averages 1.59 for $\beta = 1$ and 0.93 for $\beta = 255/256$.

- (iii) The difference in the effect of the errors between the two input levels is higher for faster adaptation. For instance, passing again from $P(e) = 0$ to $P(e) = 1/256$, the difference between the losses at the two levels averages 1.0 when $\alpha = 0.75$, while it averages 0.45 when $\alpha = 0.96875$.

8.2 Optimum coders

Given a fixed number of bits per sample, a combination of decay constant β and adaptation constant α provides the best output quality for a given probability of error.

In the case of error-free transmission, optimum quality should be attained with no error dissipation mechanism, i.e., $\beta = 1$ which produces theoretically infinite dynamic range. The parameter α is not very critical in that case.⁵

When the coder operates under noisy conditions and the probability of error is in the range of the values encountered in a normal telephone connection or even higher than that (as is the case of $P(e) = 1/256$), a very slow error dissipation associated with fast adaptation provides good robustness to channel errors, without impairing the dynamic range. Actually, in this experiment, the combination $\beta = 255/256$ and $\alpha = 0.75$ provided optimum performance even under error-free conditions.

If the probability of error is as great as $1/32$, more typical of mobile radio communications, the best compromise between dynamic range and error sensitivity is obtained by a slow adaptation constant, combined again with a slow error dissipation rate. However the use of faster error dissipation and faster adaptation could be almost as good for this very high error rate.

8.3 Objective measures

One aim of the experiment was to examine a certain number of objective measures of coder performance and to compare them in the light of the actual subjective quality ratings, obtained under very different conditions. With the results of correlation and regression analyses reported in Table II, it is possible to observe strengths and weaknesses of the different measures and to derive general indications on which are the desirable properties of an objective quality measure.

A first indication that emerges from the experimental data is that the conventional long-term SNR is a very poor indicator of the quality of ADPCM coders under noisy channel conditions; this confirms the results obtained with PCM¹⁸ and ADPCM coders⁵ in the case of error-free transmission. Therefore, the use of SNR can be completely misleading, when comparing different coders operating under noisy conditions. A noticeable improvement in prediction accuracy is obtained simply by

measuring signal-to-noise ratio segmentally. Being time-segmental is a necessary property of any successful objective quality measure of coded speech.

Table II demonstrates also that, when a coder incorporating an adaptive quantizer is operating under noisy channel conditions, an objective performance measure must not be sensitive to changes or fluctuations of the output speech level. These fluctuations can be measured separately and the value obtained can be combined with the performance measure to improve the accuracy of the subjective quality prediction. It should be noticed, for instance, that \bar{G} and σ_g have a positive coefficient when combined with SNR_{seg} (formula 5 in Table II). This indicates that the level mismatching is weighted too much in SNR_{seg} , which did not incorporate any level compensation.

An important consideration on the subject of objective quality measures is that results from recent experiments^{7,9} indicate that frequency-weighted signal-to-noise ratios improve the prediction accuracy, especially when largely different noise spectra are produced by the coders.^{7,9} In this experiment, actually the frequency-weighted SNRs did not predict the subjective ratings as accurately as the simpler level-compensated SNRs, namely, SNR_{com} and SNR_{max} . This may depend on the fact that the level compensation is effected by minimizing the rms error on the whole bandwidth; this fact may worsen the measure in same band. If this is the case, it would be a weakness of the frequency-weighted SNR's when measuring coder performance in the presence of channel errors.

A final remark on frequency domain measures: These measures are more general than time domain ones because they are insensitive to short delays or to phase distortion.^{9,14} Therefore, they are more easily applicable to the test of coders whose input and output signals are in analog form or which include digital filters. On the other hand, the performance measures used here incorporate a spectral noise weighting that cannot be directly controlled. However, it is encouraging to see that the newly defined spectral signal-to-distortion ratio, SDR_2 , provides a very good prediction of subjective ratings when combined with the level mismatching measures.

More work is needed in the field of objective prediction of coder quality. In particular, frequency weighted SNRs should be evaluated more carefully, to derive the appropriate frequency weighting mechanism. In addition, the difference in behavior of the various spectral distortion measures need to be analyzed in more depth.

8.3.1 Estimation of the subjective effect of level mismatching

8.3.1.1 A Physical Interpretation of Prediction Formulas. The results of Table II lend themselves to a nice interpretation. The quality of an

adaptive coder operating in a noisy channel environment (high probability of error) may be considered as composed of two terms: (i) “intrinsic” goodness of the speech reproduction, which takes into account the noise due to the coding and to the errors, but not the level mismatching, (ii) loss in quality due to the level variations caused by the sensitivity of the adaptation algorithm to channel errors. In formulas, we can write

$$\hat{R} = \hat{R}_I - \hat{R}_L. \quad (28)$$

\hat{R}_I can basically be estimated by any of the performance measures which incorporate gain compensation or, in any case, which are not sensitive to alterations in the output signal level. \hat{R}_L is instead estimated by a linear combination of the gain fluctuation and the average gain, modified according to eq. (27).

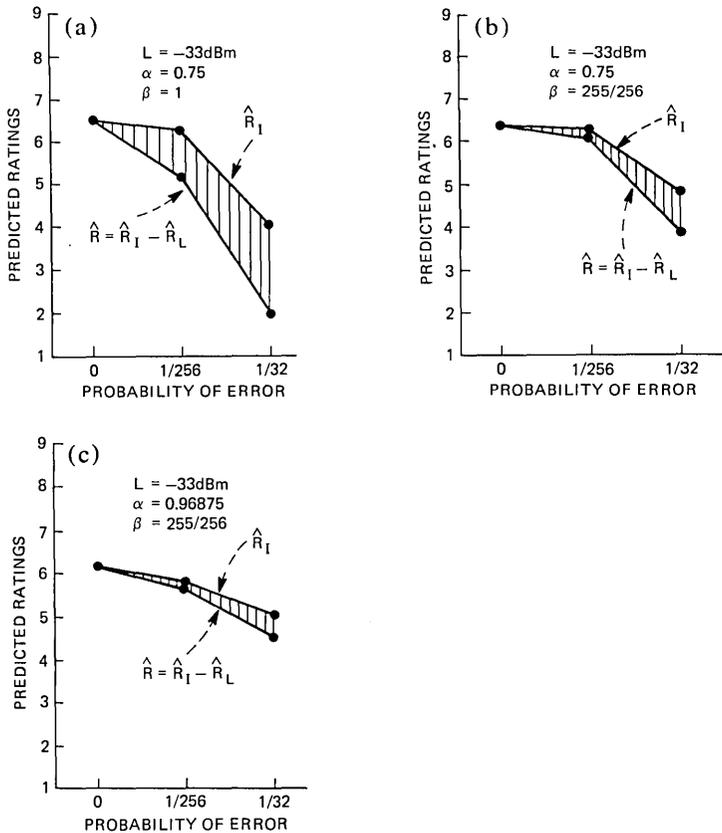


Fig. 7—Predicted overall rating \hat{R} and “intrinsic” goodness \hat{R}_I , as a function of probability of error, for 4-bit ADPCM and different combinations of design parameters: (a) $\alpha = 0.75$, $\beta = 1$. (b) $\alpha = 0.75$, $\beta = 255/256$. (c) $\alpha = 0.96875$, $\beta = 255/256$.

8.3.1.2 An Example. In the light of the interpretation given in the previous section, it is possible to give at least a qualitative answer to the question on which is the subjective effect of level mismatching. Figure 7 shows the predicted overall rating \bar{R} and the "intrinsic" goodness \bar{R}_I as a function of the probability of error, for the 4-bit ADPCM coder with low level input and three different combinations of design parameters. \bar{R} was computed according to formula 41 in Table II, while \bar{R}_I was computed discarding the terms involving \bar{G} and σ_g from the same formula:

$$\bar{R}_I = 0.785 SDR_2 + 0.998.$$

In the case of fast adaptation and absence of error recovery (Fig. 7a), the loss in quality due to level mismatching can be estimated as half category (1 point) for the intermediate error rate and 1 category (2 points) for the high error rate. In the case of fast adaptation and slow error recovery (Fig. 7b), the loss \bar{R}_L is instead reduced to about only half category (1 point) for the high error rate.

Finally, when slow adaptation and slow error recovery are used, simultaneously (Fig. 7c), the loss due to level mismatching can be estimated as only about 0.4 point, i.e., less than a quarter of a category.

IX. ACKNOWLEDGMENTS

The author wishes to thank B. J. McDermott, B. H. Bharucha, R. E. Crochiere, and J. M. Tribolet for several useful discussions and for having provided subroutines for computing some of the objective measures. The author thanks also J. Coker for recruiting the subjects and running the listening tests.

REFERENCES

1. D. J. Goodman and R. M. Wilkinson, "A Robust Adaptive Quantizer," *IEEE Trans. Commun., COM-23*, No. 11 (November 1975), pp. 1362-1365.
2. C. Scagliola, "An Adaptive Speech Coder with Channel Error Recovery," International Conference on Communications, Chicago, Ill., June 1977.
3. D. Mitra, "An Almost Linear Relationship Between the Step-Size Behavior and the Input Signal Intensity in Robust Adaptive Quantization," *IEEE Trans. Commun., COM-27*, No. 3 (March 1979).
4. P. Castellino, G. Modena, L. Nebbia, and C. Scagliola, "Bit Rate Reduction by Automatic Adaptation of Quantizer Step-size in DPCM Systems," International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.
5. B. J. McDermott, C. Scagliola, and D. J. Goodman, "Perceptual and Objective Evaluation of Speech Processed by Adaptive Differential PCM," *B.S.T.J.*, 57, No. 5 (May-June 1978), pp. 1597-1618.
6. P. Noll, "Adaptive Quantizing in Speech Coding Systems," International Zurich Seminar on Digital Communications, Zurich, Switzerland, March 1974.
7. R. E. Crochiere, L. R. Rabiner, N. S. Jayant, and J. M. Tribolet, "A Study of Objective Measures of Speech Waveform Coders," International Zurich Seminar, Zurich, Switzerland, March 1978.
8. A. A. Afifi and S. P. Azen, *Statistical Analysis a Computer Oriented Approach*, New York: Academic Press, 1972.

9. J. M. Tribolet, P. Noll, B. J. McDermott, and R. E. Crochiere, "Complexity vs. Quality for Speech Waveform Coders," IEEE International Conference on Acoustics Speech and Signal Processing, Tulsa, Oklahoma, April 1978.
10. N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," *J. Acoust. Soc. Amer.* 19 (January 1947), pp. 90-119.
11. J. D. Markel and A. H. Gray, *Linear Prediction of Speech*, New York: Springer Verlag, 1976.
12. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Trans. Acoust. Speech and Sig. Proc.*, ASSP-23, (February 1975), pp. 67-72.
13. A. H. Gray and J. D. Markel, "Distance Measures for Speech Processing," *IEEE Trans. Acoust. Speech Sig. Proc.*, ASSP-24, (October 1976), pp. 380-391.
14. B. H. Bharucha, "An Objective Measure of Codec Speech Quality," unpublished paper, 1976.
15. N. S. Jayant, "Adaptive Quantization with One-Word Memory," *B.S.T.J.*, 52, No. 7, (September 1973), pp. 1119-1144.
16. P. Slater, "Analysis of Personal Preferences," *Brit. Journal of Statistical Psychology*, 13 (November 1960), pp. 119-135.
17. J. D. Carroll, "Individual Differences and Multidimensional Scaling," in *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, Vol. I, Shepard, Romney, Nerlove (Eds.), New York: Seminar Press, 1972, pp. 105-155.
18. D. J. Goodman, B. J. McDermott, and L. H. Nakatani, "Subjective Evaluation of PCM Coded Speech," *B.S.T.J.*, 55, No. 8, (October 1976), pp. 1087-1109.
19. D. L. Richards, *Telecommunications by Speech*, London: Butterworths, 1973, Ch. 4.
20. L. Nebbia and P. Usai, "Influence of Some Types of Noise on Telephone Digital Transmissions," Symposium, "Speech Intelligibility," Liege, November 1973.

An Adaptive Intraframe DPCM Codec Based Upon Nonstationary Image Model

By N. F. MAXEMCHUK and J. A. STULLER

(Manuscript received November 21, 1978)

This paper introduces a nonstationary model for images and develops an adaptive intrafield DPCM codec based upon the model. The codec attempts to minimize the mean-square coding error at each sample point in the picture. The quantizer in the resulting adaptive codec is found to be similar to that previously obtained from visual masking considerations. Comparative simulation results using 256×256 pixel rasters are given for two- and three-bit/pixel versions of the adaptive codec, the three-bit/pixel Graham codec, and three-bit/pixel previous element DPCM.

I. INTRODUCTION

This paper introduces a nonstationary model for images and develops an adaptive intrafield codec based upon the model. The codec adaptively estimates both the mean and probable range of values of the next picture sample to be encoded and adapts the predictor and quantizer accordingly. In so doing, the coder attempts to minimize the mean-square coding error (MMSE) at each sample point in the picture. The MMSE distortion measure is generally acknowledged to be a poor indicator of image quality.¹ However, when it is applied on a *point* (rather than area) basis in conjunction with the image model presented here, the coder adaptation and resulting coding quality are found to be comparable to that previously obtained from visual masking considerations. This result follows from a property of human vision, stressed by Graham,² concerning the strong connection that exists between image chaos (unpredictability) and the visual system's tolerance to noise-like coding distortion. Because of this property, we obtain good image quality at two bits per pel and excellent quality at three bits per pel in a DPCM codec designed solely using the MMSE criterion—masking phenomena are in large part accounted for automatically when the

source model more adequately represents actual images and when the distortion criterion is applied on a point basis.

II. SOURCE MODEL

This section introduces a nonstationary causal source model for the intrafield video process that will be used to develop the adaptive predictive intrafield codec of Section IV. Motivation for the model is intuitive and follows from an examination of a representative video signal of the type we wish to encode, such as that shown in Fig. 1. This is a frame of two interlaced fields, each having 256 pixels per line and 128 lines, with amplitudes stored as 8-bit quantities. The essential characteristic of this (and any) image is that it is a projective transformation of a collection of physical objects. As a consequence, the image is partitioned into regions of luminance elements whose amplitudes are interrelated by the physical structure of the objects they represent. The result is an array of pixels composed of distinct regions having slowly varying "brightness" and "texture" with abrupt boundaries (the picture outline) separating one region from another. We find it natural to view this array as a field that is partitioned into regions of independent quasi-stationary subfields. Two underlying random phenomena are involved: the random amplitudes of picture elements within a



Fig. 1—Checker girl original.

given subfield and the random selection of the subfield with respect to raster coordinates. A source model that incorporates both phenomena is shown in Fig. 2.

Figure 2 models the image generation process as a composite of Q autoregressive sources, $q = 1, 2, \dots, Q$, and one white source, $q = 0$. Switches S1 and S2 determine which source generates output luminance $s_{mn} \equiv s_t$, where m and n are, respectively, the line number in the field and the column number of the pixel and t is the time the pixel is encountered during conventional line scanning. The autoregressive sources, characterized by predictors 1 through Q and "innovations" process $w_{mn} \equiv w_t$, provide a set of Q possible processes from which the regions of slowly varying brightness and texture of a subfield in an actual image can be approximated. The random variables w_t are assumed zero mean, independent, and characterized by a single known probability density function (pdf) $g(w)$. The predictors F_q in sources $q = 1, 2, \dots, Q$ are taken to be linear functions of pixels from the local past neighborhood of coordinate $(m, n) \equiv t$. Section IV discusses the specific predictors chosen for the codec of this paper (Table I). Source 0 models those pixels of an actual image that either have no structural relation to previous pixels or whose relation to these pixels is not adequately modeled by sources 1 through Q . Such pixels tend to occur in highly chaotic regions of the image and at certain boundaries at which new subfields are initiated. Since this source represents the extreme of chaos possible in an image, its output is taken to be a

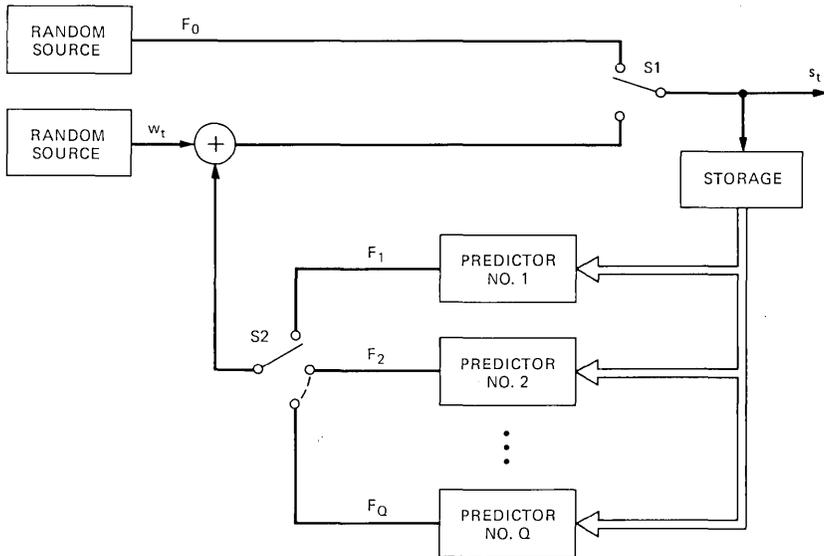


Fig. 2—Source model. Switches S1 and S2 are governed by eq. (2) in the text.

sequence of random variables $F_0 \equiv F_0(m, n)$, each uniformly distributed over $[0, 255]$.

Switches S1 and S2 of the model determine which source is used for final output at each raster coordinate and thus determine image outline as well as more subtle structural changes. "Outline" and "subtle structural changes" are subjectively perceived qualities of an image that are difficult to quantify probabilistically. However, in an actual image "structure" tends to vary slowly: boundaries between regions are an exception, but even here the discontinuity is generally only along one dimension. This suggests that probabilistic information regarding the source in operation at time t can be inferred by appropriate processing of the pixels in the local past vicinity (in the same field) of the pixel in question. To arrive at a source model that characterizes this quasistationary in the simplest way, we model the image source as choosing sources, $q = 0, 1, 2, \dots, Q$, *independently* according to *unknown* first-order probabilities $P[q; (m, n)]$ that are slowly varying functions of coordinates (m, n) . We further assume that the Q textures generated by sources $q = 1, 2, \dots, Q$ are *a priori* equally likely for a *random* choice of coordinate (m, n) :

$$E\{P[q; (m, n)]\} = c \quad q = 1, 2, \dots, Q \quad (1a)$$

and

$$E\{P[q; (m, n)]\} = \epsilon \ll c \quad q = 0, \quad (1b)$$

where c and ϵ are constants satisfying $\epsilon + Qc = 1$ and the expectation is taken over the raster coordinates.

An alternative approach would be to model the sequence of q_{mn} as stationary Markov. However, this approach was not taken since the assumption of nonstationary *independent* q leads to a relatively simple codec that is robust with respect to both varied picture inputs and channel errors. The assumption of equality of expectations in (1b) leads to mini-max performance with respect to variations in textural content of the picture to be encoded. Biasing this *a priori* distribution toward one predictor would make the codec more susceptible to poor performance on a picture which does not match this distribution. The problem faced by the codec is to estimate probabilities $P[q; (m, n)]$ by suitable processing of past pixel outputs and use the estimates to best advantage for bandwidth compression.

To summarize, the proposed model of the video process has the form (Fig. 2)

$$s_t = \begin{cases} F_0, & \text{with probability } P(0; t) \\ F_q + w_t, & \text{with probability } P(q; t), \quad 1 \leq q \leq Q, \end{cases} \quad (2)$$

where F_0 is an independent random variable uniform over $[0, 255]$, F_q is a given linear function of pixels in the local past neighborhood of s_t

(Table I), and w_t is an independent zero mean random variable characterized by probability density function (pdf) $g(w)$. Probabilities $p(q; t)$ $q = 0, 1, \dots, Q$ vary slowly with respect to at least one coordinate of the raster and satisfy (1); otherwise, these probabilities are unknown. Note that the model embeds the elusive variety of gross image structure in the unknown probabilities $P(q; t)$, $0 \leq q \leq Q$. These represent the probabilistic information that the encoder hopes to learn by suitable processing of past image source outputs.

Figure 3 illustrates a representative output generated by the model. In obtaining this output, $g(w)$ was assumed Laplacian, and the $P(q; t)$ were estimated from the image of Fig. 1 by a procedure described in Section III. Significant increases in structural similarity to Fig. 1 are possible by modeling the sequence of w_t as nonstationary. In the interest of codec simplicity, however, this additional complexity is not included in the model.

III. ANALYTICAL DEVELOPMENTS

The design of the DPCM codec of Section IV requires specification of both the quantizer and the predictor. Complete statistical information pertinent to this design is contained in the conditional pdf of s_t given



Fig. 3—Representative output of source model. Output of the source model of Fig. 2, where w_t is Laplacian and probabilities $P(q; t)$ of eq. (2) are estimated from Fig. 1.

the set of past pixels $\{s_i, t < t\} \equiv S_t^-$. This section describes how this conditional pdf can be estimated at the source output.

It can be easily shown that, for *given* $P(q; t)$, $q = 0, 1, \dots, Q$, the probability density of s_t conditioned on S_t^- is (for $0 \leq s_t \leq 255$)

$$p(s_t | S_t^-) = \frac{P(0; t)}{255} + \sum_{q=1}^Q g(s_t - F_q(S_t^-))P(q; t), \quad (3)$$

where $F_q(S_t^-)$ denotes the q th predictor F_q of s_t as an explicit function of past pixels S_t^- . An estimate of density function (3) is obtained by replacing $P(q; t)$ in the above by its estimate, as described below.

Let the number of times the q th source had been output in a local past region R_t of N points neighboring $(m, n) = t$ (Fig. 4) be denoted by $n(q)$. Due to the nature of the source model, $n(q)$ cannot be measured at the source output. However, a reasonable and computable approximation to it is given by the expectation $E\{n(q) | S_t^-\}$, where the expectation assumes a random selection of (m, n) and is over the density (w_i) . By the quasi-stationarity of $P(q; t)$, we then set

$$\hat{P}(q; t) = \frac{E\{n(q) | S_t^-\}}{N}, \quad (4)$$

which becomes (appendix)

$$\hat{P}(q; t) = \frac{1}{N} \sum_{j=1}^N \hat{P}(q_j = q | S_t^-), \quad (5)$$

where $\hat{P}(q_j = q | S_t^-)$ is the conditional probability that the j th pixel in R_t (Fig. 4) was output by source q based upon *a priori* probabilities $E\{P[q; (m, n)]\}$ of (1). Further manipulations (appendix) give

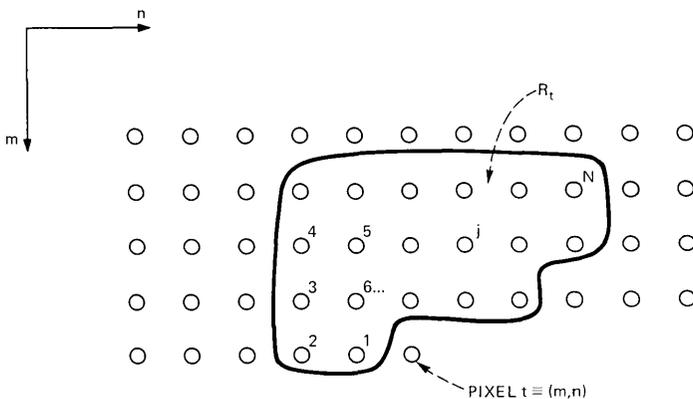


Fig. 4—Illustration of region R_t . This region consists of N pixels in a local past vicinity of pixel (m, n) . Note that the numbering of coordinates $j = 1, 2, \dots, N$ is arbitrary, as are the region boundaries.

$$\hat{P}(q; t) = K \frac{\epsilon}{255} \quad q = 0 \quad (6a)$$

$$\hat{P}(q; t) = \frac{KC}{N} \sum_{j=1}^N g(e_q(t; j)) \quad 1 \leq q \leq Q, \quad (6b)$$

where K satisfies

$$\sum_{q=0}^Q \hat{P}(q; t) = 1. \quad (7)$$

Equation (5) interprets $\hat{P}(q; t)$ as an arithmetic average of *a posteriori* probabilities of q over region P_t , and eq. (6) show how this average can be computed. The term $e_q(t; j)$ in (6b) is the difference between the actual value of the j th pixel in region R_t and the predicted value of this pixel given by predictor F_q and is therefore the implied value of the j th innovations variable in R_t under the hypothesis that predictor q was in operation at the source. Explicitly,

$$e_q(t; j) = s_t^j - F_q(S_t^{j-}), \quad (8)$$

where s_t^j is the j th pixel in R_t and S_t^{j-} is the set of pixels previous to s_t^j . Equation (6b) estimates $\hat{P}(q; t)$, by summing the relative probabilities of the innovations implied under the hypothesis that source q was in operation over region R_t . Note that if $g(\cdot)$ has its peak at zero, then $\hat{P}(q; t)$, $1 \leq q \leq Q$, will be large for those q corresponding to small prediction error $e_q(t, j)$ over the N point region. If none of the Q predictors is consistent with past local data, then all terms in the sum of (5b) will be small for $1 \leq q \leq Q$, and the normalization in (6) will make $\hat{P}(0; t)$ large. Further description of (4) to (7) is included in the derivation in the appendix.

The codec described in Section IV predicts s_t by the estimated mean of predictable source outputs:

$$\hat{s}_t = \frac{\sum_{q=1}^Q F_q(S_t^-) \hat{P}(q; t)}{\sum_{q=1}^Q \hat{P}(q; t)}. \quad (9)$$

An important characteristic of this prediction rule is its insensitivity to small variations in data S_t^- regardless of the relative values of N and Q . This is in contrast to the covariance method in linear prediction described in a review paper by Makhoul³ in which small sample size can lead to an ill-conditioned system of equations whose inversion is the adapted predictor. Since (9) is a weighted average of stable (and generally good) estimates F_q , stability persists even for $N < Q$, and some thought indicates that the resulting prediction of \hat{s}_t works in an intuitively reasonable way even if N is only unity.

IV. THE CODEC

In this section, a codec resulting from the source model is described. A block diagram of the encoder is shown in Fig. 5. The codec has been used to code pictures using two and three bits per pel.

The encoder operates by forming Q estimates $F_q(X_t^-)$, $1 \leq q \leq Q$, of source output S_t based upon the previously reconstructed field elements X_t^- . Estimates of source probabilities $P(q; t)$, $0 \leq q \leq 1$, are made with eq. (6) to (7) using previously reconstructed pixels X_t^j in place of S_t^j . Estimates $F_q(X_t^-)$ and probabilities $P(q; t)$ are used to predict the next encoder input pixel s_t , according to (9) and the most likely distribution of values s_t according to (3), with X_t^- replacing S_t^- .

The encoder has been implemented using an $N = 4$ point learning region R_t (Fig. 6) and $Q = 6$ predictors. The predictors used are given in Table I.

Note that with these six predictors the form of predictor (7) can be any one of the most common fixed predictors used in intrafield coders. This varies over the picture so that the best predictor (or best weighted sum) considering the recent past will be used at each sample point.

The pictures which were encoded consisted of 256 lines in two interleaved fields and 256 samples per line. The previous line elements were taken from the previous line in the same field. In this environment, no advantage was obtained by including elements more than one line away in the estimates. Similarly, no visible improvement was obtained using elements that were more than two elements away on the same line. The slope estimator, F_5 , and the planer estimator, F_6 , were found to be particularly useful in the system which uses two bits

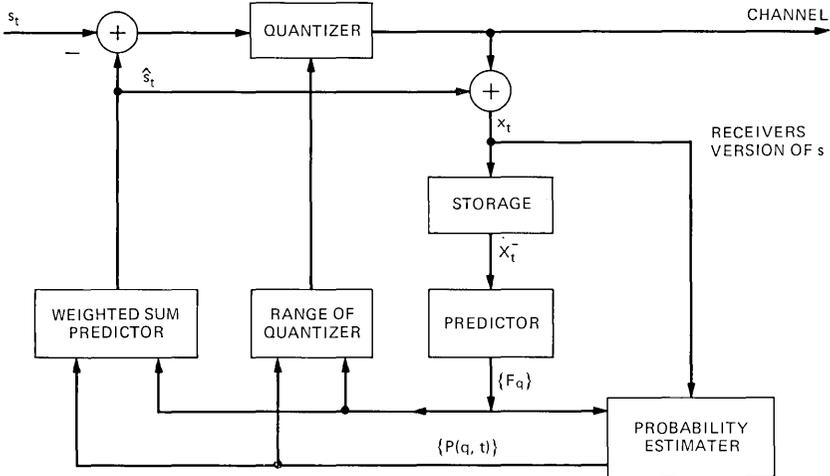


Fig. 5—The encoder.

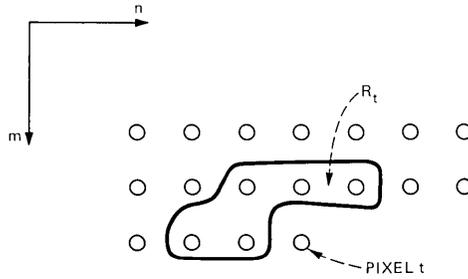


Fig. 6—Four-point region R_t used by codec.

Table I

$F_1(n, m) = \hat{x}(n - 1, m).$
$F_2(n, m) = \hat{x}(n - 1, m - 1).$
$F_3(n, m) = \hat{x}(n, m - 1).$
$F_4(n, m) = \hat{x}(n + 1, m - 1).$
$F_5(n, m) = 2\hat{x}(n - 1, m) - \hat{x}(n - 2, m).$
$F_6(n, m) = \hat{x}(n - 1, m) + \hat{x}(n, m - 1) - \hat{x}(n - 1, m - 1).$

per pel. These estimators allowed the coder to respond more quickly to edges within the picture, and reduced slope overload.

Ideally, the quantizer should be adapted at each point to the estimated probability distribution of s_t . In view of the complex form of (3), this type of redesign is not feasible, and the following ad-hoc curve-fitting technique was used to simplify the adaptation algorithm. The density function $g(w)$ was taken as Laplacian, $g(w) = \alpha/2 \exp(-\alpha |w|)$. The Max quantizer⁴ for this was determined. Each side of the distribution (3) about the mean \hat{s}_t was then approximated by an exponential distribution, and the axis was simply scaled appropriately in codec operation to place the quantization levels. The parameter of each exponential distribution was selected so that it had the same first moment about \hat{s}_t as the corresponding portion of the actual distribution as described in Fig. 7.

When the estimated probability of occurrence of the random estimator is near zero and the estimators $q = 1$ through 6 are identical corresponding to a perfectly flat region in the picture, the parameter of the exponential defining the quantizer assumes its smallest value. In this situation, the parameter of the exponential defining the quantizer is approximately α , the parameter of the Laplacian distribution defining the innovation term in the model. Therefore, α determines the minimum values of the levels of the quantizer, and these, in turn, determine the amount of granularity due to quantization noise in flat regions of the picture and the ability of the coder to respond to unexpected edges. The smaller the value of α , the lower the granular quantization noise; the larger the value of α , the quicker the coder can

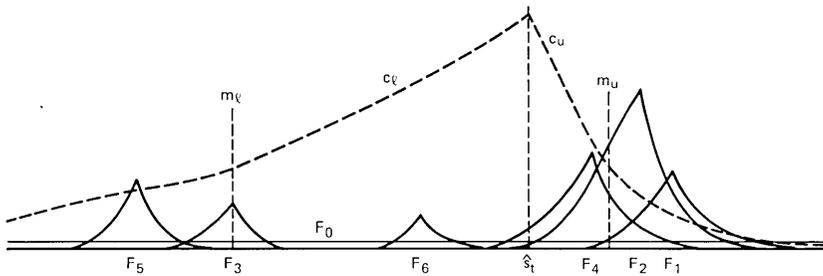


Fig. 7—Illustration of encoder's derivation of predictor and quantizer. F_0 refers to the distribution of the white source output; \hat{s}_t is the weighted sum predictor; m_u and m_l are the upper and lower first moments of the actual distribution about \hat{s}_t ; and c_u and c_l are the exponential distributions used to determine the quantizer.

respond to edges. Because of this interaction, the values of α were selected experimentally based upon visual examination of a sequence of coded pictures for the two- and three-bit/pel quantizers. For the two-bit/pel quantizer, α was selected so that the minimum value of the inner quantizer level is equal to two picture levels, when the picture is initially quantized into 256 levels. For the three-bit/pel quantizer, the inner quantization level was selected so that the inner quantization level is equal to one picture level.

The random variable F_0 in the source model of Fig. 2 is uniformly distributed over the range of possible values the sample can assume. In implementing the codec, it was found to be desirable to assume that the range of F_0 is somewhat reduced. Limiting the span of F_0 is particularly necessary in the system which transmits two bits per pel. This can be seen as follows. Assume that probability $P(0, t)$ is estimated by the encoder to be close to unity. In this situation, if F_0 has range $[0, 255]$ the four quantization levels will be spread over the entire range of possible sample values. It is then likely that none of the estimators will be close to the reconstructed value x_t even though an estimator can have closely approximated the actual value s_t . Thus, the random estimator may be used for the next sample. This creates an instability in the coder which can propagate into flat regions of the picture. To eliminate this type of instability, the maximum range of the quantizer was limited. To be consistent with limiting the maximum range of the quantizer, the span of F_0 was limited to a symmetrical region about \hat{s}_t of (9). The maximum span of the quantizer was also set experimentally. In the two-bit/pel system, the maximum span of the quantizer was set so that the inner level of the quantizer is eight picture levels. And in the three-bit/pel system, the maximum span of the quantizer was set so that the inner level in the quantizer is four picture levels. In the two-bit/pel system, there are only two quantization levels on each side of the predicted value. In this system, the maximum span of the

quantizer determined the ability of the encoder to track sudden changes in the picture. Therefore, it is necessary to make the maximum quantizer span as large as possible, without making the encoder unstable. In the three-bit/pel system, four quantization levels are on each side of the predictor. In this system, restricting the maximum quantizer span was necessary to prevent the quantizer span from frequently exceeding the range of possible picture levels and wasting quantization levels. This is why a smaller maximum value of the inner quantization level was selected for the three-bit/pel system than for the two-bit/pel system.

In Figs. 8 and 9, the quantization span for various parts of the picture in the two- and three-bit/pel systems is shown. In these pictures, the average of the upper and lower quantization spans is displayed. The white areas correspond to the smallest span of the quantizer and the black levels the largest span. It is interesting to note that the resulting quantizer adaptation is similar to that which would be expected if a masking function were used.⁵ However, this quantizer adaptation was arrived at strictly by mathematical techniques, minimizing the expected point mean-squared error with a varying probability distribution of next sample values, rather than by the psychovisual considerations used to derive masking functions. This result is



Fig. 8—Quantizer range adaptation of the two-bit/pel codec.



Fig. 9—Quantizer range adaptation of the three-bit/pel codec.

consistent with Graham's early observations concerning the strong connection between image chaos and the visual system's tolerance to noise-like distortion.²

V. RESULTS

This adaptive predictor with a two- and three-bit/pel quantizer has been implemented and compared with an adaptive predictor using Graham's rule² and the three-bit/pel fixed quantizer suggested in the Graham paper, and a previous element DPCM encoder with a fixed three-bit/pel quantizer. Our two-bit/pel adaptive predictor has considerably less slope overload than the previous element predictor having three bit/pel quantizer, but is not quite as good as the Graham predictor having a three bit/pel quantizer. Our adaptive predictor with a three bit/pel quantizer has less slope overload than the Graham predictor with a three bit/pel quantizer. In addition, the estimates at edges within the picture are accurate enough to virtually eliminate the edge business in moving sequences which is characteristic of many adaptive predictors. To demonstrate these characteristics, the difference between the original picture of the checker girl, Fig. 1, and the result of processing by these four techniques is shown in Figs. 10 to 13.

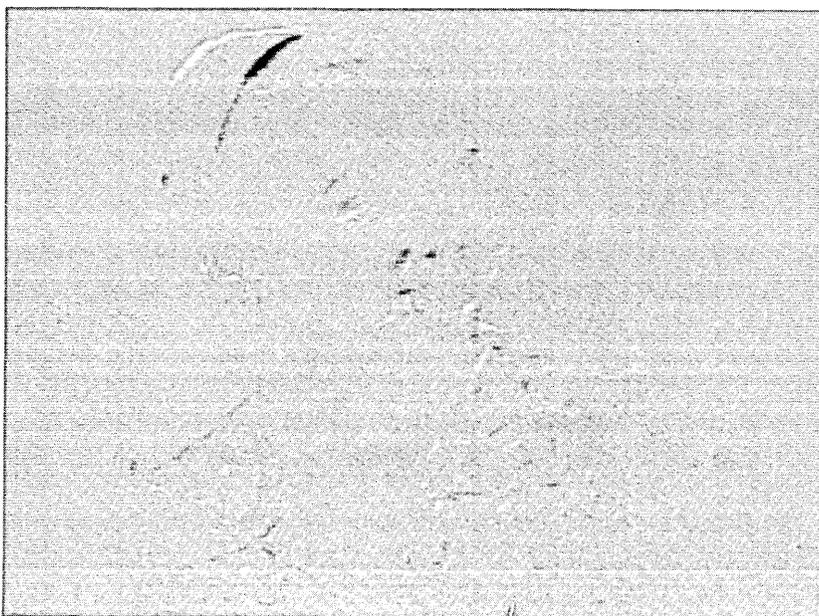


Fig. 10—Two-bit/pel codec performance. Top: Decoder output. Bottom: Difference between decoder output and original.

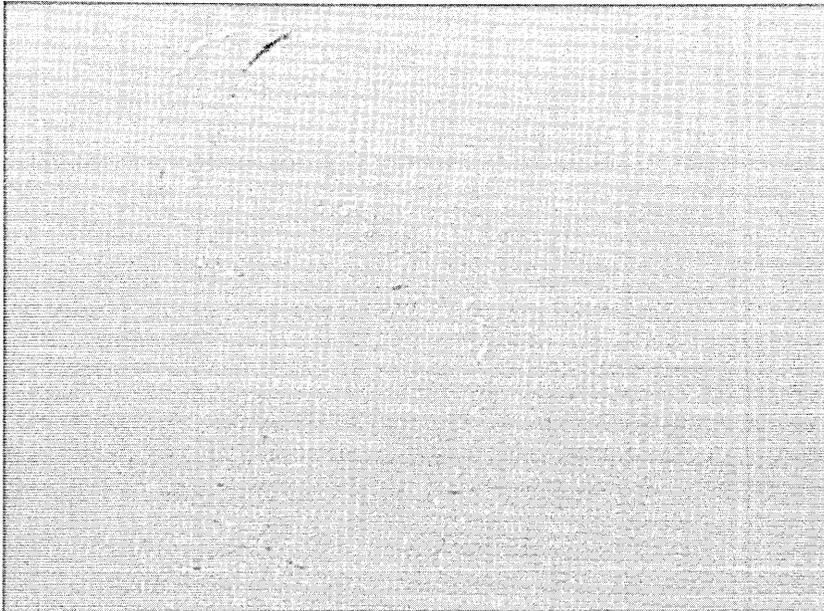


Fig. 11—Three-bit/pel codec performance. Top: Decoder output. Bottom: Difference between decoder output and original.

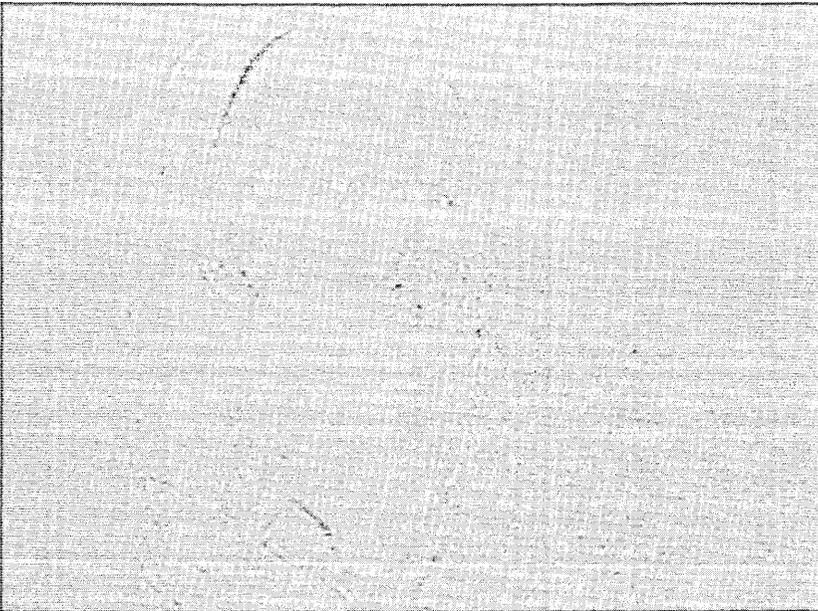


Fig. 12—Performance of three-bit Graham codec. Top: Decoder output. Bottom: Difference between decoder output and original.

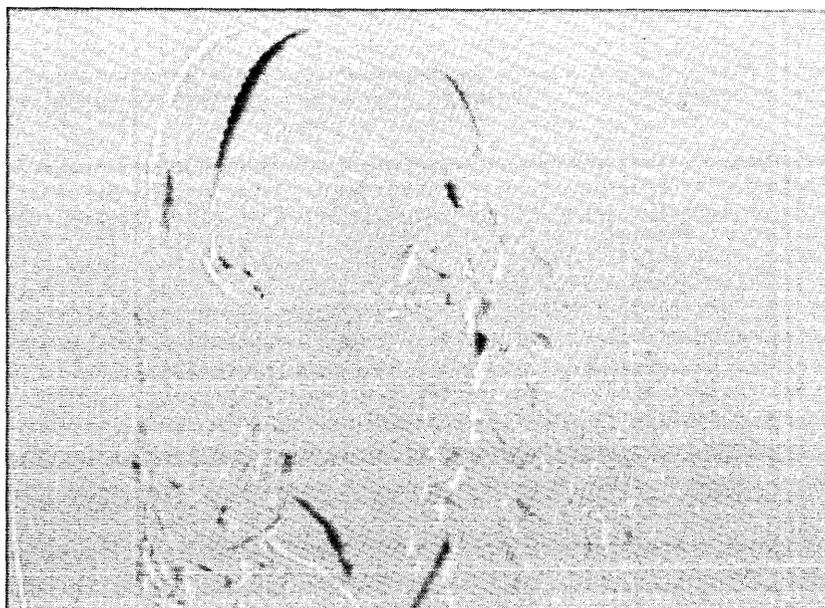


Fig. 13—Performance of three-bit previous element DPCM. Top: Decoder output. Bottom: Difference between decoder output and original.

APPENDIX

This appendix traces the development from eq. (4) to (7) of Section III.

There are $M = (Q + 1)^N$ possible vectors $\mathbf{V} = (q_1, q_2, \dots, q_j, \dots, q_N)$ of source options in the N point region R_t of Fig. 4. Number these vectors $i = 1, 2, \dots, M$ and let \mathbf{V}_i denote the i th vector. Define $n(q | \mathbf{V}_i)$ as the number of components in \mathbf{V}_i that equal the specific value q . Then (4) becomes

$$\hat{P}(q; t) = E \left| \frac{n(q)}{N} \right| \quad (10a)$$

$$= \frac{1}{N} \sum_{i=1}^M n(q | \mathbf{V}_i) \bar{P}(\mathbf{V}_i | S_t^-). \quad (10b)$$

In taking the expectation in (10a) we have treated $t \equiv (m, n)$ as a randomly chosen raster point for which $\bar{P}(q) \equiv E\{P(q; t)\}$ of eq. (1) applies. Because the q are selected independently, $\bar{P}(\mathbf{V}_i | S_t^-)$ of (10b) is related to $\bar{P}(q)$ by

$$\bar{P}(\mathbf{V}_i | S_t^-) = \frac{P(S_t^- | \mathbf{V}_i)}{P(S_t^-)} \bar{P}(\mathbf{V}_i) \quad (11a)$$

$$= \frac{P(S_t^- | \mathbf{V}_i)}{P(S_t^-)} \prod_{j=1}^N \bar{P}(q_{ij}), \quad (11b)$$

where $p(S_t^- | \cdot)$ denotes the probability density function of the vector of values in S_t^- , and q_{ij} is the j th component of \mathbf{V}_i . [The random selection of t cannot affect the independence of the components of \mathbf{V}_i in (11a); hence, (11b)].

Substituting

$$n(q | \mathbf{V}_i) = \sum_{j=1}^N \delta_{q-q_{ij}} \quad (12)$$

into (10b) and summing over i , (10b) becomes

$$\hat{P}(q; t) = \frac{1}{N} \sum_{j=1}^N \bar{P}(q_j = q | S_t^-), \quad (13)$$

where $\bar{P}(q_j = q | S_t^-)$ is the conditional probability that the j th pixel in R_t was generated by source q . Note that (13) gives $\hat{P}(q; t)$ as an average of *a posteriori* probabilities of q over the region R_t , where

$$\bar{P}(q_j = q | S_t^-) = \frac{P(S_t^- | q_j = q)}{P(S_t^-)} \bar{P}(q). \quad (14)$$

We now partition the set of pixels S_t^- into (i) a subset of pixels future to s_j but past to S_t (call it S_t^+); (ii) the pixel s_j ; and (iii) a subset of

pixels S_i^+ past to pixel s_j . The elements in S_j^+ when conditioned on s_j and S_i^+ do not depend upon q_j , and it follows after straightforward manipulations that

$$\bar{P}(q_j = q | S_i^-) = K g(s_i - F_q(S_i^+)) \bar{P}(q); \quad 1 \leq q \leq Q,$$

and

$$\bar{P}(q_j = 0 | S_i^-) = \frac{K}{255} \bar{P}(0), \quad (15)$$

where K is a normalizing constant. Substitution of eqs. (15) and (1) into (10) yields (5).

REFERENCES

1. Z. L. Budrikis, "Visual Fidelity Criterion and Modeling," Proc. IEEE, 60 (July 1972), pp. 771 to 779.
2. R. E. Graham, "Predictive Quantizing of Television Signals," IRE Wescon Convention Record, Part 4, 1958, pp. 142 to 157.
3. J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, 63 (April 1975), pp. 561 to 580.
4. J. Max, "Quantizing for Minimum Distortion," IEEE, Trans. Inform. Theory, IT-21 (July, 1975), pp. 373 to 378.
5. A. N. Netravali and B. Prasada, "Adaptive Quantization of Picture Signals Using Spatial Masking," Proc. IEEE, 65 (April 1977), pp. 536 to 548.

Reduction of Transmission Error Propagation in Adaptively Predicted, DPCM Encoded Pictures

By N. F. MAXEMCHUK and J. A. STULLER

(Manuscript received November 21, 1978)

A new technique for reducing transmission error propagation in adaptively predicted, DPCM-encoded pictures is described. The basis for the technique is a generalization of the notion of predictor output attenuation, described by Graham, to include attenuation of the adaptive prediction function. Simulation results are presented that show that application of the technique to Graham's codec results in significant reduction in error propagation without degradation of picture quality. The technique requires no increase in transmission rate.

I. INTRODUCTION

This paper presents a new and simple technique to reduce error propagation in DPCM image coders that employ adaptive switching-type prediction. An analytical performance description of this technique has not been obtained. However, simulation results using the Graham¹ adaptation algorithm are presented that demonstrate that—in this case, at least—the technique can provide substantial reduction in channel error propagation without decreasing the transmission rate.

The class of coders considered is those which adaptively choose one of Q fixed predictors F_q , $q = 1, 2, \dots, Q$, according to a decision rule that operates on the previously reconstructed pixels in the local past vicinity of the element to be predicted. If x_{ij} is the i th pixel on the j th line of the input raster, and y_{ij} is the vector of reconstructed pixels in the local past vicinity of (i, j) , then the adaptive predictor has the form

$$\hat{x}_{ij} = F(y_{ij}), \quad (1)$$

where $F(\cdot)$ is one of Q fixed functions $F_q(\cdot)$, $q = 1, \dots, Q$, with q chosen according to a decision rule operating on y_{ij} ,

$$q = D(y_{ij}). \quad (2)$$

The encoder and decoder use the same decision rule to determine q , but the decoder must base its decision upon its possibly contaminated version of the reconstructed past scene.

An example of (1) and (2) is given by Graham's predictive encoder. Here $Q = 2$ with

$$\begin{aligned} F_1(y_{ij}) &= y(i-1, j) \\ F_2(y_{ij}) &= y(i, j-1) \end{aligned} \quad (3)$$

and

$$D_{ij} = \begin{cases} 1; & \text{if } |y(i-1, j-1) - y(i, j-1)| \\ & < |y(i-1, j-1) - y(i-1, j)| \\ 2; & \text{otherwise.} \end{cases} \quad (4)$$

It is well known that, for a fixed transmission rate, adaptive prediction generally results in a more accurate coded version of the image, particularly on edges within the picture where large changes in amplitude occur along one dimension. However, a serious problem generally arising from such adaptation is the response of the system to channel errors. Generally, the effect of an error propagates over a larger area of the picture when an adaptive predictor is used than when a fixed predictor is used. This occurs because transmission errors not only (i) contaminate the value of the elements used by the receiver in the function $F(\cdot)$ when the receiver's choice of q is correct, but can also (ii) cause an error in the receiver's choice of q . Note that effect (i) is present in nonadaptive coders and is defined as occurring in adaptive coders when the correct choice of q is made by the decoder. Effect (ii) is unique to adaptive prediction and is potentially more grievous since the transmitter and receiver then use different choices for the prediction function $F(\cdot)$ —a result that, once started, can propagate. An example of the effect of transmission errors in adaptive DPCM is shown in Fig. 1. In this example, Graham's three-bit codec is used over a binary symmetric channel having bit error probability of 10^{-4} . Figure 2 shows the difference between the output of this system with and without transmission errors.

II. PREDICTOR OUTPUT LEAK

As observed by Graham and others, the effect of transmission errors can be reduced by attenuating the predictor output by a constant α , $0 \leq \alpha \leq 1$. In general, a bias term can be introduced so that $\hat{x}_{i,j}$ assumes the form:

$$\hat{x}_{ij} = \alpha F(y_{ij}) + (1 - \alpha)\eta, \quad (5)$$



Fig. 1—Effect of 10^{-4} channel bit error probability on output picture: Graham 3 bit/pel codec.

where η is a constant in the span of possible picture values. A possible choice of η is the mean of x_{ij} ,

$$\eta = E\{x_{ij}\}. \quad (6)$$

Other choices, however, can give subjectively better results depending upon context and system nonlinearities.

Equation (5) can be viewed as a weighted combination of *locally inferred* and *globally given* knowledge about x_{ij} . A value of $\alpha < 1$ has the effect of decreasing the memory of the closed reconstruction loop, and the bias term causes the output to tend toward η . The quantity quantized and transmitted has the form

$$x_{ij} - \hat{x}_{ij} = \alpha[x_{ij} - F(y_{ij})] + (1 - \alpha)[x_{ij} - \eta], \quad (7)$$

which is seen to consist of both DPCM and PCM information. As α varies from one to zero, the system changes from DPCM to PCM. Therefore, attenuation of predictor output as in (5) trades error-propagation attenuation with transmission rate. It should be observed that the technique described by (5) will also reduce error propagation in non-adaptive codecs. Also, the technique does not directly address the problem of the *choice* of q and is therefore a remedy more closely connected to type (i) errors than to type (ii).

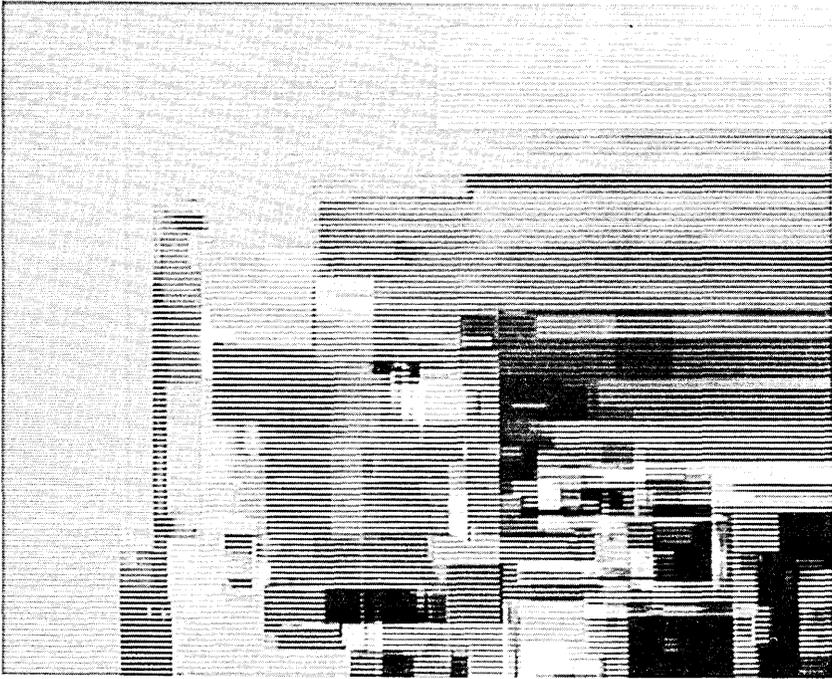


Fig. 2—Difference between output of Graham codec with and without transmission errors.

Using the Graham system, it was found that α could be reduced to $15/16$ before degradation in the output caused by quantization noise becomes visible. The reduction in error propagation resulting from (5) (with η set to 128) is shown in Fig. 3. The difference of the coded picture with and without channel errors is shown in Fig. 4. Although an improvement is obtained with this approach, the next section shows that substantially greater improvement is possible.

III. PREDICTION FUNCTION LEAK

Since the second effect of channel errors in an adaptive codec is to make the value of q uncertain, the receiver loop must in fact estimate the function $F(\cdot)$ as well as x_{ij} . In analogy with (5) we introduce a constant β , $0 \leq \beta \leq 1$, and set (at both transmitter and receiver)

$$\hat{F}(\cdot) = \beta F(\cdot) + (1 - \beta)\bar{F}(\cdot), \quad (8)$$

where $\bar{F}(\cdot)$ is a fixed predictor. A reasonable choice for $\bar{F}(\cdot)$ is the mean of F_q ,

$$\bar{F}(\cdot) = \sum_{q=1}^Q F_q(\cdot)P(q), \quad (9)$$



Fig. 3—Reduction of transmission error propagation using predictor output leak ($\alpha = 1/16$, $\eta = 128$, 10^{-4} bit error probability).

where $P(q)$ is the *a priori* probability of q . Other choices for $\bar{F}(\cdot)$ are possible.

Note that, as β varies from one to zero, a system using $\hat{F}(\cdot)$ will change from fully adaptive DPCM to nonadaptive DPCM. The smaller the value of β , the closer the predictor to being fixed, and the smaller the effect of error propagation due to using the wrong predictor. Note also that (9) is an approach that is applicable only to adaptive codecs. Because of this, we view this technique as a remedy for the second error class (ii) described in Section I.

The concept of prediction function leak has been applied to the Graham predictor and has successfully reduced error propagation. The predictor used in this experiment is:

$$\hat{x}_{ij} = \hat{F}(y_{ij}), \quad (10)$$

where $\hat{F}(\cdot)$ is given by (8) using (3) to (4) and

$$\bar{F}(y_{ij}) = \frac{1}{2}(y(i-1, j) + y(i, j-1)). \quad (11)$$

It was found experimentally that β could be reduced to be between $3/4$ and $1/2$ (depending upon the picture) before the ability of the adaptive predictor to respond to edges within the picture was compromised.

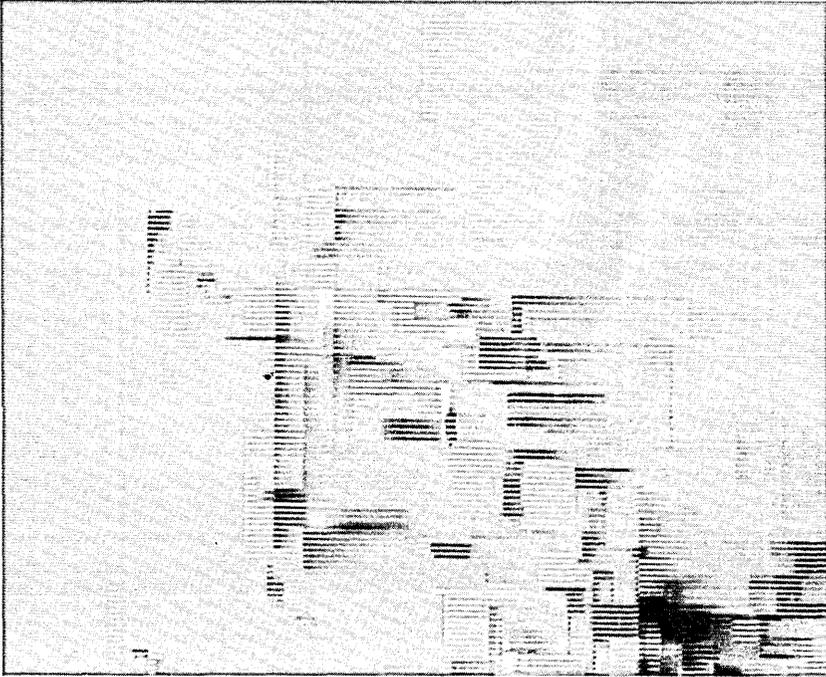


Fig. 4—Difference between output of codec with and without transmission errors ($\alpha = 15/16, \eta = 128$).



Fig. 5—Reduction of transmission error propagation using predictor output and prediction function leak ($\alpha = 15/16, \beta = 3/4, \eta = 128, 10^{-4}$ bit error probability).

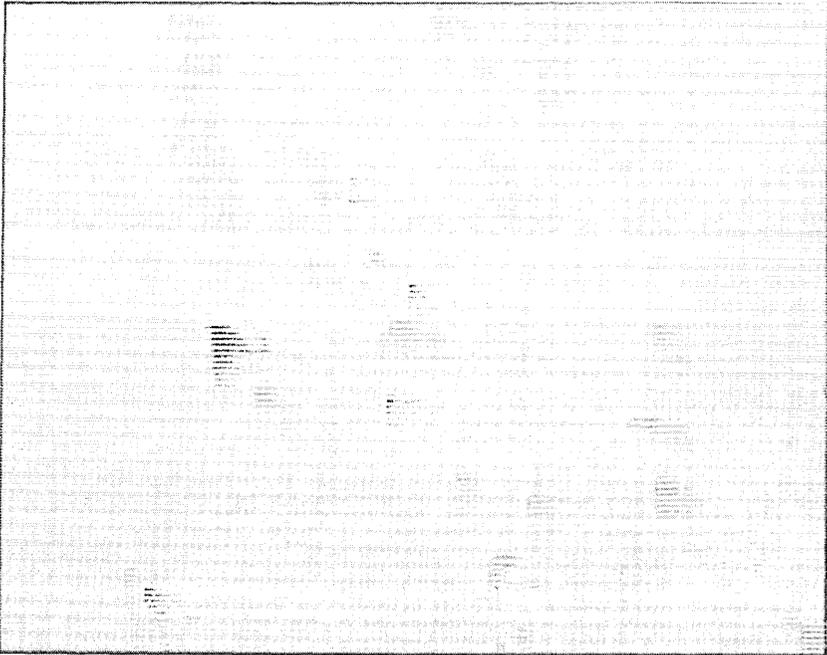


Fig. 6—Difference between output of codec with and without transmission errors ($\alpha = 1/16$, $\beta = 1/4$, $\eta = 128$).



Fig. 7—Reduction of transmission error propagation using predictor output and prediction function leak ($\alpha = 1/16$, $\beta = 1/2$, $\eta = 128$, 10^{-4} bit error probability).

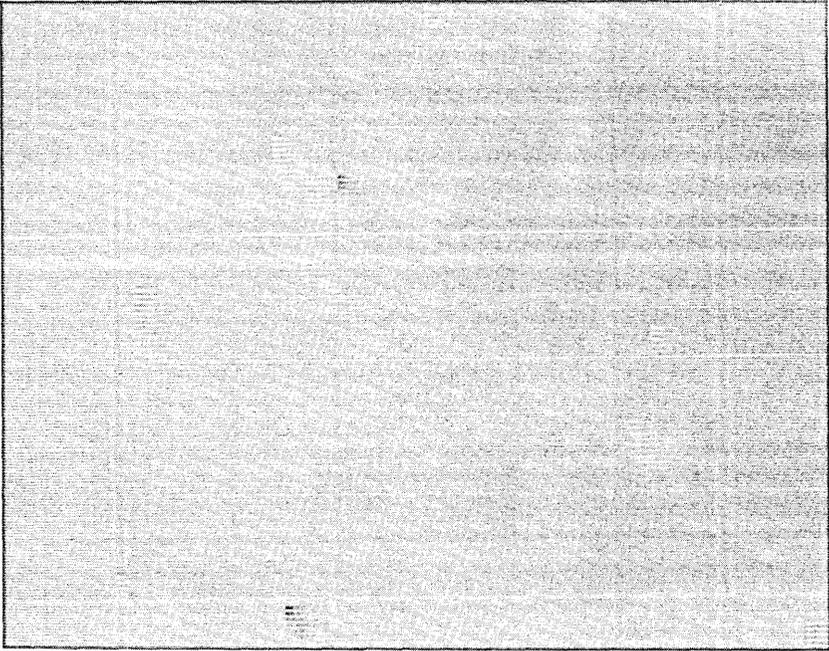


Fig. 8—Difference between output of codec with and without transmission errors ($\alpha = 15/16$, $\beta = 1/2$, $\eta = 128$).



Fig. 9—Quantizing noise of Graham codec.



Fig. 10—Quantizing noise for $\alpha = 15/16$, $\beta = 1$.



Fig. 11—Quantizing noise for $\alpha = 15/16$, $\beta = 3/4$.

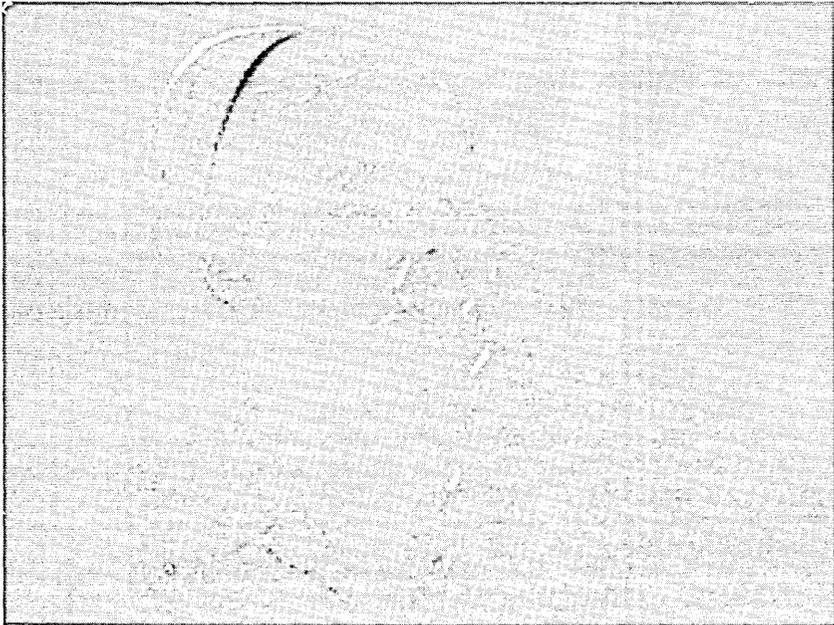


Fig. 12—Quantizing noise for $\alpha = 15/16$, $\beta = 1/2$.

When prediction function leak is used, it is possible to use predictor output leak to further reduce the effects of errors. The estimator is formed as:

$$\begin{aligned}\hat{x}_{ij} &= \alpha \hat{F}(y_{ij}) + (1 - \alpha)\eta \\ &= \alpha\beta F(y_{ij}) + \alpha(1 - \beta)\bar{F}(y_{ij}) + (1 - \alpha)\eta.\end{aligned}\quad (12)$$

Pictures with the same transmission error patterns as those that occurred in the pictures produced using the original Graham predictor were processed using (12). In Fig. 5, $\alpha = 15/16$, $\beta = 3/4$ and $\eta = 128$. Figure 6 is the difference between pictures processed with this predictor with and without transmission errors. In Fig. 7, $\alpha = 15/16$, $\beta = 1/2$, and $\eta = 128$. Figure 8 is the error difference picture of this coder. As seen from these pictures, leaking the prediction function significantly reduces the effect of transmission errors. Analysis of this effect has been hindered by the nonlinear nature of the equations and the fact that the quantities involved exist on a two-dimensional field.

Figures 9 through 12 show the encoding (quantizing) noise for each of the systems previously described in this paper. By comparing Fig. 9 with 11 and 12, it can be seen that, for β equaling $3/4$ and $1/2$, prediction function leak *reduces* encoding noise along edges within the picture. In fact, pictures transmitted with these values of β over an error-free

channel are preferable to those transmitted with the original Graham codec since edge serration, which sometimes occurs when a switching-type predictor is used, is reduced. Figure 10 shows that the introduction of predictor output leak alone does not produce the same beneficial effect.

REFERENCE

1. R. E. Graham, "Predictive Quantizing of Television Signals," IRE Wescon Convention Record, Part 4, 1958, pp. 142 to 157.

Application of Optimization Theory to the Control of the Optical Fiber Drawing Process

By D. H. SMITHGALL

(Manuscript received January 15, 1979)

The optical fiber drawing process is examined and a feedback control loop identified. The incremental dynamic response of each loop component is determined, and the sensitivity of loop response to system parameters is examined. The control loop is optimized, based upon a mean square error criterion with constraints imposed for periodic disturbances. An expression is derived for the effectiveness of the control loop with respect to sources of system disturbance and found to correlate well with experimental results.

I. INTRODUCTION

With the advent of fiber optics technology has come the potential for use of this technology in high quality telecommunications systems. Such systems require sources, detectors, and fibers superior to those used in the present applications. The fibers in high quality systems must have low loss and dispersion and yet be economically produced. One factor influencing transmission loss in the fiber, particularly at splice locations,¹ is the diameter uniformity. Diameter uniformity is directly related to the manufacturing process and is influenced by the environment in which it is drawn² as well as the material from which it is drawn. In addition, large variations in diameter occur during the startup operation, resulting in a material loss of up to 10 percent of the potential fiber.

Much of this wastage can be eliminated and a high degree of fiber uniformity maintained by the judicious design and application of a feedback control on the fiber drawing process. Optimization of such a control requires identification of the distributed, nonlinear drawing process and a quantification of the sensitivities of the process to changes in process parameters.

1.1 The fiber drawing process

Optical fiber is formed by locally and symmetrically heating a cylindrical preform, typically 7 to 25 mm in diameter and 30 and 60 cm in length, to a temperature in the neighborhood of 2000°C. As the preform of diameter D_p is fed into the heat zone at a velocity V_p , the fiber is drawn from the molten material at a velocity V_f , as shown in Fig. 1. Due to the temperatures involved and the tolerances required, the fiber cannot be drawn through a die, and consequently the surface of the molten material is a free boundary whose shape is determined by an equilibrium between the velocity shear gradients and the restraining surface tension. The diameter of the fiber, D_f , is determined, then, by the principle of conservation of mass, which may be written as

$$D_f^2 V_f = D_p^2 V_p + \int_{-\infty}^t dw, \quad (1)$$

where w is a random process representing mechanical and thermally-induced disturbances as well as the variations in diameter which occur while the process is establishing its equilibrium condition. The nature of the disturbing influences is illustrated in Fig. 2, where the fiber diameter is plotted for a 500-m length of fiber with constant D_p , V_p , and V_f . An additional source of diameter variation results from changes in preform diameter which are of a slowly varying nature. Once the process "equilibrium" has been established, the noise (diameter vari-

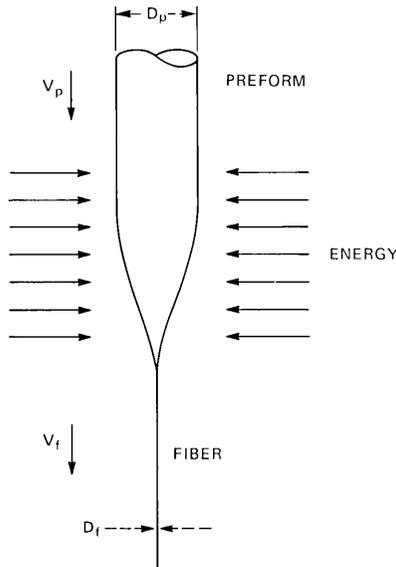


Fig. 1—Geometry of the optical fiber drawing process.

ation) which appears on the fiber can be characterized as band-limited noise riding on a slowly varying bias.

1.2 Identification of process dynamics

Since the fiber diameter is related directly to the two manipulable variables V_f and V_p by the mass conservation principle, and only indirectly to other manipulable variables such as heat source temperature or heat flux, these are the system variables through which a control signal can most effectively be coupled into the process. Experimental results show that the dynamic response of the fiber is two orders of magnitude faster with respect to the drawing velocity than to the feed velocity. Consequently, the control loop shown in Fig. 3 has been determined to be the most effective means of controlling fiber diameter.

The motor-drawing mechanism is typically a pinch-wheel device in which one wheel is driven by a DC motor. The dynamic response of the mechanism used in this investigation is modeled as

$$\frac{V_f(s)}{u(s)} = \frac{0.31 \text{ meters/second}}{0.0045s^2 + 0.030s + 1 \text{ volt}}, \quad (2)$$

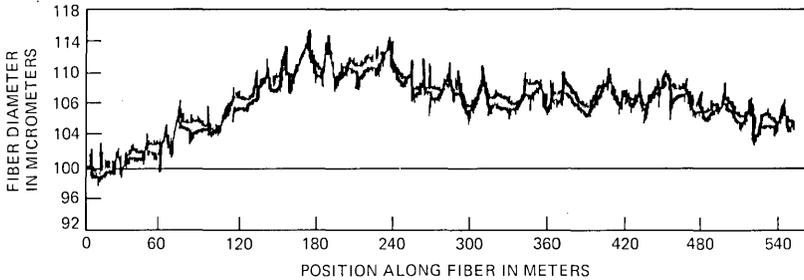


Fig. 2—Diameter profile of uncontrolled fiber with constant feed and drawing speeds.

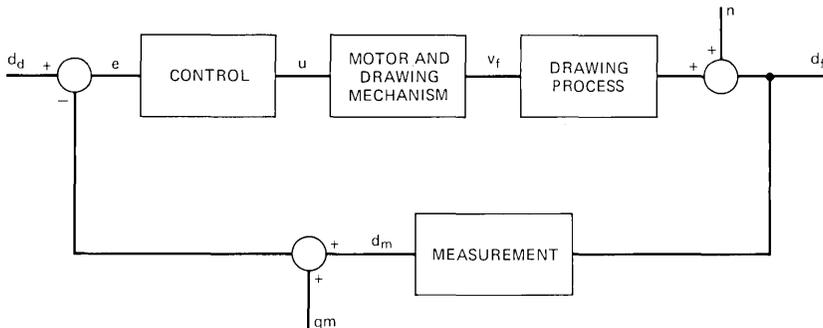


Fig. 3—Block diagram of diameter feedback control loop.

where u is the control signal to the motor. The response of the motor-drawing mechanism is dominated by the inertia of the pinch wheels, while the drawing tension on the fiber has little effect on the response of this component.

The drawing process is a nonlinear, distributed process. Since the diameter is to be regulated about a fixed set point, a linear perturbation model can be used to represent incremental system dynamics. The parameters for this model, as well as the model structure, must be determined experimentally. To obtain response characteristics, the drawing mechanism is excited with a sinusoidal perturbation and the change in draw speed and fiber diameter are measured. Using the technique of Fourier filtering,³ the relative gain and phase of the drawing mechanism and the process diameter response at each excitation frequency can be determined, and a model can be constructed from the data. It is found that, for nominal velocities up to 1 m/s, the process can be modeled by

$$d_f(s) = \frac{D_f/2V_f}{a_p s^2 + b_p s + 1} v_f(s) + n(s), \quad (3)$$

where d_f and v_f now represent incremental changes in fiber diameter and velocity, D_f and V_f represent the nominal process values, and a_p , b_p are the parameters representing process dynamics. The source of diameter variations, n , is a band-limited, Gaussian process with $E\{n\} = 0$, $E\{n^2\} = \sigma_n^2$. The parameters a_p and b_p are sensitive to certain process parameters and insensitive to others. They have been found to be insensitive to preform diameter, fiber diameter, draw velocity, and the temperature of the heat source over the ranges

$$\begin{aligned} 1950^\circ\text{C} &< T_s < 2150^\circ\text{C} \\ 7 \text{ mm} &< D_p < 19 \text{ mm} \\ 0 \text{ m/s} &< V_f < 1 \text{ m/s} \\ 80 \text{ }\mu\text{m} &< D_f < 125 \text{ }\mu\text{m}. \end{aligned}$$

The response is generally insensitive to changes in preform feed velocity over the range of speeds commensurate with the above values of D_f , V_f , D_p , but is known to change for draw velocities higher than 1 m/s. The dynamic response parameters are also sensitive to the length of the heat zone as shown in Fig. 4. The long heat zone was obtained in a furnace and the short heat zone with laser heat source.⁴

Due to physical limitations, the fiber diameter is measured at some point below the heat zone. Using a forward scattering interference fringe counting technique,⁵ the fiber can be measured with an accuracy of 0.25 μm at a rate of 1000 measurements per second. Due to the high measurement rate, the measurement process has no dynamic response to contribute to the loop dynamics. As a result of the digital fringe

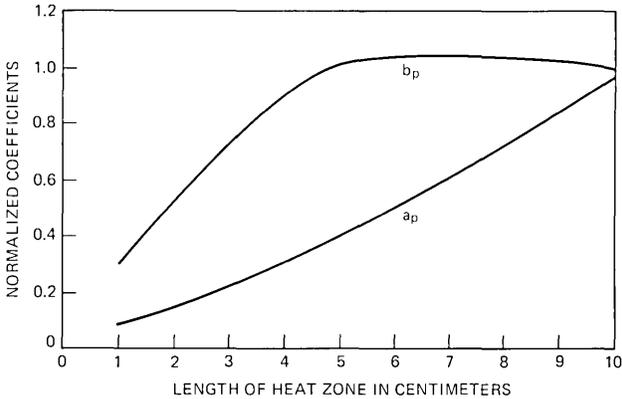


Fig. 4—Variation of process parameters with the length of heat zone.

counting technique, however, a quantization noise, q_m , perturbs the control loop. The effect of q_m can be seen from Fig. 3 to be equivalent to a loop with perfect measurement and a set point which changes $\pm q_m$ at random times. The effect of the noise on the fiber diameter depends upon the rate at which the step changes occur. Experience has shown that the quantization noise becomes significant only when the standard deviation of other system disturbances have been reduced to the level q_m and represents a lower bound of achievable performance for the drawing system.

Since the measurement process must be physically located beneath the heat source, it must also be located some distance away from the point, or region, where the diameter of the molten zone changes in response to variations in the drawing velocity. The distributed nature of the process leads to the expectation that the response to a step change in V_f would result in a distributed change in the boundary of the molten material, as well as a change in fiber diameter. Experience indicates that small perturbations in draw velocity result in diameter variations in the molten zone which are confined to a very small region, near the point where the fiber is formed. Consequently, a “point” at which the molten material changes diameter in response to changes in draw velocity can be defined, and a measurement time delay, T , results which relates the delay distance and nominal drawing velocity.

The resulting model of the measurement process is

$$\frac{d_m(s)}{d_f(s)} = 0.040 e^{-sT} \text{ volts}/\mu\text{m}, \quad (4)$$

where the quantization noise has been reassociated with the diameter set point. The delay time is typically 0.04 to 0.10 second for a drawing velocity of 1 m/s.

II. DESIGN OF THE CONTROL

The control problem is to reduce the effects of the noise source, n , upon the fiber diameter d_f . The relationship between the two quantities when the feedback loop of Fig. 3 is used is given by

$$\frac{\delta d_f(s)}{n(s)} = \frac{1}{1 + G(s)} = H(s), \quad (5)$$

where $\delta d_f(s) = d_f(s) - d_a$, and $G(s)$ is the forward transfer function

$$G(s) = \left. \frac{d_f(s)}{v_f(s)} \right|_{n=0} \times \frac{v_f(s)}{u(s)} \times \frac{u(s)}{d_m(s) - d_d} \times \frac{d_m(s) - d_d}{d_f(s)}. \quad (6)$$

The form of the control circuit is

$$\frac{u(s)}{d_m(s) - d_d} = \frac{K_c a_c s + 1}{s b_c s + 1}. \quad (7)$$

The integral term is required to remove the slowly varying components of the noise source, and the lead-lag term is used to shape the response curve. Using this control circuit and the component responses (2) to (4), the response curve

$$\left| \frac{\delta d_f(s)}{n(s)} \right|_{s=j\omega}^2$$

as a function of ω is shown in Fig. 5. The response curve shows that low frequency disturbances can be effectively suppressed. There is a loss of control effectiveness as frequency increases to a point where the control loop has no effect upon the disturbances. In the region of the corner frequency ω_c , the noise is amplified. Figures 5 and 6 illustrate the role of the control circuit parameters upon the response curve. The suppression of low frequency variations is affected only by the loop gain. The degree of response peaking around the corner frequency is affected by the gain as well as the lead-lag network parameters. Some peaking resulting from high loop gain must be allowed, since it can only partially be compensated for by adjustment of a_c and b_c . It has been found that +2 dB is an acceptable peak gain for the response curve, from the standpoint of the stochastic disturbances.

The performance of the control loop can be optimized by choice of the parameters K_c , a_c , b_c to minimize the performance index

$$J = \int_0^{\infty} |H(j\omega)|^2 d\omega, \quad (8)$$

subject to the constraint

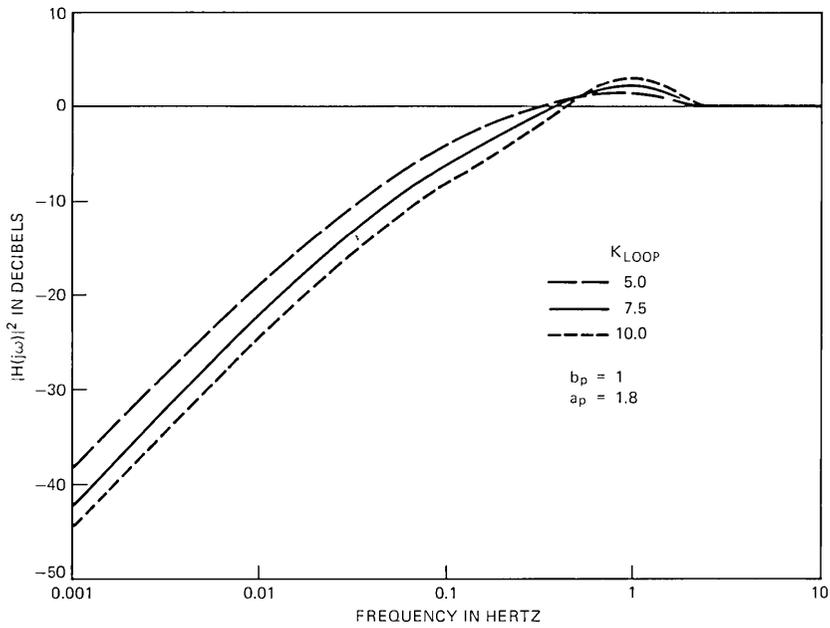


Fig. 5—Sensitivity of control loop performance to loop gain.

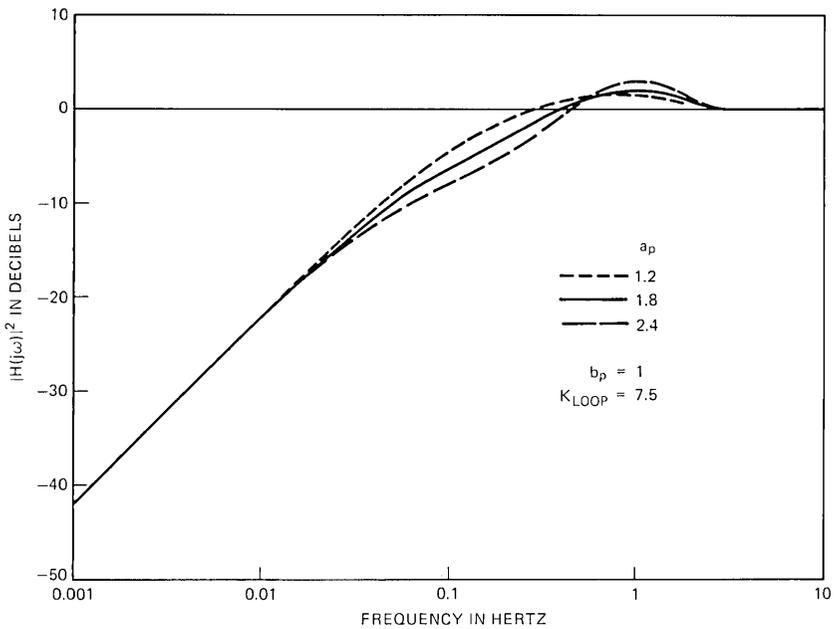


Fig. 6—Sensitivity of control loop performance to lead-lag compensation.

$$\left\{ \omega = \omega_0: 0 < \omega < \infty \right\} |H(j\omega)|^2 \leq 2 \text{ dB.} \quad (9)$$

The performance index (8) is a measure of the relative noise power transmitted to the fiber diameter through the control loop. Minimization of (8) is equivalent to minimizing

$$J_1 = E \left\{ \delta d_f^2(t) \right\}, \quad |n(\cdot)| = 1, \quad (10)$$

for $n(\cdot)$ modeled by white noise. This relationship is established by the relationship between the power spectral densities

$$\Gamma_{d_f} = |H(j\omega)|^2 \Gamma_n, \quad (11)$$

resulting in

$$J_1 = \int_{-\infty}^{\infty} \Gamma_{d_f} d\omega = 2\sigma_n^2 J. \quad (12)$$

Due to the constraint (9), which seeks to reduce the effect of periodic disturbances upon the loop design, optimization is most effectively performed in the frequency domain.

The optimum performance for a typical set of drawing system parameters is shown in Fig. 7, for the case of two measurement delay times. Referring to Figs. 5 to 7, of all the system parameters, the effectiveness of the control loop is most sensitive to loop gain and the measurement delay time. Examination of the phase characteristics of the system components reveals that over the range of controllable disturbances the measurement delay contributes the largest phase lag to the loop dynamics. Consequently, it is imperative in drawing system design to minimize the measurement delay. The loop response is sensitive to loop gain because the destabilizing tendencies of the measurement delay pull the root loci toward the $j\omega$ axis.⁶

To determine the effectiveness of the control, account must be taken of both the process-related noise, n , and the measurement quantization noise q . The fiber diameter variation is related to these two quantities in the frequency domain by the expression

$$\delta d_f(s) = H(s)n(s) + G(s)H(s)q(s). \quad (13)$$

The power spectral densities are related by

$$\Gamma_{d_f}(\omega) = |H(j\omega)|^2 \Gamma_n(\omega) + |F(j\omega)|^2 \Gamma_q(\omega), \quad (14)$$

where $F(s) = G(s)H(s)$. The noise generating process associated with the fiber drawing process has been experimentally determined and can be modeled as

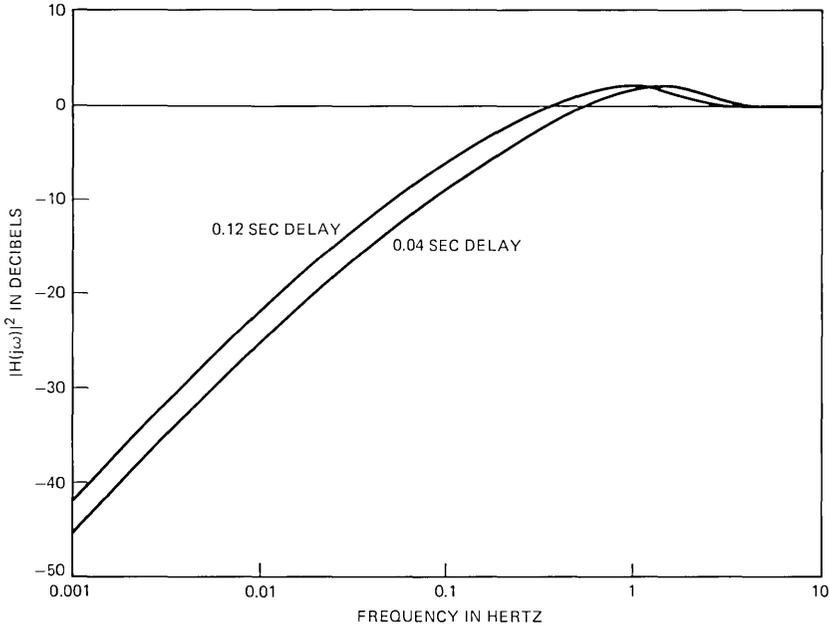


Fig. 7—Sensitivity of control loop performance to measurement time delay. System parameter values are $a_p = 0.0032$, $b_p = 0.080$.

$$\Gamma_n(\omega) = \frac{\frac{\omega_n}{\pi} \sigma_n^2}{\omega^2 + \omega_n^2}, \quad (15)$$

where ω_n is the corner frequency of the noise spectrum.

The quantization noise can be modeled as a random telegraph signal⁷ for which the power spectral density function is

$$\Gamma_q(\omega) = \frac{\lambda q_m^2}{\lambda^2 + \pi^2 \omega^2}, \quad (16)$$

where λ is the mean number of switchings per unit time. The control function can be approximated as

$$\begin{aligned} |H(j\omega)|^2 &= \frac{\omega^2}{\omega^2 + \omega_c^2} \\ |F(j\omega)|^2 &= \frac{\omega_c^2}{\omega^2 + \omega_c^2}, \end{aligned} \quad (17)$$

where ω_c is the corner frequency of the closed loop transfer function. The variance of the fiber diameter is then determined by computing the auto-covariance function⁸ with zero lag:

$$\sigma_d^2 = \frac{1}{1 + \frac{\omega_c}{\omega_n}} \sigma_n^2 + \frac{1}{1 + \frac{\lambda}{\pi\omega_c}} q_m^2. \quad (18)$$

Equation (18) illustrates an interesting tradeoff. Whereas it is desirable to reduce the process-induced fiber diameter variation by increasing ω_c insofar as possible, it is done at the expense of allowing additional quantization noise to affect the fiber diameter. The latter source may only be reduced by increasing the resolution of the measurement system.

Experimentally, the effectiveness of the control loop is illustrated in Fig. 8 by the distribution of mean and standard deviations measured on 500-m lengths of controlled and uncontrolled fibers, after the process has reached equilibrium. The mean diameter of the uncontrolled fibers varies over a 4- μm range due to variations in preform diameter, nominal drawing velocities, etc. Using the feedback control, the mean diameter is held to within 0.1 μm of the set point. This deviation from set point is due, in part, to the 0.25 μm measurement resolution. The standard deviations of uncontrolled fibers are distributed over a wide range with the majority of the samples having a

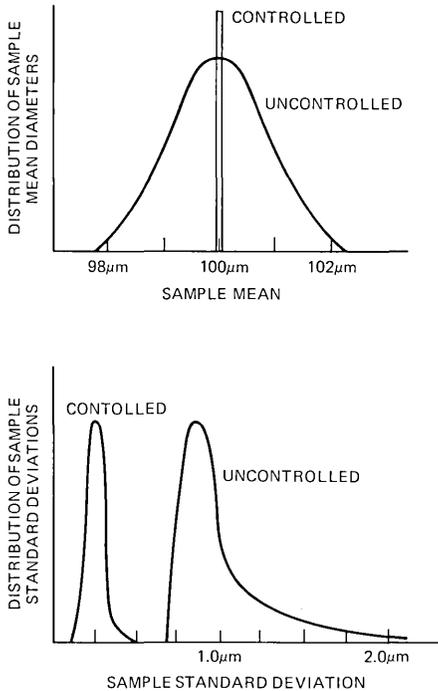


Fig. 8—Measured statistical characteristics of controlled and uncontrolled fibers.

standard deviation of $0.8 \mu\text{m}$. Use of the feedback control improves fiber quality with the result that standard deviations below $0.30 \mu\text{m}$ can be repeatably attained.

Using experimentally determined system parameters, the controlled variance of the fiber diameter given in eq. (18) can be compared to experimental values. If $\omega_n = 0.02$, $\sigma_n = 0.8 \mu\text{m}$, $q_n = 0.25 \mu\text{m}$, $\lambda = 5$, and from Fig. 7 $\omega_c = 0.2$, eq. (18) yields a theoretical standard deviation of $0.27 \mu\text{m}$. This value is very close to the median of the experimentally measured sample standard deviations.

III. CONCLUSION

Optimization theory has been applied to the optical fiber drawing process, resulting in a diameter feedback control loop which effectively reduces fiber diameter variations. In addition to examining the sensitivities of control loop performance to system component parameters, the sensitivities with respect to the noise sources were also examined. The expression derived to describe system performance with respect to process-induced diameter variation and measurement quantization noise showed good agreement with experimental results.

REFERENCES

1. C. M. Miller, "Transmission vs. Transverse Offset for Parabolic—Profile Fiber Splices with Unequal Core Diameters," *B.S.T.J.*, 55, No. 7 (September 1976), pp. 917–927.
2. M. Nakahara et al., "Drawing Techniques for Optical Fibers," Review of the Electrical Communication Laboratories, 26, No. 3–4 (March–April 1978), pp. 476–483.
3. P. Eykhoff, *System Identification*, New York: John Wiley & Sons, 1974, pp. 378–382.
4. R. E. Jaeger, "Laser Drawing of Glass Fiber Optical Waveguides," *Ceramic Bulletin*, 55, No. 3 (1976), pp. 270–273.
5. D. H. Smithgall, L. S. Watkins, and R. E. Frazee, "High Speed Noncontact Fiber Diameter Measurement Using Forward Light Scattering," *Appl. Opt.*, 16, No. 9 (September 1977), pp. 2395–2402.
6. K. Ogata, *Modern Control Engineering*, Englewood Cliffs: Prentice-Hall, 1970, pp. 346–350.
7. W. B. Davenport and W. L. Root, *Random Signals and Noise*, New York: McGraw-Hill, 1958, pp. 61, 104.
8. G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*, New York: Holden-Day, 1968, Ch. 6.

On the Structure of Real-Time Source Coders

By H. S. WITSENHAUSEN

(Manuscript received January 30, 1979)

The outputs of a discrete time source with memory are to be encoded ("quantized" or "compressed") into a sequence of discrete variables. From this latter sequence, a receiver must attempt to approximate some features of the source sequence. Operation is in real time, and the distortion measure does not tolerate delays. Such a situation has been investigated over infinite time spans by B. McMillan. In the present work, only finite time spans are considered. The main result is the following. If the source is k th-order Markov, one may, without loss, assume that the encoder forms each output using only the last k source symbols and the present state of the receiver's memory. An example is constructed, which shows that the Markov property is essential. The case of delay is also considered.

I. INTRODUCTION

The outputs of a discrete time source with memory are to be encoded ("quantized" or "compressed") into a sequence of discrete variables. From this latter sequence, a receiver must attempt to approximate some features of the source sequence. Operation is in real time, and the distortion measure does not tolerate delays. Such a situation has been investigated over infinite time spans in Ref. 1. In the present work, only finite time spans are considered.

The main result is the following. If the source is k th-order Markov, one may, without loss, assume that the encoder forms each output using only the last k source symbols and the present state of the receiver's memory.

An example is constructed, which shows that the Markov property is essential.

II. THE MODEL

2.1 The causal structure

A source produces a random sequence X_1, X_2, \dots, X_T where for each $t \in \{1, \dots, T\}$, X_t is a vector in n_t -dimensional real space. The source

is characterized by the sequence distribution: A given probability measure on the Borel sets of the product space of dimension $\sum_{t=1}^T n_t$.

For each t , there is an opportunity for noiseless transmission of a signal Y_t taking q_t possible values. This signal is produced from the X sequence by an encoder. As we consider the problem in real time, causality allows the encoder at t to see only the values X_1, \dots, X_t . The encoders are thus characterized by functions $f_t: R^{n_1+\dots+n_t} \rightarrow \{1, \dots, q_t\}$, Borel measurable, $t = 1, \dots, T$.

At the receiving end, the most that could be accessible at stage t is the subsequence Y_1, \dots, Y_t . However, we also want to consider the case of limited memory, as the receiver might not be able to store this whole sequence for large t . The model will be the following.

At $t = 1$, only Y_1 is available, and a discrete variable $Z_1 = r_1(Y_1)$ taking m_1 values is stored in memory. For each $t > 1$, the memory is updated by

$$Z_t = r_t(Z_{t-1}, Y_t), \quad t = 2, \dots, T - 1,$$

where Z_t takes values in $\{1, \dots, m_t\}$ and

$$r_t: \{1, \dots, m_{t-1}\} \times \{1, \dots, q_t\} \rightarrow \{1, \dots, m_t\}$$

is the memory update function.

The purpose of the receiver is to generate a variable V_t in R^{s_t} by

$$V_1 = g_1(Y_1),$$

where $g_1: \{1, \dots, q_1\} \rightarrow R^{s_1}$,
and for $t > 1$

$$V_t = g_t(Z_{t-1}, Y_t),$$

where

$$g_t: \{1, \dots, m_{t-1}\} \times \{1, \dots, q_t\} \rightarrow R^{s_t}.$$

The interpretation of V_t is that it represents an approximation to something we wish to know at the receiving end about X_t . In particular, one may have $s_t = n_t$ and consider V_t as approximating X_t itself.

The functional relationships described above are symbolized in Fig. 1.

The case of *full receiver memory* is included in this model. One need only identify Z_t with (Y_1, \dots, Y_t) and r_t with the concatenation function "append."

Furthermore, in this case,

$$m_t = \prod_{k=1}^t q_k.$$

2.2 The criteria

The performance of the system is defined by way of a sequence of distortion functions. For each t , a Borel measurable function

$$\psi_t: R^{n_t} \times R^{s_t} \rightarrow [0, \infty)$$

is given. Then

$$J_t = E\{\Psi_t(X_t, V_t)\}$$

measures the distortion at stage t . It is possible that J_t be infinite, but it is always well defined, as the composition of Borel functions is Borel and $\Psi_t \geq 0$.

2.3 The optimization problem

The problem to be considered is the following:

Given: The integers: $T; n_1, \dots, n_T; q_1, \dots, q_T; m_1, \dots, m_T; s_1, \dots, s_T$.

The distribution of the X sequence.

The distortion measures Ψ_1, \dots, Ψ_T .

Choose: The functions $f_1, \dots, f_T; g_1, \dots, g_T; r_1, \dots, r_T$ (the latter are redundant in the full memory case, i.e., when $m_t \geq q_1 q_2 \dots q_t$ for all t). A choice of a system of such functions will be called a "design."

In order to:

Minimize (exactly or within ϵ) the sum

$$J = \sum_{t=1}^T J_t.$$

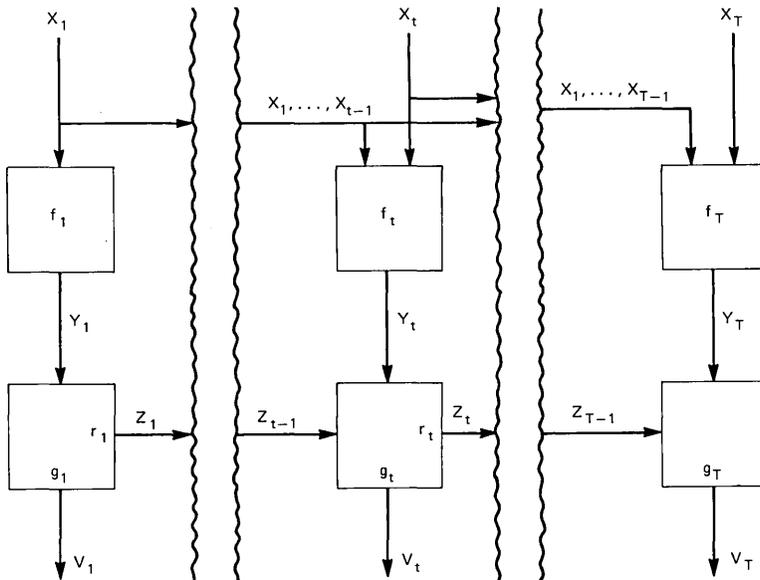


Fig. 1—General system.

Remark that nothing would be gained by having J as a nonnegative linear combination $\sum c_t J_t$ (for instance, with $c_t = e^{-\lambda t}$, a discount factor) because such $c_t \geq 0$ can simply be absorbed into the definition of Ψ_t .

It should be said that the freedom of having $n_t, q_t, m_t, s_t, \psi_t$ depend upon t is not a matter of extra generality, but is essential to the proof techniques used in the sequel.

A design producing the values (J_1, \dots, J_T) is *at least as good* as a design producing (J'_1, \dots, J'_T) when $J_t \leq J'_t$ for all $t \in \{1, \dots, T\}$. This, of course, implies the much weaker statement that $J = \sum J_t \leq J' = \sum J'_t$.

A design may exist which is at least as good as any other; it is called a *dominant* design. In general, however, no dominant design exists because the set in R^T of achievable vectors (J_1, \dots, J_T) does not have a corner (J_1^*, \dots, J_T^*) such that all other points of this set lie in the shifted orthant $J_t \geq J_t^*, t = 1, \dots, T$. Instead, the set may have a *Pareto frontier* of "admissible" vectors, i.e., vectors (J_1, \dots, J_T) such that no vector (J'_1, \dots, J'_T) is achievable that has $J'_t \leq J_t$ for all t with strict inequality for some t .

2.4 Special encoder structures

The encoder f_t at a specific stage $t > k$ is said to have *memory structure of order k* , if there is a Borel function

$$\hat{f}_t: \{1, \dots, m_{t-1}\} \times R^{n_{t-k+1} + \dots + n_t} \rightarrow \{1, \dots, q_t\}$$

such that

$$f_t(X_1, \dots, X_t) = \hat{f}_t(Z_{t-1}, X_{t-k+1}, \dots, X_t) \quad \text{a.s.}$$

This is equivalent to the assertion that Y_t is measurable on the σ -field generated by $Z_{t-1}, X_{t-k+1}, \dots, X_t$. In other words, the encoder elaborates Y_t using only the k most recent source outputs X_{t-k+1}, \dots, X_t and the receiver's current memory Z_{t-1} .

III. THE MAIN THEOREM

The sequence X_1, X_2, \dots, X_T is said to be k th-order Markov, when, given any block of k consecutive X_t , the parts of the sequence preceding and following this block are conditionally independent. For $k = 1$, this is the ordinary Markov property. Note that the k th-order Markov property holds in a vacuous way if $T < k + 2$.

Most sequences can be approximated by k th-order Markov sequences for sufficiently large k . If this k is small compared to T , then the following main theorem provides a substantial simplification of the encoder optimization problem.

Theorem 1: Suppose the source is k th-order Markov. Then, given any design, there is another design with the following properties:

(i) *The new design differs from the given one only in the choice of encoders.*

(ii) *All encoders of the new design have memory structure of order k . (The last encoder f_T can even be made to have memory structure of order 1.)*

(iii) *The performance index J of the new design does not exceed the index of the old design.*

Postponing the proof of Theorem 1 to Section V, we comment here on its significance. It says, in particular, that for a Markov source and a receiver with perfect memory, one need only consider encoders which generate each code symbol Y_t using only the current source symbol X_t and the past code sequence Y_1, Y_2, \dots, Y_{t-1} . This result is essentially dependent on the Markov property of the source as can be seen from the following example.

Take $T = 3$ and, for $t = 1, 2, 3$, let $n_t = s_t = 1$, $q_t = 2$, $\psi_t(X_t, V_t) = (X_t - V_t)^2$. Suppose the receiver has perfect memory. Suppose that the source sequence (X_1, X_2, X_3) takes just eight equally probable values, namely $(13, 1, 3)$, $(12, 1, 2)$, $(11, 1, 1)$, $(10, 1, 0)$, $(-10, -1, 0)$, $(-11, -1, 1)$, $(-12, -1, 2)$, $(-13, -1, 3)$.

At the first stage, if one considers only the minimization of J_1 , one has a classical quantization problem for X_1 . As X_1 takes its values in two separate equiprobable clusters, the minimum of J_1 is attained by letting Y_1 signal the sign of X_1 to identify the cluster. Then $V_1 = \pm 11.5$ and $J_1 = 1.25$. Any other choice of the first encoder yields a strictly larger value for J_1 . Furthermore, Y_1 is already sufficient for the attainment of $J_2 = 0$, the second-stage receiver need not even look at Y_2 . However, Y_2 can be used to help the third-stage receiver. If one lets Y_2 signal the parity of X_1 , then $J_3 = 0$ is attainable by letting Y_3 signal whether $|X_3| \leq 1$ or not.

The design so obtained minimizes J_t for each t (it is "dominant"); *a fortiori*, it minimizes $J = \sum_1^3 J_t$, giving $J = 1.25$. However, the second-stage encoder does not have memory structure of order one.

Is it possible to achieve $J = 1.25$ with memory structure of order one although the source is not Markov? The answer is no for, if one changes the first-stage encoder, this alone will drive J_1 and, *a fortiori*, J above 1.25. But if the first encoder signals the sign of X_1 and the second encoder must have first-order structure, then the second encoder is useless. Indeed, X_2 contains no information not already contained in Y_1 , and the receiver remembers Y_1 . Now Y_1 is useless to the third-stage receiver,† and a single binary signal Y_3 is insufficient to distinguish among the four possible values of X_3 . The best that can be

† $Y_1 = \text{sgn } X_1$ and X_3 are independent.

done is to form Y_3 as in the previous design, giving $J_3 = 0.25$; hence, $J = 1.5$.

The optimum design requires encoder f_2 to “signal ahead” features of X_1 for the later benefit of receiver g_3 . This phenomenon is ruled out for sources with the Markov property.

IV. TWO BASIC LEMMATA

All the results in this paper will be derived from two basic lemmata: a “two-stage lemma” and a more complex “three-stage lemma.” Once these are obtained, the use of induction and of the technique of “repackaging” random variables will suffice.

4.1 The two-stage lemma

This lemma uses what is, in fact, the basic line of reasoning in Ref. 1. Consider a system with $T = 2$ and any joint distribution of the pair of random vectors (X_1, X_2) . Observe that the content Z_1 of the receiver’s memory at the beginning of stage 2 is a certain function of X_1 ; that is,

$$Z_1 = \phi(X_1),$$

where ϕ is a Borel function (in fact, it is the composition of f_1 and r_1). The second (and last) stage is characterized by the functions f_2 and g_2 with (Fig. 2)

$$Y_2 = f_2(X_1, X_2),$$

$$V_2 = g_2(Z_1, Y_2).$$

Lemma 1: Given a two-stage system with a design in which f_2 does not have memory structure of order 1, one can change f_2 (and only f_2) so that it has this structure and the new design is at least as good as the given design.

Proof: If only f_2 is changed, then J_1 , ϕ , and g_2 remain as given. We have to show that, for a suitable change in f_2 , J_2 can only decrease. Consider the function

$$F((Z_1, X_2), Y_2) \equiv \psi_2(X_2, g_2(Z_1, Y_2)).$$

As Y_2 is discrete and F is measurable (by its construction), a measurable function \hat{f}_2 exists (see the appendix) such that

$$F((Z_1, X_2), \hat{f}_2(Z_1, X_2)) \leq F((Z_1, X_2), Y_2)$$

for all values of Z_1, X_2, Y_2 . Hence, by the substitution

$$Z_1 = \phi(X_1)$$

$$Y_2 = f_2(X_1, X_2)$$

$$\psi_2(X_2, g_2(\phi(X_1), \hat{f}_2(\phi(X_1), X_2))) \leq \psi_2(X_2, g_2(\phi(X_1), f_2(X_1, X_2)))$$

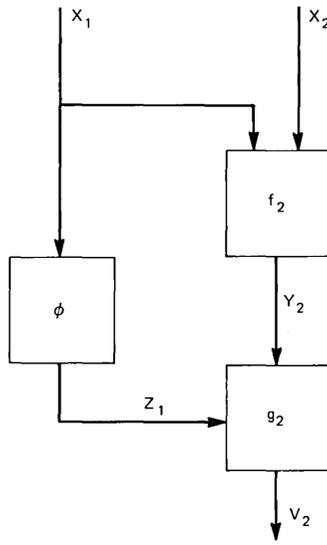


Fig. 2—Two-stage lemma.

holds for all X_1, X_2 . As the functions ϕ_1, f_2 and \hat{f}_2 are measurable, both sides of the inequality are measurable. Since they are also nonnegative, the inequality persists when taking the expectation of both sides, whether finite or not. This establishes that J_2 can only decrease as claimed.

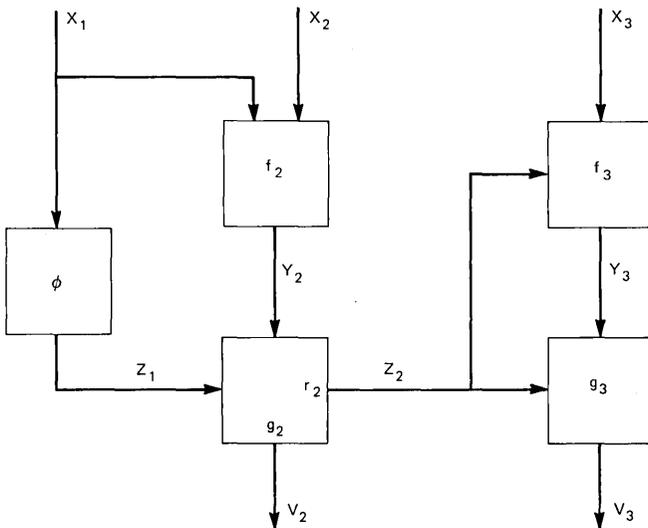


Fig. 3—Three-stage lemma.

4.2 The three-stage lemma

Consider a three-stage system ($T = 3$) with a Markov source. Assume that the last encoder f_3 already has first-order memory structure, while f_2 does not (Fig. 3).

Lemma 2: Under the above assumptions, one can replace f_2 by an encoder \hat{f}_2 having memory structure of order one, without increasing the total cost $J = J_1 + J_2 + J_3$.

Proof: The first-stage cost J_1 is unaffected by changes in f_2 and the effect of the first-stage design is to generate the receiver memory Z_1 as a certain measurable function $Z_1 = \phi(X_1)$, where ϕ is the composition $r_1 \cdot f_1$. By assumption, f_3 can be written in the form

$$Y_3 = f_3(Z_2, X_3)$$

where

$$Z_2 = r_2(Z_1, Y_2).$$

The cost incurred in the last two stages can thus be written

$$\begin{aligned} \psi_2(X_2, g_2(Z_1, Y_2)) \\ + \psi_3(X_3, g_3(r_2(Z_1, Y_2), f_3(r_2(Z_1, Y_2), X_3))) \\ \equiv F(Z_1, X_2, X_3, Y_2), \end{aligned}$$

defining the measurable function F .

Consider now the conditional expectation

$$E\{F(Z_1, X_2, X_3, Y_2) | X_1, X_2\}.$$

Because X_3 is a finite dimensional random vector, a regular conditional distribution of X_3 exists for any condition. In view of the Markov property, conditioning on the pair (X_1, X_2) is equivalent to conditioning on X_2 only. Let $P(dX_3 | X_2)$ be a regular version of this conditional distribution.

Then the conditional expectation under consideration can be written

$$\int P(dX_3 | X_2) F(Z_1, X_2, X_3, Y_2),$$

where Z_1 and Y_2 , which depend only on the conditioning variables X_1, X_2 , can be treated as fixed. This integral defines a measurable function (nonnegative and possibly extended real-valued)

$$G(Z_1, X_2, Y_2).$$

For any choice of f_2 , the sum $J_2 + J_3$ will be given by the expectation of G . Note that X_1 enters G only by way of Z_1 and Y_2 .

As in Lemma 1, a measurable function \hat{f}_2 exists such that, for all Z_1, X_2 and Y_2 ,

$$G(Z_1, X_2, \hat{f}_2(Z_1, X_2)) \leq G(Z_1, X_2, Y_2).$$

Substituting $Z_1 = \phi(X_1)$, $Y_2 = f_2(X_1, X_2)$ and taking the expectations of both sides of this inequality, implies, by the chain rule, that $J_2 + J_3$ cannot increase when f_2 is replaced by \hat{f}_2 .

V. PROOF OF THE MAIN THEOREM

To begin with, the situation of the last stage is always a special one, as the following lemma shows.

Lemma 3: For any source statistics and any design, one can replace the last encoder by one having memory structure of order one, without performance loss.

Proof: The given T -stage system can be considered as a two-stage system, by setting

$$\begin{aligned}\bar{X}_1 &= (X_1, X_2, \dots, X_{T-1}) \\ \bar{X}_2 &= X_T \\ \bar{Z}_1 &= Z_{T-1} = \phi(\bar{X}_1) \\ \bar{Y}_2 &= Y_T \\ \bar{f}_2(\bar{X}_1, \bar{X}_2) &= f_T(X_1, X_2, \dots, X_{T-1}, X_T) \\ \bar{g}_2(\bar{Z}_1, \bar{Y}_2) &= g_T(Z_{T-1}, Y_T) \\ \bar{V}_1 &= (V_1, \dots, V_{T-1}) \\ \bar{V}_2 &= V_T \\ \bar{\psi}_1(\bar{X}_1, \bar{V}_1) &= \sum_{t=1}^{T-1} \psi_t(X_t, V_t) \\ \bar{\psi}_2(\bar{X}_2, \bar{V}_2) &= \psi_T(X_T, V_T),\end{aligned}$$

which amounts to a change in notation. Of course,

$$\bar{n}_1 = \sum_{t=1}^{T-1} n_t,$$

a substantial increase in dimension.

By Lemma 1, there exists an encoder \hat{f}_2 which has the structure

$$\bar{Y}_2 = \hat{f}_2(\bar{Z}_1, \bar{X}_2)$$

and whose use does not increase \bar{J}_2 . Reverting to the original notation, this corresponds to an encoder \hat{f}_T with the structure

$$Y_T = \hat{f}_T(Z_{T-1}, X_T)$$

whose use does not increase J_T . As the other J_t are unchanged, the lemma is proved.

The above fact is the starting point for the proof of the main theorem with $k = 1$.

Lemma 4: The main theorem holds for $k = 1$. That is, for a Markov source and any design, one can replace the encoders by appropriate encoders having first-order memory structure without increase in the expected cost J .

Proof: Using backward induction, one can first replace f_T by \hat{f}_T , as in Lemma 3. Now suppose the encoders for stages $t + 1, t + 2, \dots, T$ already have memory structure of order one. It must be shown that f_t can be replaced by \hat{f}_t with such structure, without increase in expected total cost. To this effect, the T -stage system can be considered as a three-stage system, in which the third stage has first-order memory structure and the source is Markov, as follows. Set

$$\tilde{X}_1 = (X_1, \dots, X_{t-1}) \left(\text{thus, } \tilde{n}_1 = \sum_{i=1}^{t-1} n_i \right)$$

$$\tilde{Z}_1 = Z_{t-1} = \phi(\tilde{X}_1)$$

$$\tilde{X}_2 = X_t$$

$$\tilde{Y}_2 = Y_t$$

$$\tilde{Z}_2 = Z_t = \tilde{r}_2(\tilde{Z}_1, \tilde{Y}_2) = r_t(Z_{t-1}, Y_t)$$

$$\tilde{V}_2 = V_t = \tilde{g}_2(\tilde{Z}_1, \tilde{Y}_2) = g_t(Z_{t-1}, Y_t)$$

$$\tilde{\psi}_2(\tilde{X}_2, \tilde{V}_2) = \psi_t(X_t, V_t)$$

$$\tilde{X}_3 = (X_{t+1}, \dots, X_T), \quad \left(\tilde{n}_3 = \sum_{i=t+1}^T n_i \right)$$

$$\tilde{Y}_3 = (Y_{t+1}, \dots, Y_T), \quad \left(\tilde{q}_3 = \prod_{i=t+1}^T q_i \right)$$

$$\tilde{V}_3 = (V_{t+1}, \dots, V_T) = \tilde{g}_3(\tilde{Z}_2, \tilde{Y}_3), \quad \left(\tilde{s}_3 = \sum_{i=t+1}^T s_i \right).$$

The latter relation follows from the fact that each $V_\theta, \theta > t$, is given by $g_\theta(Z_{\theta-1}, Y_\theta)$ and the variables $Z_{\theta-1}, Y_\theta$ are known functions of $Z_t, Y_{t+1}, Y_{t+2}, \dots, Y_T$ using the memory update functions. Then

$$\tilde{\psi}_3(\tilde{X}_3, \tilde{V}_3) = \sum_{\theta=t+1}^T \psi_\theta(X_\theta, V_\theta).$$

As the encoders for stages $t + 1, \dots, T$ already have first-order

memory structure, their effect is to define an encoder

$$\tilde{Y}_3 = \tilde{f}_3(\tilde{Z}_2, \tilde{X}_3)$$

because each of the Y_θ , $\theta > t$, included in \tilde{Y}_3 is given by a function $f_\theta(Z_{\theta-1}, X_\theta)$ and the variables $Z_{\theta-1}, X_\theta$ are known functions of \tilde{Z}_2, \tilde{X}_3 ; i.e., of Z_t, X_{t+1}, \dots, X_T using the memory update functions and recursion. The given encoder at stage t has the general form $Y_t = f_t(X_1, X_2, \dots, X_{t-1}, X_t)$ which translates to

$$\tilde{Y}_2 = \tilde{f}_2(\tilde{X}_1, \tilde{X}_2).$$

The new source $(\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$ is Markov since $\tilde{X}_1 = (X_1, \dots, X_{t-1})$ and $\tilde{X}_3 = (X_{t+1}, \dots, X_T)$ are conditionally independent given $\tilde{X}_2 = X_t$, by the assumed Markov property of the original source.

Thus, the three-stage system satisfies the assumptions of Lemma 2, and \tilde{f}_2 can be replaced without loss of total expected cost by \hat{f}_2 , which has the structure

$$\tilde{Y}_2 = \hat{f}_2(\tilde{Z}_1, \tilde{X}_2).$$

This translates to an encoder

$$Y_t = \hat{f}_t(Z_{t-1}, X_t)$$

for the original problem. Since the notational changes do not influence total cost, the inductive step, and therefore the lemma, is proved.

Note that the above induction is carried out down to $t = 2$ because f_1 cannot help but have the desired structure, albeit trivially so.

Turning to the case of general k , observe that the encoders for the first k stages have memory structure of order k in a trivial way, whatever their design, and for the last stage, Lemma 3 applies. Thus the conclusion of the main theorem holds for $T \leq k + 1$ trivially, as does the assumption on the source. Hence, assume $T \geq k + 2$.

The essence of the proof is a "sliding block" repackaging of the source variables.

Let
$$\tilde{X}_t = (X_t, X_{t+1}, \dots, X_{t+k-1})$$

for
$$t = 1, \dots, \tilde{T} \quad \text{where} \quad \tilde{T} = T - k + 1 \geq 3.$$

Then the sequence $(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_{\tilde{T}})$ is Markov. For the variables, let

$$\tilde{Y}_1 = (Y_1, \dots, Y_k),$$

$$\tilde{Y}_t = Y_{t+k-1}, \quad \text{for } t = 2, \dots, \tilde{T}$$

$$\tilde{Z}_t = Z_{t+k-1}, \quad \text{for } t = 1, \dots, \tilde{T}$$

$$\tilde{V}_1 = (V_1, \dots, V_k),$$

and
$$\tilde{V}_t = V_{t+k-1}, \quad \text{for } t = 2, \dots, \tilde{T}.$$

The functions relating these variables are written as follows:

$$\tilde{Y}_1 = \tilde{f}_1(\tilde{X}_1)$$

summarizes the action of the first k encoders, which will remain unchanged as they already (trivially) have memory structure of order k . For $t > 1$,

$$\tilde{Y}_t = \tilde{f}_t(\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_t) = f_{t+k-1}(X_1, \dots, X_{t+k-1})$$

where \tilde{f}_t is not uniquely defined by this relation. It can be made unique and measurable by requiring (for example) that for $\theta = 2, \dots, t$, the function \tilde{f}_t depends on the argument $\tilde{X}_\theta = (X_\theta, \dots, X_{\theta+k-1})$ only through its last component $X_{\theta+k-1}$. However, any measurable \tilde{f}_t satisfying the identity is acceptable.

The receivers are characterized by their memory updating functions:

$$\tilde{Z}_1 = \tilde{r}_1(\tilde{Y}_1)$$

summarizes the recursive buildup of Z_k from (Y_1, \dots, Y_k) using r_1, \dots, r_k . For $t > 2$, \tilde{r}_t is defined by

$$\tilde{Z}_t = \tilde{r}_t(\tilde{Z}_{t-1}, \tilde{Y}_t) = r_{t+k-1}(Z_{t+k-2}, Y_{t+k-1}).$$

Likewise,

$$\tilde{V}_1 = \tilde{g}_1(\tilde{Y}_1)$$

summarizes the action of the first k decoders (including their memory updating). For $t > 2$, \tilde{g}_t is defined by

$$\tilde{V}_t = \tilde{g}_t(\tilde{Z}_{t-1}, \tilde{Y}_t) = g_{t+k-1}(Z_{t+k-2}, Y_{t+k-1}).$$

Finally, $\tilde{\psi}_1(\tilde{X}_1, \tilde{V}_1) = \sum_{i=1}^k \psi_i(X_i, V_i)$ and for $t > 2$

$$\tilde{\psi}_t(\tilde{X}_t, \tilde{V}_t) = \psi_{t+k-1}(X_{t+k-1}, V_{t+k-1}),$$

where $\tilde{\psi}_t$ depends on argument \tilde{X}_t only through the component X_{t+k-1} . Now Lemma 4 can be applied to this \tilde{T} stage system with Markov source. Without increase in total cost, the encoders $\tilde{f}_2, \dots, \tilde{f}_T$ can be replaced by encoders \hat{f}_t with first-order memory structure, i.e.,

$$\tilde{Y}_t = \hat{f}_t(\tilde{Z}_{t-1}, \tilde{X}_t), \quad \text{for } t = 2, \dots, \tilde{T}.$$

Expressing this in terms of the original variables, the functions f_t for $t = k + 1, \dots, T$ are replaced by functions \hat{f}_t satisfying

$$Y_{t+k-1} = \hat{f}_{t+k-1}(Z_{t+k-2}, X_t, X_{t+1}, \dots, X_{t+k-1}) \quad \text{for } t = 2, \dots, T - k + 1$$

or equivalently

$$Y_t = \hat{f}_t(Z_{t-1}, X_{t-k+1}, \dots, X_t) \quad \text{for } t = k + 1, \dots, T.$$

These \hat{f}_t exhibit memory structure of order k , so that the main theorem is proved.

VI. THE CASE OF DELAYED DISTORTION MEASURES

The basic model of Section II can be modified as follows for the case in which a delay of $\delta > 0$ steps is included in the definition of distortion. The first change is that the variables V_1, \dots, V_δ are simply not generated, the receiver spends its first δ periods just accumulating observations of Y_1, \dots, Y_δ and updating its memory accordingly.

The second change is that, for $t > \delta$, distortion is measured by a function $\psi_t(X_{t-\delta}, V_t)$ whose expectation defines J_t . The design objective is to minimize

$$J = \sum_{t=\delta+1}^T J_t.$$

For this situation, the following structure simplifying result holds.

Delay Theorem: Suppose that the source is k th order Markov and that the distortion is defined with delay δ . Then any given design can be replaced, without loss, by one in which the encoders have memory structure of order $\max(k, \delta + 1)$.

Proof: In case $k \geq \delta + 1$, one can perform the same transformation of the point of view as in the proof of the main theorem. Indeed, this transformation gives cost functions of the form

$$\tilde{\psi}_t(\tilde{X}_t, \tilde{V}_t), \quad t = 1, \dots, T - k + 1,$$

where

$$\tilde{X}_t = (X_t, \dots, X_{t+k-1}), \quad \tilde{V}_1 = (V_1, \dots, V_k), \quad \tilde{V}_t = V_{t+k-1}.$$

This is compatible with the delay criterion, as follows:

$$\tilde{\psi}_1(\tilde{X}_1, \tilde{V}_1) = \sum_{t=1}^{k-\delta} \psi_t(X_t, V_{t+\delta})$$

and for $t = 2, \dots, T - k + 1$

$$\tilde{\psi}_t(\tilde{X}_t, \tilde{V}_t) = \psi_{t+k-1}(X_{t+k-\delta-1}, V_{t+k-1}),$$

where it happens that $\tilde{\psi}_t$ depends upon \tilde{X}_t only through the component $X_{t+k-\delta-1}$.

Therefore the argument of the main theorem applies: One can use encoders with memory structure of order k . In fact, the above shows that this conclusion is valid for any criteria of the form

$$\sum_t \psi_t(X_t, X_{t+1}, \dots, X_{t+k-1}, V_{t+k-1}).$$

As for the case $k < \delta + 1$, observe that the source is then, *a fortiori*, Markov of order $\delta + 1$. Hence, the first case applies to yield memory structure of order $\delta + 1$, as claimed.

VII. CONCLUDING REMARKS

A few extensions of the results are of interest.

(i) The proof of the three-stage lemma goes through under the weaker assumption that f_3 depends upon Z_2 , X_2 , and X_3 .

(ii) All the results in this paper remain true for V_t restricted to given subsets of R^s . This would correspond to quantization levels fixed in advance, as opposed to their selection as part of the design.

(iii) Suppose $\delta = 0$, $k = 1$, and the encoder is restricted *a priori* to be a finite state machine of the type

$$W_t = h_t(W_{t-1}, X_t),$$

$$Y_t = f_t(W_{t-1}, X_t),$$

where W_t is a discrete variable representing the contents of the encoder's memory. Then the main theorem implies that it is optimal to take $Z_t = W_t$ and $h_t = r_t$ since this simulation of the receiver's memory produces the argument required for the generation of Y_t . This result was obtained independently by N. T. Gaarder.

VIII. ACKNOWLEDGMENTS

The author is indebted to B. McMillan for bringing this problem to his attention and to D. Slepian for stimulating discussions.

APPENDIX

Let X be a set and \mathcal{B} a σ -algebra of subsets of X . Let Y be a finite set $\{1, \dots, q\}$. A function $F: X \times Y \rightarrow \mathbb{R}$ is called measurable if for each y in Y , the function $F(\cdot, y): X \rightarrow \mathbb{R}$ is \mathcal{B} -measurable.

Then it follows that a function $f: X \rightarrow Y$ exists such that

$$F(x, f(x)) \leq F(x, y)$$

holds for all $x \in X$ and $y \in Y$ and the function f is \mathcal{B} -measurable (which means that $\{x | f(x) = y\}$ is in \mathcal{B} for each y).

Since y takes only finitely many values, it is evident that, for each x , one can select an $f(x)$ to satisfy the inequality. However, there may be many x for which the minimizing y is not unique. This creates the need for a choice of values in defining f , and if such a choice were made in a totally arbitrary manner, it is possible that the resulting f not be \mathcal{B} -measurable. What is needed is the (elementary) proof that, for a reasonable way to resolve ambiguous choices, the resulting f is automatically \mathcal{B} -measurable.

Given $y \in Y$, consider the set A_y of all x for which y is among the minimizing values, this set is measurable because it is defined by a finite number of inequalities among measurable functions, namely, for each $y' \in Y$,

$$F(x, y) \leq F(x, y').$$

The sets A_y cover X but with overlaps. To remove the overlaps, use the numerical indexing of Y to define

$$B_1 = A_1$$

and, for $y > 1$,

$$B_y = A_y - \bigcup_{i=1}^{y-1} A_i.$$

This construction preserves measurability and removes overlap. Thus, if f is defined to take value y on B_y , the desired result is attained. This amounts to stipulating that, when the minimum is attained for more than one element of Y , $f(x)$ is defined as the element with the smallest label.

REFERENCE

1. B. McMillan, "Communicating Systems Which Minimize Coding Noise," B.S.T.J., 48, No. 9 (November 1969), pp. 3091-3113.

Current-Access Magnetic Bubble Circuits

By A. H. BOBECK, S. L. BLANK, A. D. BUTHERUS, F. J. CIAK,
and W. STRAUSS

(Manuscript received March 9, 1979)

Experimental and theoretical results from our work on current-access technology show promise for high-density, $\sim 10^7$ bits/cm², and high-frequency, $f > 1$ MHz, bubble devices. We have operated current-access devices where the bubble-driving fields derive from two patterned conducting sheets instead of orthogonal field coils. Margins for generation, propagation, and transfer were studied on 8- μ m periods at 1 MHz. These 8- μ m period structures typically required 1.5 mA/ μ m per conducting sheet and dissipated 14 μ W/bit. Single conducting-sheet, current-access circuitry also propagates bubbles but offers less design flexibility. We present design criteria, magnetic field equations, and design curves. Implementation of these devices required new magnetic materials with quality factor Q comparable to available garnets, yet higher mobility and lower dynamic coercivity. Of the three systems, $(\text{YLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$, $(\text{LaLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$, and $(\text{LaLuSm})_3(\text{FeGa})_5\text{O}_{12}$, the last appears best suited; some room temperature characteristics of the composition $\text{La}_{0.6}\text{Lu}_{2.1}\text{Sm}_{0.3}\text{Ga}_{0.9}\text{Fe}_{4.1}\text{O}_{12}$ are $4\pi M_s = 470\text{G}$, $\mu = 750 \text{ cmS}^{-1}\text{Oe}^{-1}$, $d = 1.6 \mu\text{m}$, and $\Delta H_c = 2 \text{ Oe}$ across a bubble for the threshold of motion. Necessary improvements in processing were made with a radio-frequency, chlorine-containing plasma etch which produced metal patterns identical to those of the etch mask. We anticipate that current-access devices, when compared to conventional field-access devices, will achieve higher data rates, lower power consumption per bit, and greater storage densities with existing processing technologies.

I. INTRODUCTION

We have been investigating the potential of conductor-access bubble circuitry¹ to meet the need for a high-density, high-capacity, low-cost, nonvolatile memory device. There are two basic ways to move bubbles: field-access² and current-access (conductor-access). Field-access de-

vices, the only type in use at this time, require a pair of orthogonal drive coils to provide an in-plane magnetic field which, coupled to a bubble data stream via structured permalloy or ion-implanted features,³ provides the necessary bubble drive forces.

It is difficult, however, to take advantage of the data rates intrinsic in field-access bubble devices as they are scaled down in size. In particular, the volt-ampere product increases with increasing frequency, making economical coil drive circuits difficult to design. Present field-access devices are limited to about a 250-kHz rotating field rate and are thus too slow for applications such as television frame storage. A bit rate per chip of 1 Mb/s or greater would make bubbles attractive to a wider range of potential customers.

In traditional current-access bubble circuitry, meandering conductor strips of relatively high resistance, rather than drive coils, make it possible to provide the necessary currents to move bubbles at stepping rates of 1 MHz and greater with relative ease but give rise to a concern that current-access chips would dissipate excessive power, especially at the higher bit storage densities such as 10^7 b/cm².

The disadvantage of the current-access circuits disclosed to date has been the complexity of the conductor patterns themselves. We have developed new configurations, however, that use rather simple apertured sheet geometries and which should dissipate less power as well. Bubble circuits with an 8- μ m period, a four-times improvement in storage density over present circuits, are expected to be achieved while maintaining the same design rules of the 16- μ m period field-access circuits now in manufacture at Western Electric, Reading.⁴

II. HISTORY OF CURRENT ACCESS DEVELOPMENT

It is not surprising that the very first experiments to control bubble movement used conductor drive. In these first experiments, adjacent posts of a "waffle-iron" ferrite base plate were looped by conductors as shown in Fig. 1.¹ Conductors ϕ_1 , ϕ_2 , and ϕ_3 driven by unipolar current pulses produced bubble motion with the direction determined by the pulse sequence. Bubble domains had to overlap from one post to another before proper operation could be obtained.

A continued evolution in circuit design next led to the single conductor design of Copeland et al.,⁵ shown in Fig. 2. Note the conductor which meanders back and forth across a groove which has been etched into the surface of the epitaxial garnet layer. Undulations in the width of the groove define preferred bubble rest positions which supplement a bipolar current drive to produce bubble motion. It is a basic requirement of this design, however, that a rather precise geometrical fit be maintained between the bubble and the groove pattern. The ideal bubble diameter is, in fact, one-half of the conductor period making

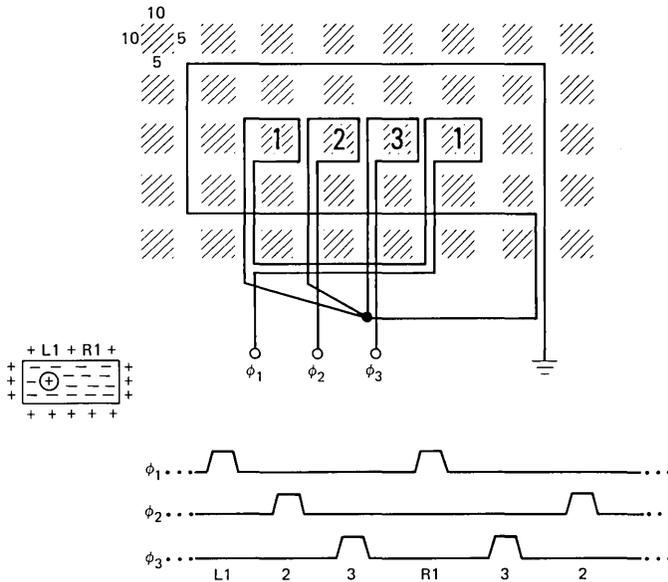


Fig. 1—Schematic of the first bubble shift register circuit. Bubbles in orthoferrites were moved three steps with a waffle-iron base plate.

bubble-to-bubble interactions prohibitive at the maximum storage density and hence the circuit could only be operated with data in every other position. A variation in which grooved tracks were replaced by ion-implanted tracks of similar shape was also operated.³ About this time, effort in conductor propagation was terminated in favor of “field-access” propagation.

Conductor propagation re-emerged in 1974 when Walsh and Charap⁶ proposed a perforated conducting sheet driven by either a rotating or oscillating current as a novel bubble drive. Their computer simulation indicated that operation comparable to that of field-access devices could be accomplished. Their paper went unnoticed.

In 1975, the bubble lattice storage device was announced by Voegeli et al.⁷ Conductors were used to move bubbles in the main storage area, the entrance-exit line, and the detector. In the most recent paper on the bubble lattice, Hu et al.⁸ state “Ironically, the “simple” shift register propagation of an isolated-bubble has one of the smallest margins in the chip.” Hu et al. were somewhat handicapped in their design since bias field compatibility had to be maintained with the main storage area. Nonetheless, their report was not encouraging.

Then in 1977, Dekker et al.,⁹ Phillips Research Labs, introduced a design which combined conductor and field access on a major-minor chip. They replaced the usually field-driven major track by a conductor-driven track that operated at a ten-times-higher stepping rate. In

this way, they achieved a fivefold reduction in access time. Typical of the Dekker designs is that illustrated in Fig. 3. The “conductor” is in reality a dual-layer, permalloy-gold sandwich in which the gold layer has been included to lower the sheet resistance and, consequently, the



Fig. 2—An SEM of a turn of a conductor-groove circuit. The circuit period is $16\ \mu\text{m}$.

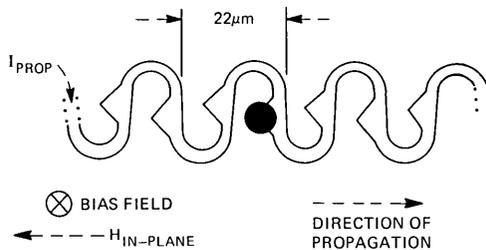


Fig. 3—A current-driven major track. An in-plane field generates poles at points of the “conductor,” which is in reality a conductor-permalloy sandwich.

drive dissipation. A nonrotating in-plane field, supplied by the very coils that operate the field-access portion of the chip, magnetizes the triangular permalloy points, converting them to low energy resting positions for bubbles. The primary drive for bubbles comes from areas defined by the conductor loops driven by a bi-polar current to alternately attract and repel bubbles. As a result, bubbles step from loop to loop but always move to the right because of the directionality induced by the triangular points described above. Bubbles attracted to these points offset to the right from their otherwise neutral position within a loop.

Dekker reported a power dissipation of 0.6 mW/b for this structure. It is thus difficult to conceive that their design could be used in an all-conductor chip, since the power dissipation for even a 100-kb chip would be 60 watts. We see in Section VII that chip partitioning can be used to reduce the dissipation considerably, but unlikely, in this case, to a useful level. Copeland's design, on the other hand, did operate at very low power per bit, but it was both wasteful of storage area and difficult to process. It was difficult to process because its minimum feature dimension was a very small portion of the circuit period.

This, then, was the reported status of current access bubble circuits. The expected advantages of higher data rates, ease of driving, and simple package design had not been realized.

III. APERTURED SHEET FUNDAMENTALS

It will be helpful to the understanding of the device structures covered later if we digress at this point to some design considerations peculiar to apertured sheet devices. First, let us consider some analogies that exist between the more familiar permalloy structures and the conductor structures that we introduce in this and the following sections.

An analogy exists between the stray field of an isolated permalloy bar immersed in a uniform magnetic field H and the stray field generated by the distortion of an otherwise uniform current I in a sheet conductor with a slot.⁶ In either case, an isolated magnetic dipole is formed. This situation is shown in Fig. 4, where it can be seen that a "plus"-pole-seeking bubble would be attracted to the upper opening of the slot just as it would be attracted to the upper end of the permalloy bar. Note also that the directions of the magnetic field H and the current I are at right angles while producing identically oriented dipoles. We must be careful not to carry this equivalence too far, since the field H is applied over a volume, whereas the current I is confined to a plane. The practical significance is that a slot can completely block a current whereas a permalloy bar cannot completely shunt an applied field.

Next, it is useful to define a sheet current density $J(\text{mA}/\mu\text{m}) = I/W$,

where $I(\text{mA})$ is the current applied to a sheet of width $W(\mu\text{m})$. We can relate current density in conductor devices to inplane field intensity in permalloy devices by observing that an infinitely wide conductor carrying a current density $J = 1 \text{ mA}/\mu\text{m}$ supports an in-plane field of 6.3 Oe. This is illustrated in Fig. 5.

There are, of course, an infinite variety of aperture profiles that can replace the slot of Fig. 4. A number that we have considered are shown in Fig. 6. By fabricating large-scale models, energizing these in a left-to-right direction from a 30-kHz current source and then measuring the signals induced in a uniformly wound single-layer coil scaled in dimension to that of a hypothetical bubble, we obtained the z -field contours also included in the figure. The z -component is that field

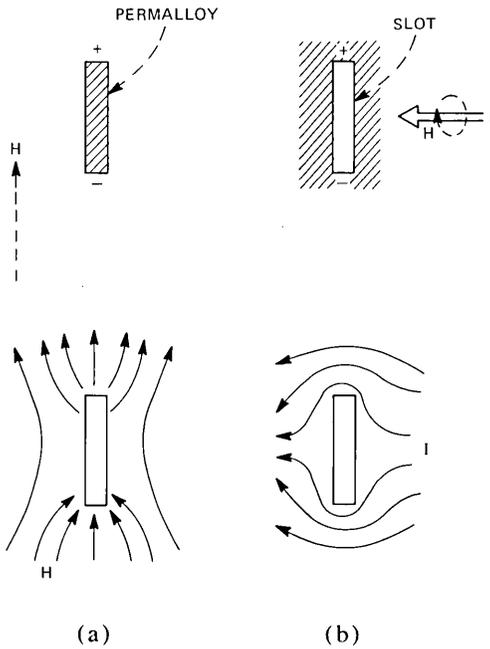


Fig. 4—Magnetic field H applied to a permalloy bar results in a dipole field. Current I applied to a conducting sheet produces a similar field.

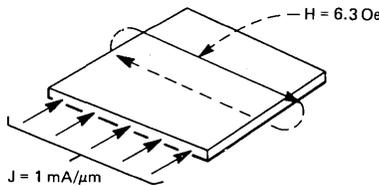


Fig. 5—A conducting sheet, infinite in extent, develops a 6.3-Oe surface field when driven at $J = 1 \text{ mA}/\mu\text{m}$.

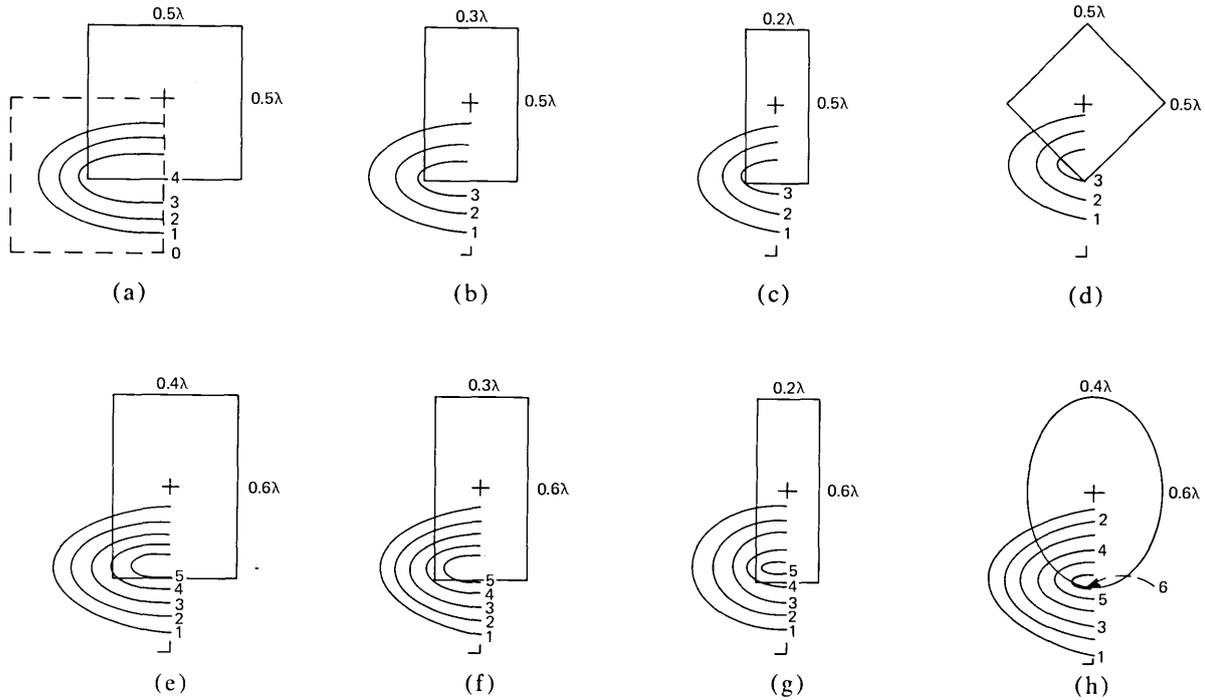


Fig. 6—Contours proportional to the z -component of field measured on scale models are given for apertures (a) through (h). Current flow was from left to right.

component normal to the sheet itself. The results actually apply to a square array of identical apertures positioned at center-to-center separations of λ . Only one quadrant of field is shown, as the others follow by symmetry considerations. Especially note how closely the field peaks and valleys are confined to the vicinity of the apertures. The oversized elements in the lower row give rise to larger fields than the shorter elements of the upper row, but at the expense of a higher power dissipation. We find the shapes "b," "f," and "h" optimal when items such as the ease of fabrication are also considered. Further details on this technique are given in Section X.

Similar results can also be obtained analytically and we have, in fact, devoted most of Section X to this end. One advantage of current-access bubble propagation is that they can be given a more precise mathematical treatment than is possible with field-access devices. Non-linear, permalloy-to-bubble interactions are not involved except, perhaps, in the detector itself. For example, we have calculated the field contours caused by current flow past a circular hole in an otherwise continuous infinite sheet. Two of the resultant design curves are given in Fig. 7, where h is the garnet thickness, a is the hole diameter, and Z_s is the elevation of the conductor sheet above the upper surface of the garnet. These contours give the average field in the garnet for an applied current density of $1 \text{ mA}/\mu\text{m}$. Conceptually, the role of the hole can be viewed as converting an otherwise uniform in-plane field (Fig. 5) into a bipolar field at about a 50-percent efficiency. This compares favorably to the conversion efficiency of a well-designed permalloy feature, which is also approximately 50 percent.

Next we consider the properties of an array of parallel conductor strips such as we encounter when detectors are discussed. The analytical results shown in Fig. 8 have been derived from design curves located in Section X. The parameters apply to a circuit period of $8 \mu\text{m}$. For $J = 1 \text{ mA}/\mu\text{m}$, the peak-to-peak bias field modulation is 9 Oe , meaning that a bubble "riding" in the field trough will have its bubble-to-strip transition field increased by 4.5 Oe . In an expander-type detector, it will be desirable to increase this transition field to at least 15 Oe so we can conclude that the current density in the detector will be increased to $\sim 4 \text{ mA}/\mu\text{m}$. In conductor circuits, there is ample opportunity to do this.

We now present a problem peculiar to the apertured sheet geometry. Current in the conductor sheet not only generates the local field inhomogeneities due to the apertures and an overall inplane field, but it also generates a gross z-field. The situation is shown in Fig. 9, where the field values pertain to $J = 1 \text{ mA}/\mu\text{m}$. As might be expected, this field peaks at the edge of the conductor. As a consequence, any bubble located near the edge of a driven conductor sheet experiences a wide

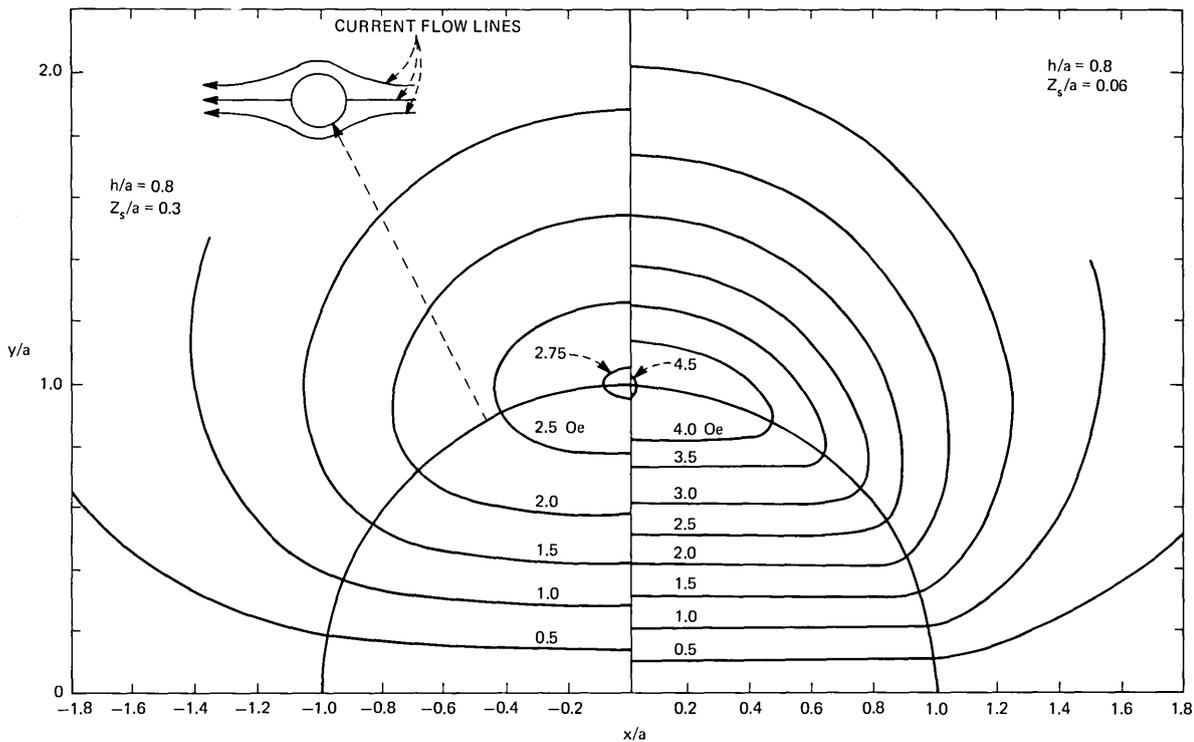


Fig. 7—Calculated z -field contours in the vicinity of a circular hole for a current density $J = 1 \text{ mA}/\mu\text{m}$.

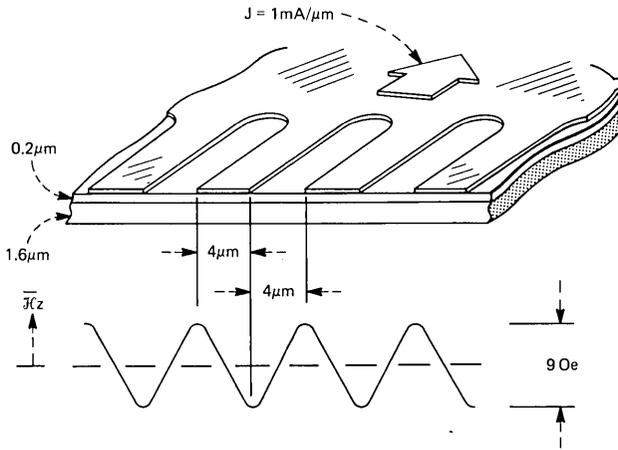


Fig. 8— \bar{H}_z variation generated by an array of parallel conducting strips. The dimensions are typical of an $8\text{-}\mu\text{m}$ period expander detector.

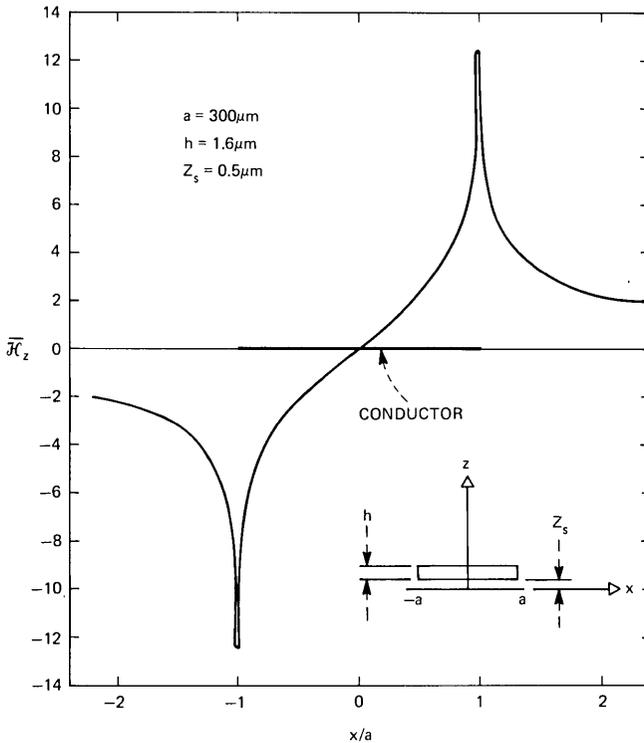


Fig. 9—Curve of the z -field due to a uniform current density $J = 1 \text{ mA}/\mu\text{m}$ flowing in a conducting sheet of width $2a$.

bias change. This problem is substantially alleviated if active bubble circuits are confined to the central 80 percent of a sheet. Other solutions such as the use of a return conductor can also be employed.

Finally, let us contrast the performance of apertures and meandering conductors with respect to field localization and power efficiency. A comparison between the field contour of the undulating conductor given in Fig. 10 and those of Fig. 6 reveal that the fields of the latter more closely resemble those of a dipole. Furthermore, the fields produced by neighboring apertures are seen to be distinct and separate. Such is not the case for the former and, in fact, the edges of the conductor are biased to nearly one-half the peak field.

The structure of Fig. 10 requires one-third the current/cell to produce the same peak field as that of Fig. 6f; however, its resistance/cell is substantially higher. Consequently, there is little difference from a power dissipation standpoint.

IV. ROTATING CURRENT

In this conductor access scheme, the propagate elements are perforations in a continuous sheet of conductor material deposited directly on the epi-garnet. Bubble propagation is accomplished by a pair of

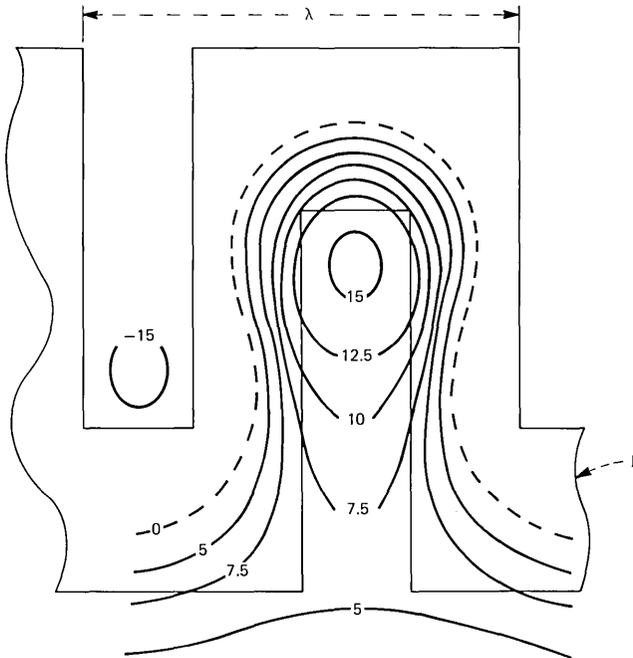


Fig. 10—Measured z -field contours for a meander-line conductor path. Note that the peak field extends toward the boundary of the conductor path.

bipolar currents applied in orthogonal directions to the conductor plane. This structure was first examined by T. J. Walsh and S. H. Charap.⁶ Their computer simulations demonstrated the principle of operation. The obvious advantage over a field access device is the absence of field coils. This permits a simpler package and higher operating frequencies, which were previously limited by the field coils.

4.1 Principle of operation

A current is applied to a patterned conductor overlay. This current produces poles around the apertures. As the current is rotated in the conductor plane, the poles rotate about the apertures (see Fig. 11). In this figure, a chevron is the aperture (propagate element). As a current I passes from left to right, poles are established, positive on the bottom and negative on the top of the chevron. The bubble is attracted to the positive poles and will rest at the lower portion of the chevrons as in Fig. 11a. In the following three illustrations, the current is shown rotated in 90-degree steps. As the current rotates, the poles it produces also shift, giving rise to bubble propagation.

4.2 Experimental

Circuits were fabricated with 16- μm and 8- μm periods. Because of the processing advantages of single-gap structures, most test elements

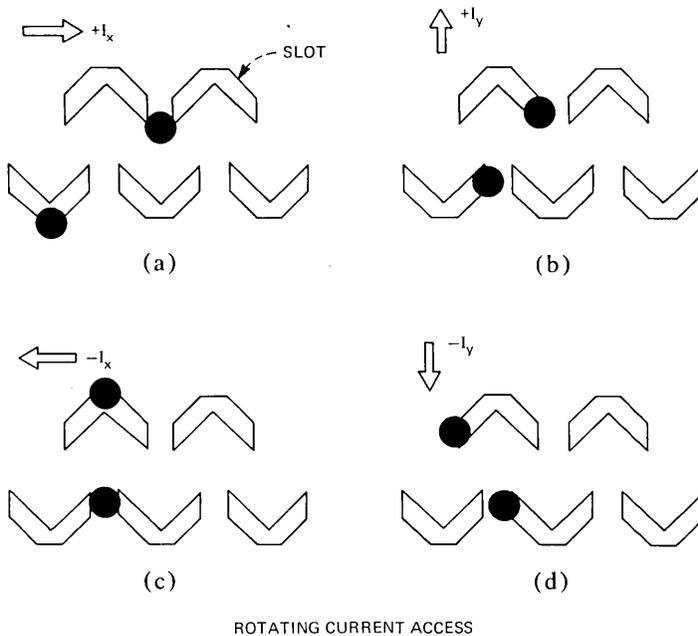


Fig. 11—Principle of operation for a circuit using rotating current for the drive. Motion, (a) through (d), proceeds much as for a single-gap permalloy circuit where chevrons are driven by a rotating field.

were chevrons. Multiple elements, such as the T-bar or X-bar, require tighter line-width control and tax processing capabilities. Portions of the masks used to make the circuits are shown in Fig. 12.

Circuits were patterned in a conductor film of 4500 Å AlCu. Bubbles were viewed in reflection with the conductor sheet serving as the reflecting surface. Four drive pulses are applied in quadrature for propagation, and a typical pulse train is shown in Fig. 13. Some overlap of adjacent current pulses is required. The amount of overlap was varied to ascertain the best operating margins with the lowest power dissipation. It was determined that 33-percent overlap was optimum. Note the slots in the arms of the test structures of Fig. 13. The slots function to keep the current flow lines parallel.

Both chevron and X-bar propagate elements were evaluated. The propagate margins for three chevron-like elements at an 8- μm period are given in Fig. 14. The margins are for various bubble patterns in straight-line propagation. Tests with different bubble patterns gave no indication of bubble-bubble interaction or reduction in propagate margin. Propagation through a 90-degree turn of Fig. 12c was marginal. The X-bar structure of Fig. 12d did not propagate domains' reliability.

4.3 Status

The high current densities required for propagation of 4 to 5 mA/ μm as well as the complexities of providing multiple current drives into a common load curtailed efforts on this single-conductor approach. Consequently, problems associated with the design of generators, transfers, turns, and detectors have not been addressed.

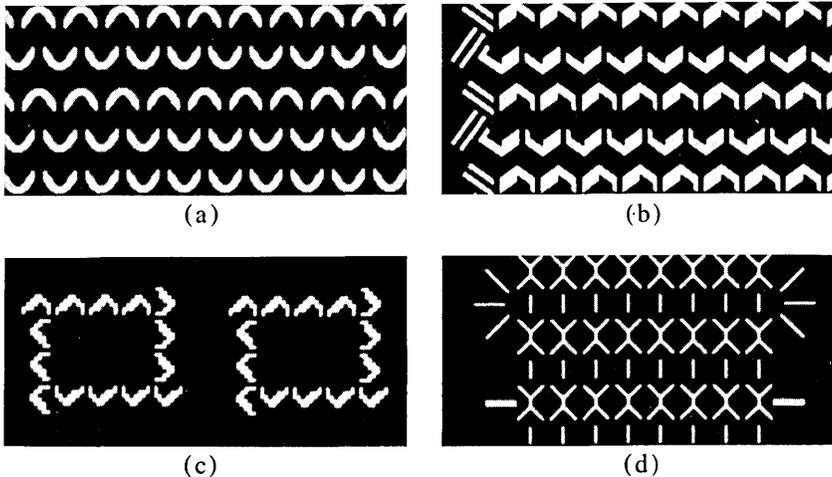


Fig. 12—Test masks used to evaluate rotating current structures.

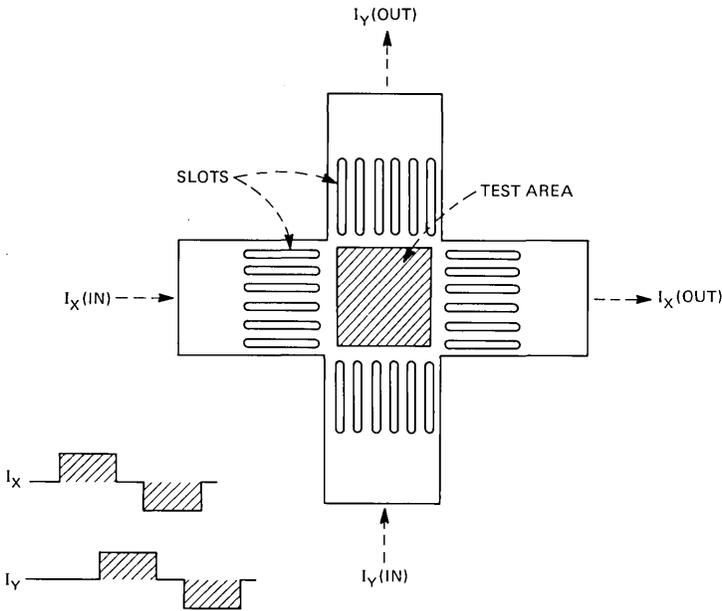


Fig. 13—Details of the overall test circuit. The slots in the lead-in arms prevent the current from “pin-cushioning.”

V. CURRENT-ACCESS PROPAGATION BASED ON A STATIC OFFSET FORCE

In this section, we take an in-depth look at propagation circuits that can be driven by a single bipolar current source. The circuits of Copeland and Dekker discussed in Section II and those introduced in the following section, in which the meandering conductor is replaced by an apertured sheet, fall into this category.

5.1 Description

An excellent discussion of the static offset force problem has already been presented by Copeland et al.¹⁰ We expand their results to include the effect of the domain wall coercivity H_C . In the usual single conductor circuit, the current flows in a conductor that crosses back and forth over the bubble track as shown in Fig. 15a. When a bipolar current is applied to the conductor, it produces an approximate standing wave field envelope:

$$H_I = N_I I \sin 2\pi x/\lambda, \quad (1)$$

where the X direction is along the bubble track and H_I is a Z -directed bubble drive field (Fig. 15b). It is obvious that a bubble can respond to

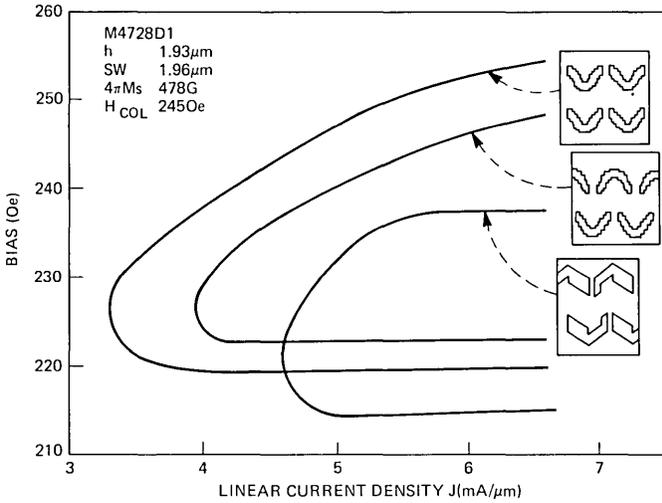


Fig. 14—Operating margins for the three aperture designs indicated.

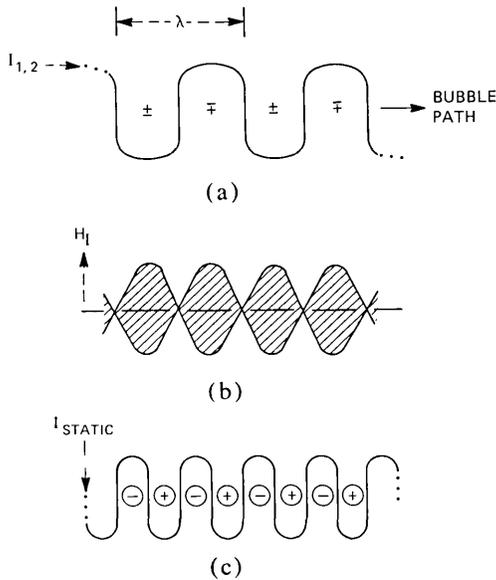


Fig. 15—Schematic layout of a single-conductor circuit. The conductor driven with a bipolar current (a) and the idealized z -field standing-wave pattern produced (b) are shown, as well as a double-frequency meander-conductor driven by a constant current (c).

this standing-wave field pattern by either oscillating back and forth or, more likely, being ejected from the track altogether. Next add the static current I_S (Fig. 15c) which produces yet another field pattern but this time with a double spatial frequency. This field is of the form

$$H_S = N_S I_S \cos(4\pi x/\lambda). \quad (2)$$

In eqs. (1) and (2), N_I and N_S are current-to-field conversion factors.

It is only illustrative, of course, that we use a static current I_S to produce H_S . We have already seen that grooves in a garnet or permalloy points can accomplish the same end and without an added power dissipation. If the dynamic (H_I) and the static (H_S) drive fields are properly proportioned, we achieve propagation as illustrated in Fig. 16. The ideal offset is $\lambda/8$ with an absolute maximum tolerance of $\pm \lambda/8$.

5.2 Conditions for bubble propagation

Before we discuss the constraints that must be imposed on H_I and H_S to obtain bubble propagation, it is helpful to review some bubble fundamentals. A field H applied to a stationary wall causes motion at a velocity v (cm/s) given by the equation

$$v = \mu(H - H_c), \quad (3)$$

where μ (cm/sec-Oe) is the wall mobility. In the case of a bubble, the change in field intensity across the bubble ΔH is used with the result

$$V = \frac{\mu}{2} [\Delta H - 8H_c/\pi]. \quad (4)$$

Since the bubble in the circuit (Fig. 15) experiences fields H_I and H_S simultaneously, we combine (1) and (2) to obtain

$$H_T = N_I I \sin 2\pi X/\lambda + N_S I_S \cos(4\pi X/\lambda). \quad (5)$$

There is no general analytical solution for the velocity of a bubble, even assuming H_I sinusoidal and $H_c = 0$. By graphical analysis we can,

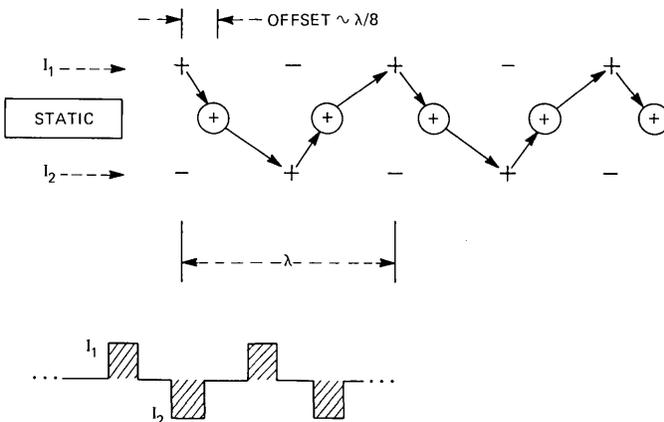


Fig. 16—The unidirectional propagation mechanism for the circuit of Fig. 15.

however, determine the minimum fields needed to overcome the coercivity H_c . From these we can determine the minimum drive currents. The mobility can then be introduced as an iteration of the coercivity under the assumption that the bubble moves at constant velocity.

The drive field H_I and the static field H_S and their sum are plotted in Fig. 17. It is assumed that positive fields increase the bias-causing bubbles to seek locations at which the applied fields are most negative. We treat the case where the bubble diameter is one-fourth the circuit period λ ; however, the analysis can be extended to any diameter. Since mobility effects are being neglected, the condition $\Delta H \geq 8H_c/\pi$ is sufficient to move a bubble. Actually, (4) is valid only for a linear gradient; however, we assume that any field difference at the extremes of the bubble diameter is equivalent to ΔH .

We start with a bubble in position 1 of Fig. 17, where 1 defines the location of the center of the bubble and, initially, with $H_c = 0$. Increasing H_I from zero to the amplitude sketched on the figure shifts the bubble from 1 to 2 and then, as H_I returns to zero, to 3. This completes a half-step of propagation. The other half-step takes place when H_I is reversed in sign. Next, we consider the conditions imposed on H_S when the coercivity H_c is included. There are "dead spots" at H_I (peak), and it is the role of the static field H_S to dislodge bubbles from those positions. The minimum values of H_S for any movement whatsoever is

$$2H_S = 8H_c/\pi. \quad (6)$$

Equation (6) therefore defines the critical minimum value for H_S . To

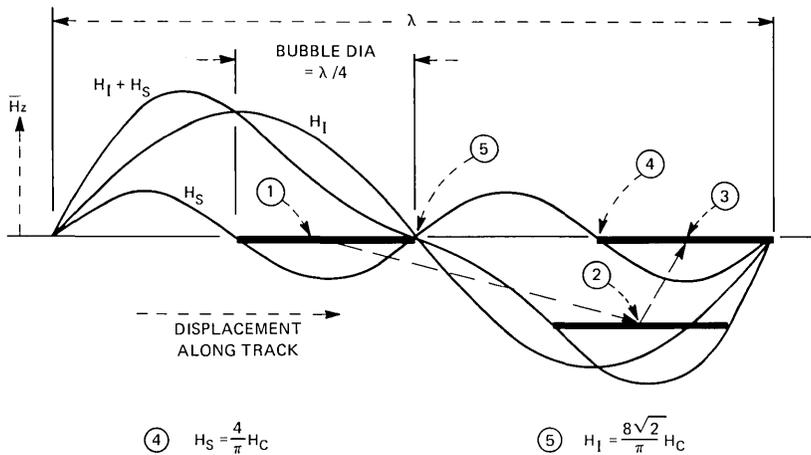


Fig. 17—Idealized z -fields H_I and H_S for a single-conductor propagation circuit with a static offset field. Bubble motion is from 1 to 3.

deduce the critical amplitude of H_I , we set $H_S = 4H_c/\pi$ and assume that there is a bubble at 1. As H_I is increased from zero amplitude, the bubble will move to the right, eventually reaching 5. Now 5 is a minimum gradient location for the bubble and, to overcome the coercivity, the amplitude of H_I must reach

$$H_I = 2\sqrt{2}H_S. \quad (7)$$

Once beyond 5, the bubble moves quickly ahead since it has excess gradient until it stops at 4, where again the field gradient is $8H_c/\pi$. When H_I returns to zero, the static field H_S completes this half-step by nudging the bubble a bit further to the right.

In summary, our two design equations are

$$H_S = 4H_c/\pi \quad (8)$$

and

$$H_I = 8\sqrt{2}H_c/\pi. \quad (9)$$

5.3 Methods to realize a static offset field

It is beyond the scope of this paper to list all the methods proposed to produce a static offset field. Permalloy dots, permalloy points, and embossed garnet features, i.e., grooves, are the three most extensively reported. In the next section, two more approaches are added to that list. They are (i) permalloy features contoured in the z -direction and polarized by the bias field and (ii) ion-implanted low-energy bubble rest positions produced by selectively implanting zones in the surface of the garnet. The remainder of this section, however, is devoted to "grooves," as they are representative of the other approaches.

In *Magnetic Bubbles* by O'Dell,¹¹ the equivalence of bubbles and current loops is discussed at length. This analogy serves as an excellent introduction to the subject of grooves. It can be shown that the stray field of a bubble and a loop current are identical if the bubble is replaced by a loop current $I_B = 2M_s h$, where the units are mA, gauss and μm , respectively. As an example, if $h = 1.5 \mu\text{m}$ and $M_s = 40$ gauss, parameters typical for an 8- μm device, then $I_B = 120$ mA. Shown in Fig. 18a, the "loop current" representation aids in the comprehension of bubble behavior.

For a cylindrical hole in a magnetic layer, a similar equivalent current is $I_H = M_s h$, which, for the parameters above, is 60 mA. For partially etched holes, i.e., craters, the expression is $I_h = M_s \Delta h$ (Figs. 18b and 18c). Since an undulating groove is nothing more than a continuous chain of craters, the analogy can readily be extended to cover that configuration.

We have also included the design curves of Fig. 19, which give the

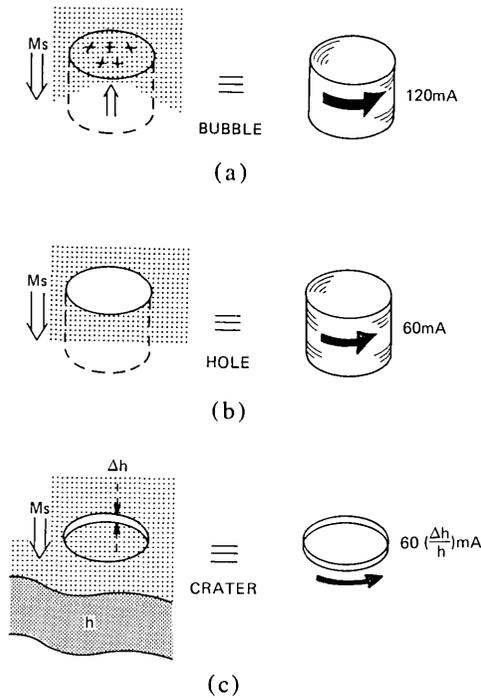


Fig. 18—The loop current magnetostatically equivalent to a typical bubble found in 8- μ m period devices is given in (a). Similar relationships hold for a hole (b) and a partial hole (c).

field of a current loop located on the surface of a plate of thickness h . These curves will be useful if a more detailed analysis is attempted. They give the z -field averaged through the plate thickness h and normalized to the field at the center of the loop.

VI. KEYHOLE AND KNOTHOLE CIRCUITS

A class of bubble devices is now introduced, i.e., those driven by currents applied to apertured sheets rather than to arrays of conductors. In this section, only those configurations with a single conductor sheet are described. In Section VII, those circuits with two conductor sheets are covered.

6.1 Keyhole circuits

The idea for keyhole circuits originated from considerations involving the well-known "conductor crossing problem" in permalloy circuits. In particular, we were concerned with the role of the bias field. The bias field necessary to stabilize 6- μ m and 1- μ m diameter bubbles is about 180 Oe and 450 Oe, respectively. Consequently, the direct

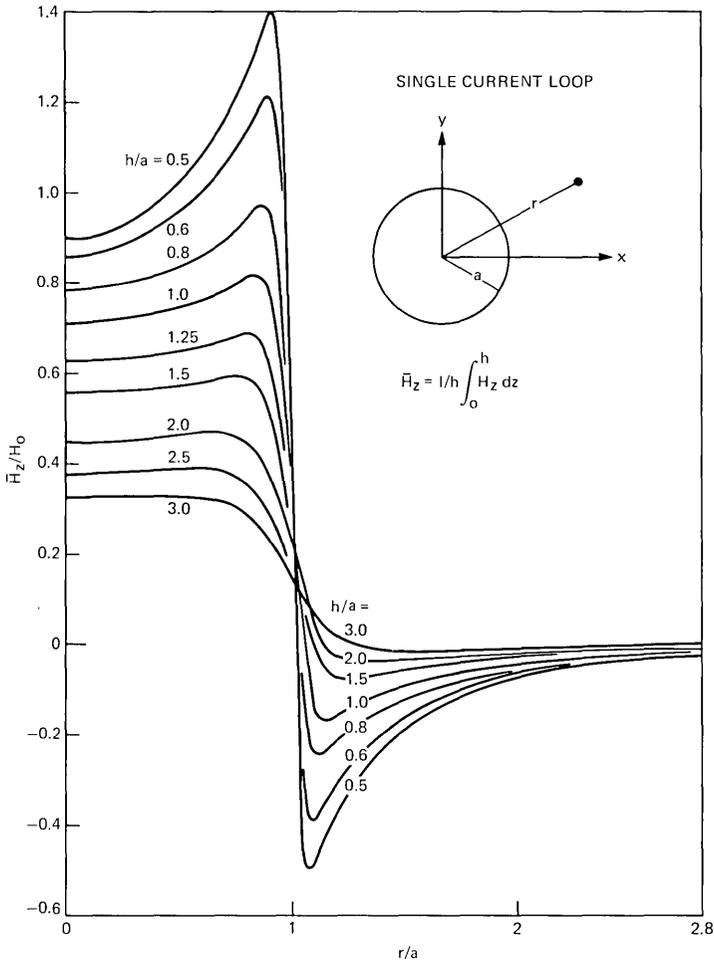


Fig. 19— Z -field contours normalized to the field H_0 at the center of a current loop. These curves are useful to characterize bubble-bubble as well as bubble-etched feature (embossed) interactions.

influence of the bias field on the magnetic state of the permalloy features cannot be neglected in circuits with small bubbles, i.e., high density circuits.

We use the bias field to produce a static offset field by the approach seen in Fig. 20. The field H_{bias} has a component H_{\parallel} directed along the body of any permalloy feature inclined to the plane of the garnet. Such is the case, then, for a permalloy feature patterned at the edge of a tapered conductor. Magnetic poles that repel bubbles develop at the end of the permalloy feature nearer the garnet, while those that attract

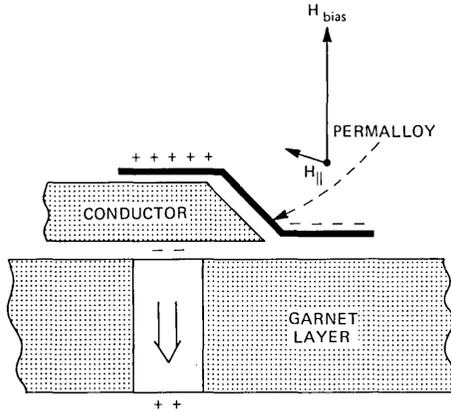


Fig. 20—A permalloy element, located on the edge of a tapered conductor, polarized by a component of the bias field.

bubbles develop at the end farther from the garnet. A reversal of the bias field direction will reverse the sign of the permalloy poles. However, the poles of the bubble also reverse, so the interaction remains unchanged. It is relatively easy to approximate the magnitude of the local field perturbations presented to the bubble. These calculations show that peak fields of a few oersteds to tens of oersteds can be expected.

Bubble propagation circuits capitalizing on this effect were designed; Fig. 21 is typical of those in which the apertures, squares fitted with a tab, are keyhole-like in design; hence, the name. Position and shape of the permalloy bar is seen in more detail in Fig. 22. For the reasons described above, bubbles are attracted to the ends of permalloy bars and repelled from their centers. Bubbles travel from left to right, moving from bar to bar in response to a bipolar current applied vertically to the sheet. They must “tunnel” through barriers as they travel from end-to-end on a given bar, but not as they jump from one bar to another. This asymmetry was apparent in the minimum drive currents of circuits that were tested.

For an 8- μm period circuit operated at 500-kHz, pulse amplitudes were required of 11.2 mA/ μm to move along the permalloy and 7.2 mA/ μm to bridge successive elements. Construction details were: 0.10- μm SiO₂ prespacer, 0.28- μm AlCu, 0.23- μm SiO₂ spacer, and 0.62- μm permalloy (ply). With a garnet material that supported 1- μm bubbles, the operating bias range was 495 ± 17 Oe. A feature of this class of circuits is that they can accommodate a very wide range of bubble diameters. One foremost disadvantage is that they are difficult to process because of the intricacies of the aperture design.

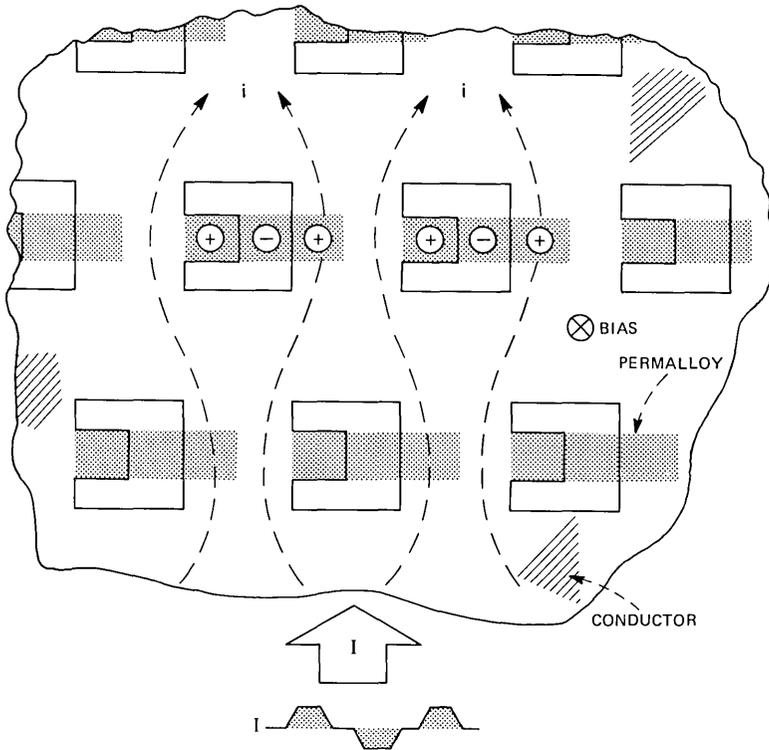


Fig. 21—Bubbles in this keyhole circuit move from left to right in response to a bidirectional current drive. The tab within the hole elevates the end of the permalloy bar but does not interfere with current flow around the aperture.

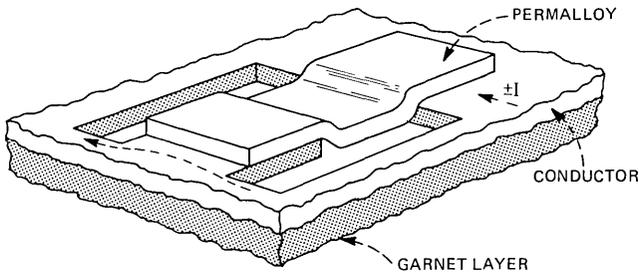


Fig. 22—Perspective view of a keyhole aperture shows the contoured permalloy.

6.2 Knothole circuits

In a knothole circuit, ion-implanted islands are combined with a single-apertured conductor sheet to form a bubble shift register. Uniform ion implantation of a garnet surface is the traditional method of

suppressing hard bubbles. Field-access bubble propagate circuits that use implanted patterns rather than permalloy elements are being considered for next-generation high-density applications. The uniaxial anisotropy of most garnet materials, and especially those treated in this paper, is lowered by any implantation that stresses the lattice. One can consider that bubbles couple to implanted areas because these areas tend to reduce the stray field energy by providing flux closure. The interaction, however, is more complex, as it can involve "charge walls" in the implant layer. Also, bubbles are especially attracted to the implant side of boundaries that partition implanted and nonimplanted areas. Phenomologically, an implanted island behaves like that of an area that has been thinned and as such can be replaced by a loop current. It follows then that the knothole circuit is a physical representation of the hypothetical structure analyzed in Section V.

A straight-line knothole shift register requires two implant islands and a single, simple aperture for each circuit period. Note that a tab in the aperture is no longer necessary and has been removed. Hence, this circuit has been given the descriptive name knothole. The general layout and operation of three parallel shift registers is illustrated in Fig. 23. Bubbles in the center row propagate to the right, those in the

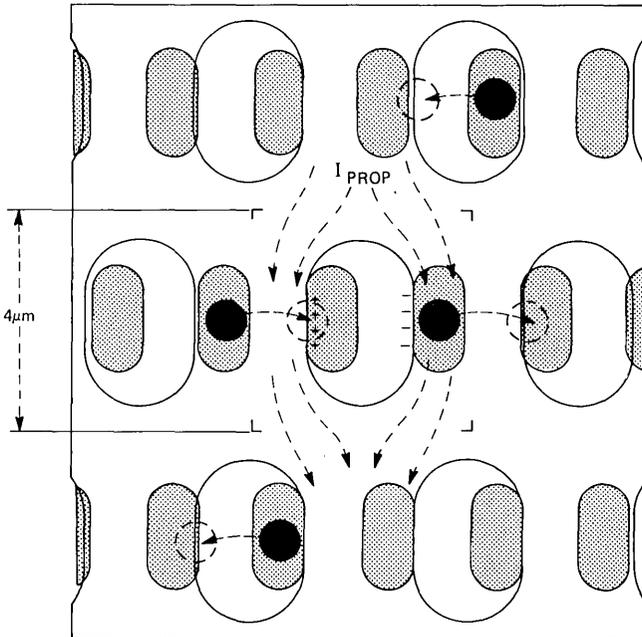


Fig. 23—Bubbles move from rest positions at the ion-implanted islands in response to current flow downward. Each circuit period requires two implanted islands.

outside rows propagate to the left. A bipolar current applied to the sheet drives the bubbles.

The performance of knothole circuits is sensitive to bubble size. To illustrate this point, refer to Fig. 24 where the data are presented for a 4- μm period test circuit using a nominal 1.7- μm bubble material. Islands were given an implant of 2.5×10^{13} Ne at 100 Kev, and no overall implant was included. A static in-plane field of 210 Oe was applied parallel to the current. Bias margin, frequency range, and drive power/bit (based on 0.1 ohms/cell) are outstanding. The direction of propagation is *opposite*, however, to the design direction. The oversize bubbles straddle pairs of implanted areas during operation.

Design of an 8- μm period closed-loop shift register is given in Fig. 25 and data in Fig. 26. Here the nominal bubble diameter is 1.6 μm , which is a good match to the nominal 2 μm by 4 μm implant islands; hence, the direction of propagation is correct in the classical sense, i.e., clockwise around the loop. The islands were implanted at 3×10^{13} with Ne at 50 Kev. Driven at $J = 3 \text{ mA}/\mu\text{m}$, the continuous power dissipation is 58 $\mu\text{w}/\text{b}$.

A variety of π -turns were tested during the course of this development, and the most successful of these is incorporated in the circuit of

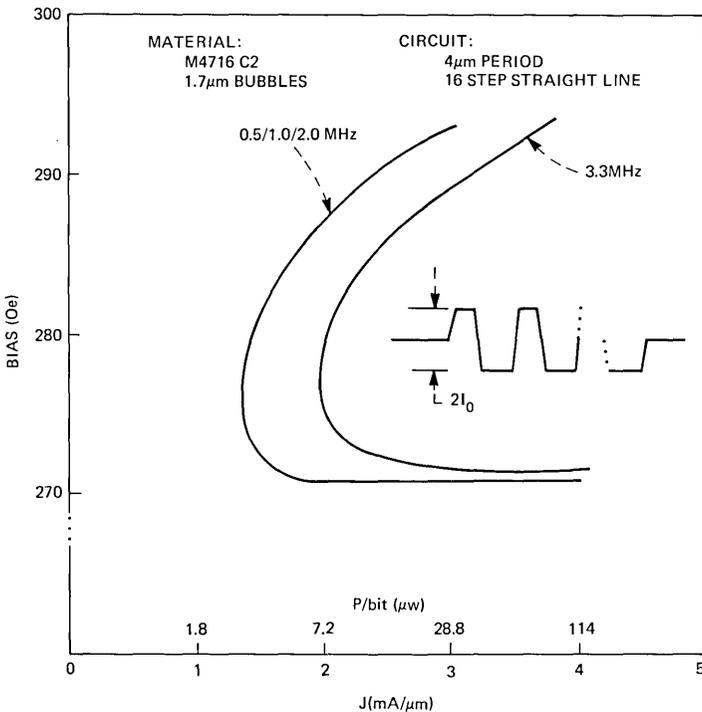


Fig. 24—Operating margins vs frequency for a 4- μm period knothole circuit.

Fig. 25. The shape of the aperture in the turn was optimized from field plots on scale models and iterated until the desired drive gradients were seen. Operating margins of this turn improve when an in-plane field is applied parallel to the direction of the drive current. For the data of Fig. 26, the in-plane field was 100 Oe.

Straight-line propagation parallel to the drive current was achieved with the circuit of Fig. 27. Operating margins are similar to those obtained for propagation normal to the current, and much of the same design criteria apply here. Although the drive currents per cell are lower, the resistance/cell is substantially higher; thus, the power/cell is essentially unchanged.

A detailed description of support functions such as generation and detection is deferred until Section VII, which treats dual conductor circuits.

VII. DUAL CONDUCTOR OPERATION

The knothole configuration with its single perforated sheet and "offset-force" zones does not have the design flexibility that can be

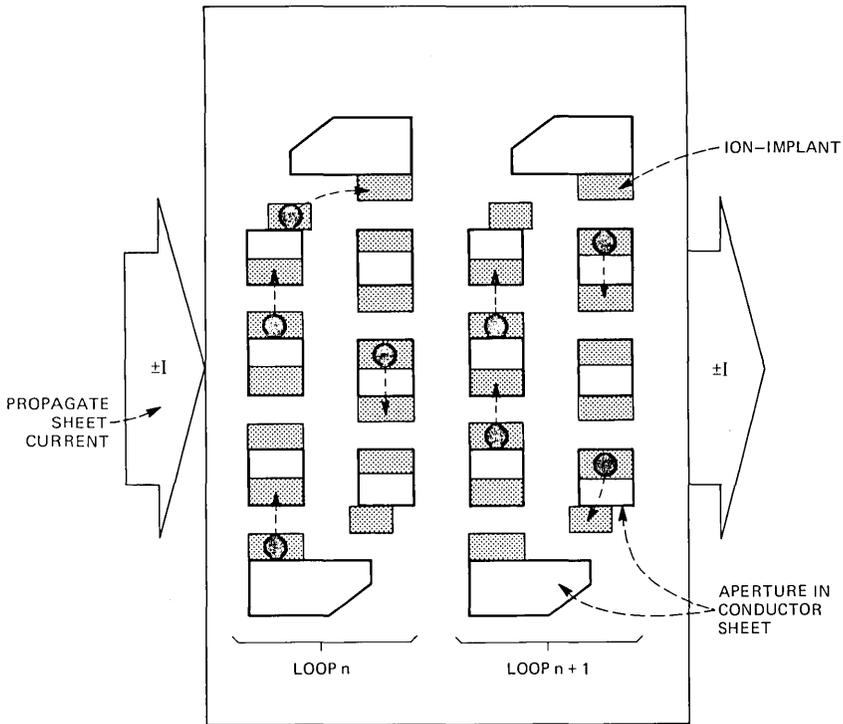


Fig. 25—Details of π -turns used in 7-step shift register loops. Clockwise turns can be mirrored to make counterclockwise turns.

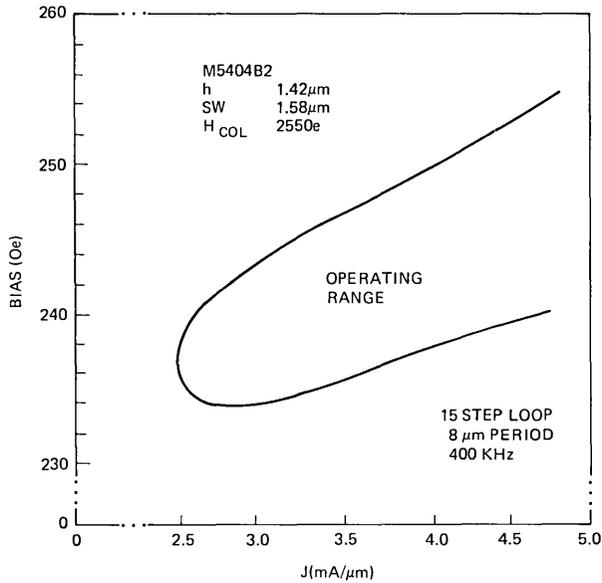


Fig. 26—Data for a 15-step knothole loop designed as in Fig. 25. Bubbles were propagated in 14-step bursts.

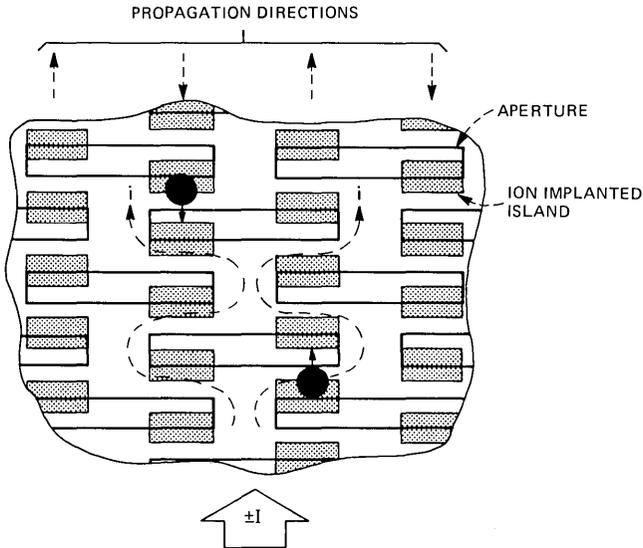


Fig. 27—Propagation parallel to the direction of general current flow.

achieved with a pair of conductor sheets. For example, with the pair of conductor sheets it is possible to transfer a bubble, i.e., switch a bubble from one path to another, or reverse the direction of propagation, simply by altering the propagate pulse sequence. This section deals

with straight-line propagation parallel and perpendicular to the direction of current flow, closed loop design, nucleate generation, detection, and transfer. Also included is a discussion of factors that influence chip design. The discussion is limited to the case where the currents in the two levels flow either parallel or anti-parallel to one another. It will be apparent to the reader, however, that, if the structures are reconfigured so that the currents flow perpendicular to one another, a new family of device structures somewhere between the rotating current structures of Section IV and those of this section could be realized.

Dual conductor circuits are fabricated with a first patterned conductor level either directly on the epitaxial garnet or onto a pre-spacer layer, an insulating layer, and finally a second patterned conductor level. Details of the actual fabrication are covered in Section IX. It will be useful at this time to describe the general layout of the test structure depicted in Fig. 28. All the test data described in the following sections were taken on circuits located within the cross-hatched region which is $320\ \mu\text{m}$ wide and $600\ \mu\text{m}$ long. The paths that lead into this region are shaped to make the current density within the region as uniform as possible.

In line drawings equivalent to processed circuits, for a view looking down onto the second conductor layer, the following conventions will be established and used throughout (see Fig. 29). All apertures in the first conductor level (that nearest the garnet) are outlined by a solid line, whereas apertures in the second level are indicated by a dot pattern. Sheet currents $I_{1,3}$ and $I_{2,4}$ flow in the first and second

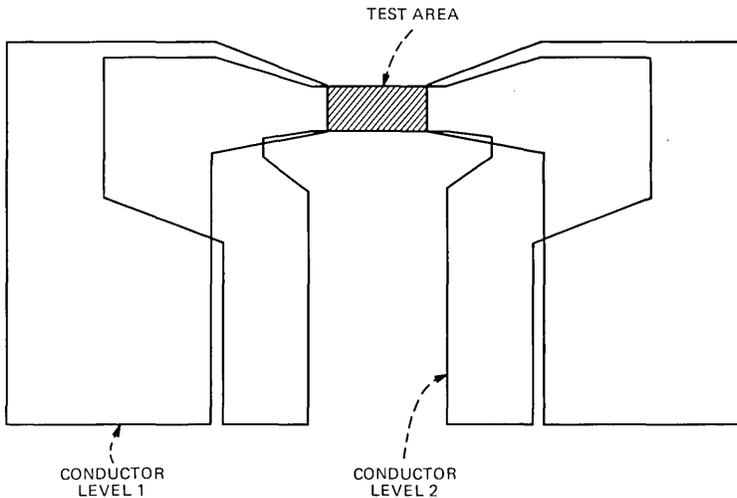


Fig. 28—Layout of the circuit used to test dual-conductor designs. Two or more probe fingers contact each pad area located along the bottom edge. Uniform current density is maintained at the test area.

conductors, respectively, where they generate local bias-field perturbations at positions 1, 3 and 2, 4. For a bubble-stabilizing field H_{bias} directed downward as shown, bubbles are attracted to location 1 for current I_1 , location 2 for I_2 , etc.

7.1 Propagation normal to the current flow

In Fig. 29, it can be seen that a bubble initially at position 1 propagates upwards (normal to the direction of current flow) when the current pulse sequence [234] is applied. Sequence [321] returns the bubble to its starting position. Straight-line propagation over a further distance is achieved by simply linking like pairs of apertures as in Fig. 30. The length of each aperture is 0.5λ , and the distance between successive like positions is the circuit period, λ . It should also be apparent that the maximum tolerable misalignment of apertures in the vertical direction is 0.25λ .

Typical current-pulse waveshapes to produce the sequence

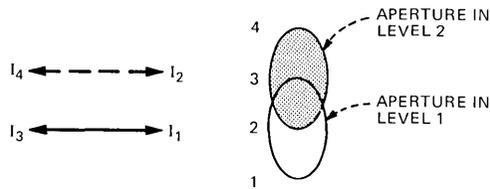


Fig. 29—A four-position, dual-conductor, bubble-stepping circuit. Bubbles are attracted to positions 1 through 4 when currents I_1 through I_4 are applied.

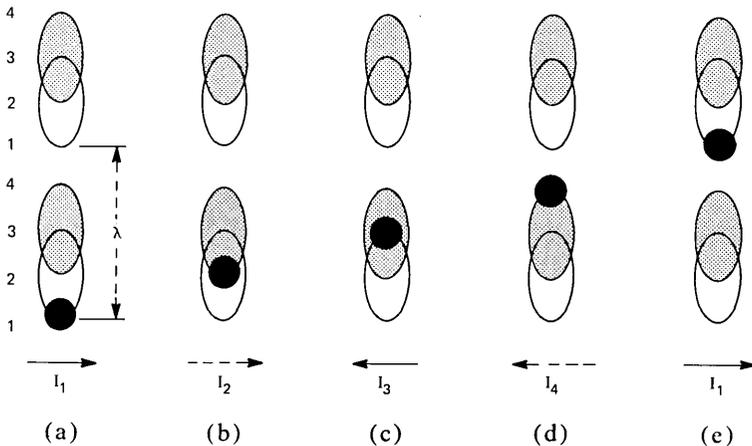


Fig. 30—A time sequence of the operation of a dual-conductor shift register. The direction of propagation will be reversed if the currents are applied in the sequence (e) through (a).

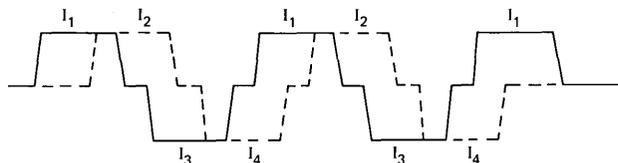
[123412341] for stepping a bubble through two periods are sketched in Fig. 31. Some overlap between successive pulses is helpful to further quantize the traveling-wave bias field minimum that drives the bubble. Also, to get a bubble fully advanced into position for, say, another two-step burst, then a final [41] overlap should be provided.

Success in aperture design is measured in terms of generating the maximum possible field gradient for the lowest power dissipation. Almost any aperture will perturb an otherwise uniform current density and give rise to bubble motion. Some of those tried were shown in Fig. 6, along with field plots obtained on scaled-up models (details in Section X). This field data information coupled with both processing and operational experience points to the rectangular shape 0.3λ by 0.5λ as the most practical choice.

The operating margins for a series of $8\text{-}\mu\text{m}$ period dual-conductor test circuits are presented in Figs. 32 to 38. Circuit data cover both unimplanted and implanted garnets as well as the influence of in-plane fields parallel to the drive currents. A frequency and temperature run are also included. Individual data points are shown wherever they aid in understanding the experiment. Bubbles used in these tests were obtained by subjecting a chip to a large in-plane field. Testing continued until the bubble supply was depleted, and a new supply was initialized.

We report, in Fig. 32, 1-mHz data on an implanted garnet test circuit driven with a two-step to and from sequence [1234123432143214]. Of special interest is the spread in data points, especially at low drive. We suspect that this is a result of the bubble states found in implanted garnets¹²⁻¹⁴ and to a greater extent in unimplanted garnets. The domain structure of the implanted layer is also suspected.

The low bias margin is limited by stripout, as is usual for most of the dual-conductor circuits tested. This is especially significant when detection is considered. Free-bubble collapse limits the absolute upper bias, since "hold currents" are not maintained between pulse bursts. If an in-plane field is applied,¹⁵ much of the uncertainty in the data points is eliminated, as is evident in Fig. 33.



CURRENT PULSE SEQUENCE

Fig. 31—The current pulse sequence used to step bubbles in a dual-conductor circuit. The overlap of pulses is used to further quantize the propagation field.

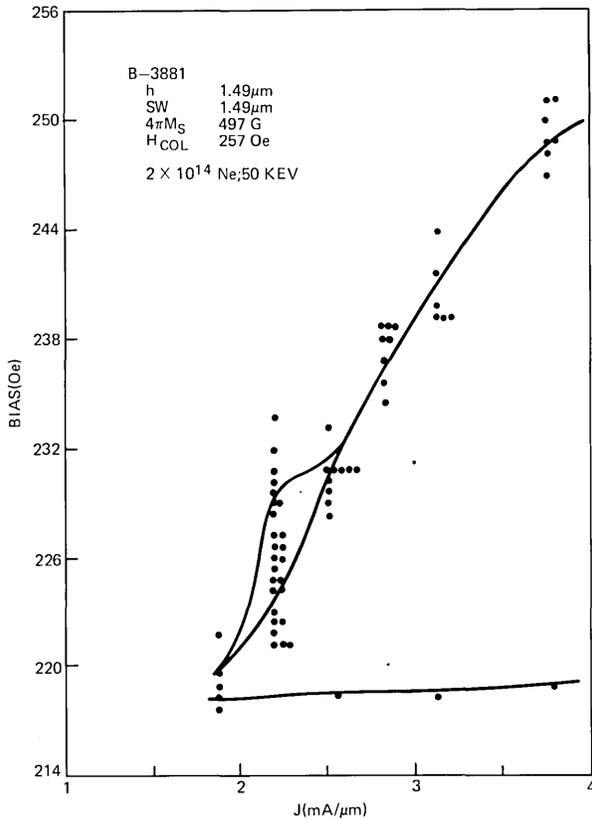


Fig. 32—Bias margin vs drive for an implanted garnet tested at a 1-MHz stepping rate in an 8- μ m period dual-conductor circuit. Each data point represents an individual check of the upper bias margin. The spread in the test data indicates that the bubble state and/or the implant layer magnetization does not reproduce from test to test.

When an unimplanted garnet is tested, the upper bias margin often shows two distinct distributions (Fig. 34). Failure modes associated with these curves indicate that two distinct bubble states, probably the $S = 0$ and $S = 1$, are involved. With a sufficiently intense in-plane field, the double distribution disappears. Refer to Fig. 35, where margins for both a 200-Oe and 300-Oe in-plane field are given.

These experiments suggest that an in-plane field is helpful to stabilize bubble propagation in dual-conductor circuits. We do expect the shape of the field gradients to continually redirect bubbles toward the desired path. However, it is known that $S \neq 0$ bubbles do not move along the maximum field gradient, but rather at angles (often near right angles) to the gradient. In Table I, the skew angle calculated for $S = 1$ bubbles are tabulated for the garnet compositions reported in this paper. Further insight into the problem comes from an analysis of

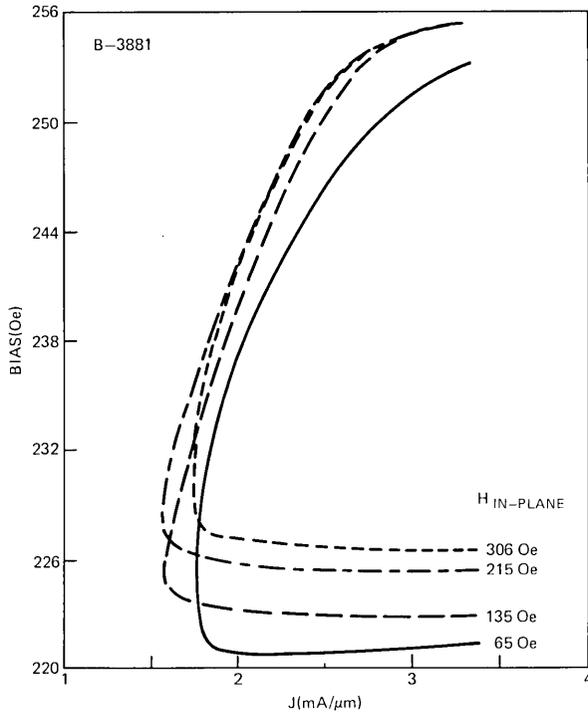


Fig. 33—Margin data for the conditions of Fig. 32, except that an in-plane field from 65 to 306 Oe was applied parallel to the current flow. The spread in data points was virtually eliminated.

Table I—Comparison of material parameters

Material Type	q (ratio)	$4\pi M_s$ (gauss)	$\Delta H_c \cdot$ (Oe)	μ (cmS/s-Oe)	$(l/\gamma \text{Oe-s})$	ψ^* (degrees)	f_{OPT}^\dagger (MHz)
YSmLuCaGe garnet	3.8	524	2.0	300	1.7×10^7	22.5	0.4
LaLuSmGa garnet	4.0	478	1.5	500-1000	1.7×10^7	35-45	0.5-1.0

* $\tan \psi = 4\mu S/\gamma d$ ($S = 1$).

† $f_{OPT} = 4\mu H_c/\pi\lambda$ ($\lambda = 8 \mu\text{m}$).

the trajectory of an $S \neq 0$ bubble moving in the dipole field generated at a circular hole. Details are presented in Section X. The choice of whether to cope with bubble states by one of the conventional hard bubble suppression methods or by an in-plane field will be dictated by the method chosen to detect bubbles. It may prove advantageous to shape the apertures, particularly in turns, to take advantage of preferred propagation angles.

In Fig. 36, the results are presented of a frequency run on an unimplanted garnet in a 200-Oe in-plane field. These results are

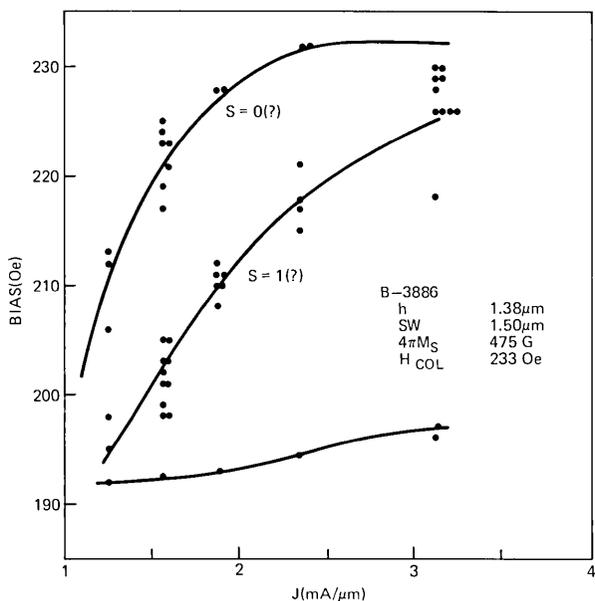


Fig. 34—Bias margins vs drive for an unimplanted garnet tested at 1 MHz in an 8- μm period dual-conductor circuit. Two distinct distributions in the upper bias margin were observed.

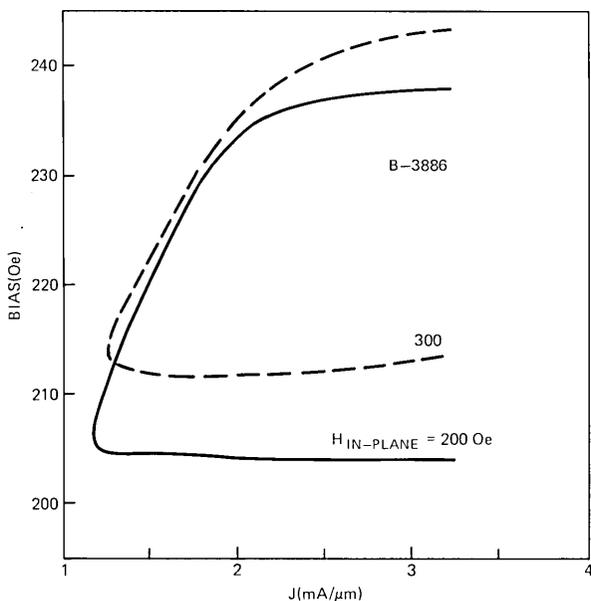


Fig. 35—The data of Fig. 34 repeated except with an in-plane field applied parallel to the current. With unimplanted garnets, substantially higher in-plane fields are needed to tighten the data points.

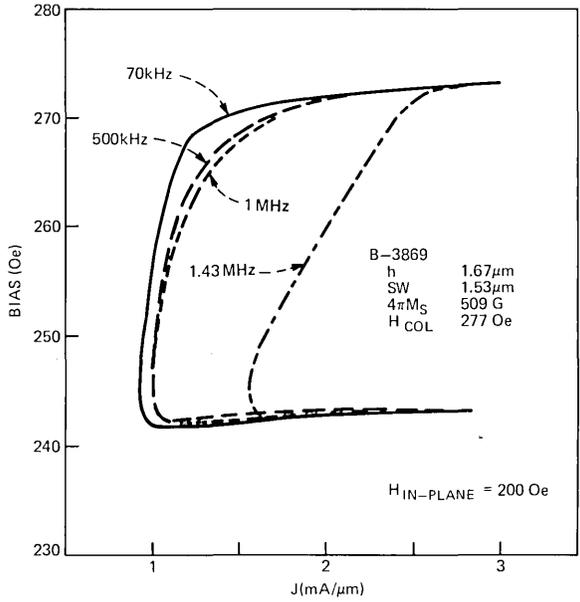


Fig. 36—Operating bias margin vs frequency for an unimplanted garnet film with a 200-Oe in-plane field.

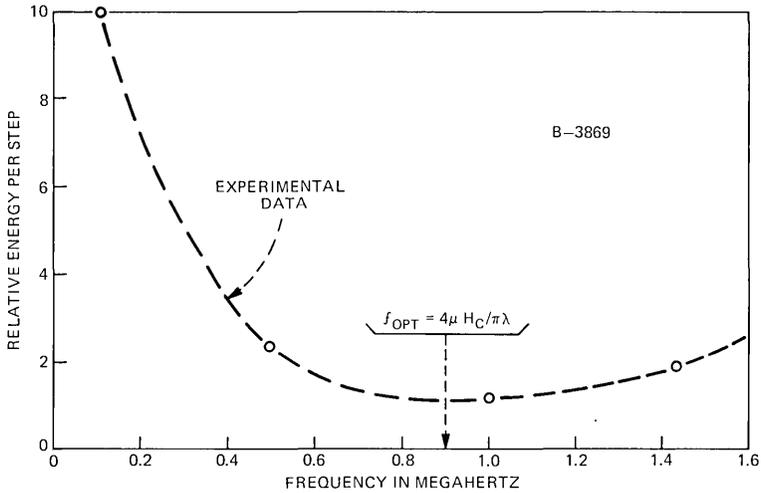


Fig. 37—It is wasteful of power to run current-access circuits except at a frequency f_{OPT} at which the energy dissipated per step is a minimum. Calculated and experimental values of f_{OPT} are compared.

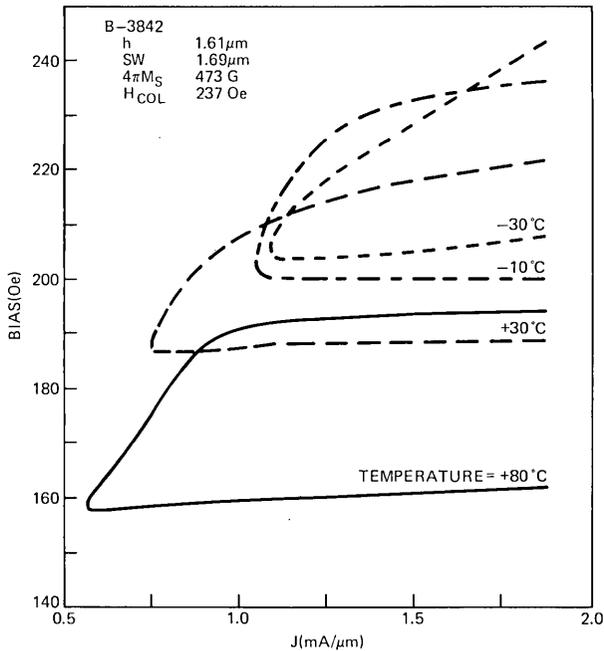


Fig. 38—Operating margins vs temperature taken at 30 kHz with a four-step propagation sequence. The garnet was unimplanted.

consistent with predictions based on calculated drive gradients and measured garnet properties. We can use these results to introduce f_{OPT} , the optimum frequency at which to step a bubble from one point to another to minimize energy dissipation. It can be shown that $f_{OPT} = 4 \mu H_c / \pi \lambda$. The relative energy per step (proportional to J^2/f) vs frequency calculated from the data of Fig. 36 is plotted in Fig. 37. The calculated f_{OPT} for wafer B-3869 ($H_c = 0.5$ Oe, $\mu = 1,000$ cm/sec-Oe) compares favorably to the frequency corresponding to the energy minimum of Fig. 37.

Operating margins at 30 kHz for an unimplanted garnet over the temperature range -30°C to $+80^\circ\text{C}$ are given in Fig. 38. These data were furnished by J. L. Smith. The minimum drive increases at low temperature much as expected due to the increase in coercivity. Note especially the very wide margins at 80°C .

7.2 Propagation around a closed loop

Details of a nine-step, bi-directional, closed loop are illustrated in Fig. 39. Note that the slots in the turn are 0.75λ in length and that elements that lead into the turn are repositioned somewhat. Bubbles propagate clockwise for the drive sequence $[\dots 12341234\dots]$ and

counterclockwise for [...43214321...]. Bias field-drive characteristics for a 6- μm period loop operated at 1 MHz are given in Fig. 40. The margins are quite adequate, as is generally the case for loops at the center of the test area. If z-field cancellation is not used, bubbles in

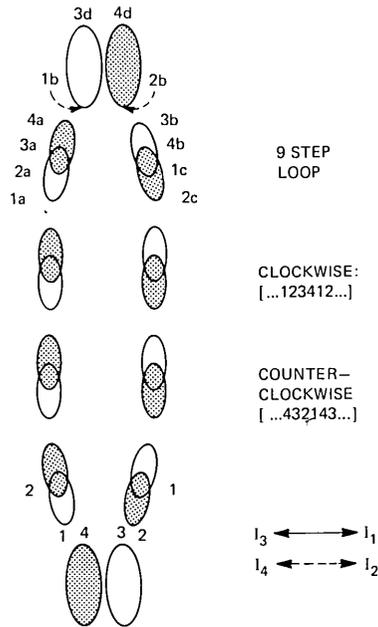


Fig. 39—A 9-step dual-conductor closed loop. The elliptical apertures in the turns have been lengthened for the reasons described in the text.

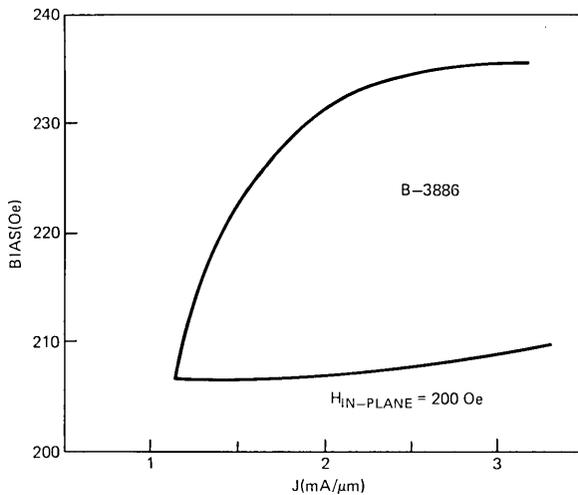


Fig. 40—Bias margin vs drive for a 6- μm period, 7-step loop operated at 1 MHz. The garnet was unimplanted.

loops located near the outer edges of the test area are biased by both an alternating and a constant field. A typical failure mode (Fig. 39) finds bubbles in clockwise motion, jumping to 3d from 2b rather than stepping to 3b. Some modifications were necessary to cope with this problem.

First, the two slots in the turn were lengthened further to 1.25λ and, second, additional apertures were placed beyond the turn itself to help equalize the current distribution. The latter is readily accomplished by interconnecting loops end to end (Fig. 41). Note that the bubbles in adjacent loops move in paths counter to each other just as the gears of a gear train.

Impetus for making the modifications given above came from test circuits in which interconnecting loops spanned the width of the test area. Loops with lengthened turn-elements outperformed all the others in positions 1 and 5, whereas all loops were satisfactory in positions 2, 3, and 4.

These experiments also gave additional evidence of the "gyro" nature of the bubbles in our materials. Consistently, the even-numbered loops outperformed the odd-numbered loops for one direction of

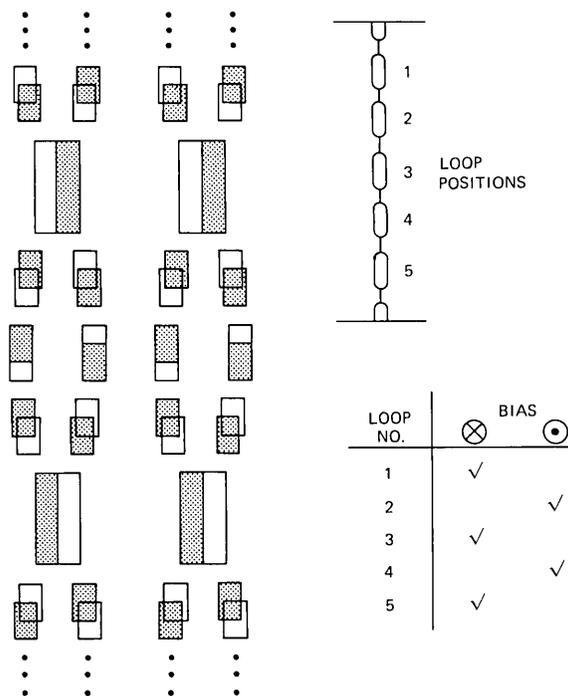


Fig. 41—Loops can be interconnected end-to-end by the method shown. The propagation direction of adjacent loops alternates, i.e., cw, ccw. Performance is related to the propagation sense and bias field direction.

bias and vice versa. One additional observation not as yet understood was made: Operation was generally better if the bias field was directed toward rather than away from the garnet surface.

7.3 Propagation parallel to current flow

In this section, propagation parallel to the direction of general current flow is considered. This is not an absolutely essential operation for a chip layout, since such motion can also be made by combinations of motions normal to current flow and the π -turns discussed in the preceding section. In fact, our earliest shift registers were designed that way. The principle of operation, from Fig. 42a, is straightforward. A bubble starting at the leftmost position 1a proceeds through 1a2a3a4a1234 in response to pulse sequence [123412341].

The actual operation is far from ideal, since propagation down the center track is stable only for a restricted (and oftentimes non-existent)

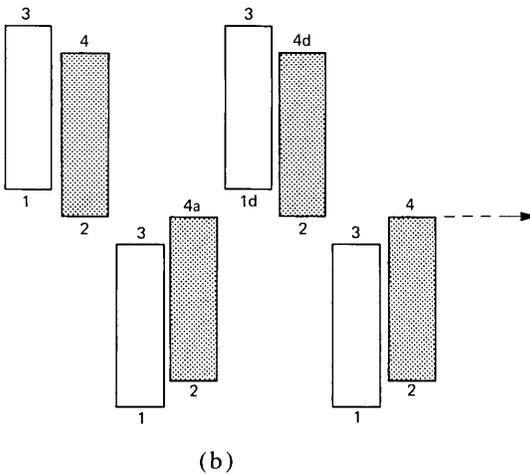
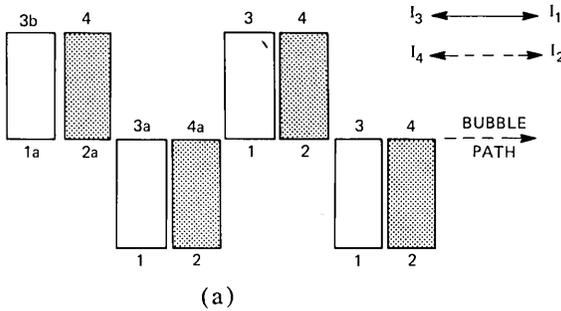


Fig. 42—Circuit for propagation parallel to the current. The original bidirectional design (a) was modified (b) to be unidirectional to improve performance.

set of drive and bias conditions. It is observed that a bubble in transit through 1a2a jumps to 3b rather than proceeding to 3a as desired. This occurs since the current density at 3b, and thus its pole strength, is substantially greater than at 3a, thereby counteracting 3b's increased distance from 2a.

Operation improves substantially if the apertures are increased in length and staggered as shown in Fig. 42b. This modification further removes the offending poles from the intended bubble track and also forces increased current along the bubble track. As a result, the circuit of Fig. 42b gives adequate margins for motion [...12341234...], however, motion along [...43214321...] is unreliable because of encounters such as 1d to 4d rather than 4a.

Operating margins for an 8- μm period shift register stepped at a 1-MHz rate are given in Fig. 43. The garnet was not implanted nor was an in-plane field applied. The performance is seen to be quite good.

7.4 Bubble generation

A nucleate bubble generator can either be patterned in one of the propagation conductor levels or it can be provided in a separate conductor level. Only the former design is considered here, since the design of the latter is obvious. Now it is possible to introduce slots in a sheet conductor if the slots are narrow ($1/4\lambda$ or less) and run essentially parallel to the current flow. Such slots can be used to concentrate current and thus nucleate bubbles, yet permit current flow during normal propagation. Just how this can be done is seen in Fig. 44.

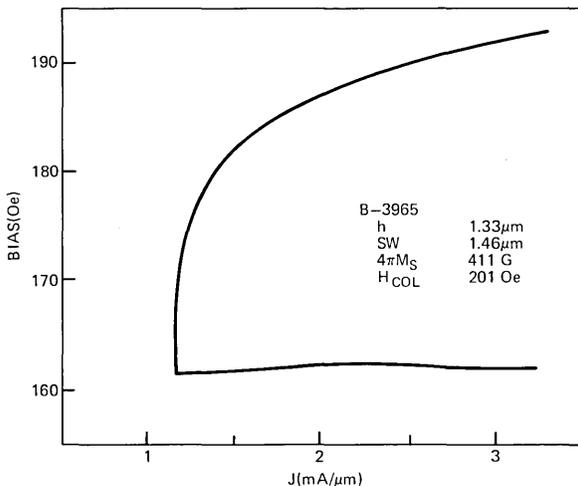


Fig. 43—Bias range vs drive at 8- μm period and 1 MHz for circuit (b) of Fig. 42. The garnet was not implanted.

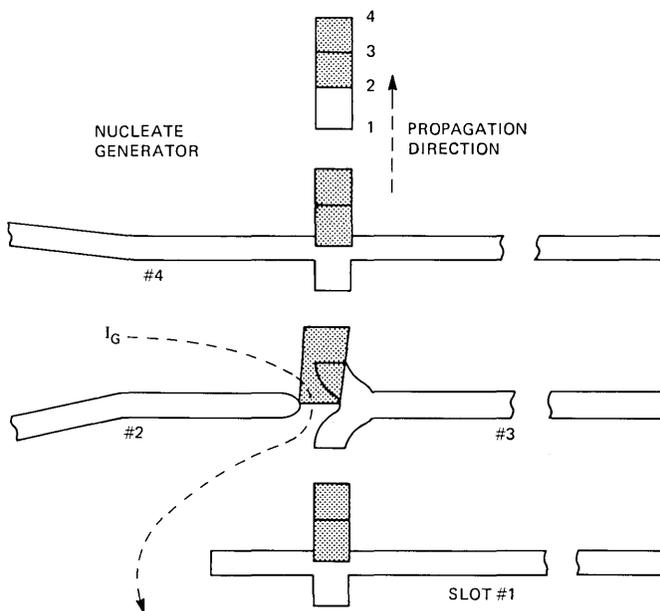


Fig. 44—A nucleate generator can be patterned as part of a conducting level. Slots guide the generator current but do not interfere with normal propagation. Bubbles are nucleated at the tip of slot #2.

The generator is incorporated into level 1 since that ensures the lowest possible generate current. Slots #2 and #4 funnel the generator current I_g from the input pad into the generator proper. A continuation of slot #4 acting in conjunction with slots #1 and #3 forms relatively high resistance paths so that most of the incident current I_g concentrates at the tip of slot #2, nucleating a bubble, before diffusing as it enters the chip area below.

Generator operation has been characterized in a nominally 8- μ m period shift register processed on wafer B3954. The drive conditions were: amplitude 1.5 mA/ μ m, width 1.5 μ s with 33 percent overlap. The generate pulse was applied in conjunction with propagate pulse 4. Subsequent propagate pulses move the newly generated bubble along the shift register track. Both adjacent and isolated domains were generated. The 200-ns wide generate pulse had an amplitude range of 230 to 360 mA.

The composite bias margins for generation at and propagation away from the generator were 19 Oe, with the lower end of the margin at stripout. For propagation through the generator, the margins reduce to 15 Oe with most of the margin loss at the low end. To ensure proper motion of bubbles as they pass upwards through the generator structure, a current density $J_{1,3}$ must be maintained between slots #3 and

#4. In our experiments, a single resistor diverted a fraction of the drive current $I_{1,3}$ into the generator pad.

7.5 Detection

Field-access permalloy devices use a chevron expander detector in which a bubble domain is stretched laterally before detection takes place in a magnetoresistance sensor. This approach has the advantage that it permits unlimited expansion without any effect on the data rate of the chip. The exit portion of the detector can be designed either to shrink the strip domain back to a bubble or to discharge the strip domain into a guard rail. The detector we report is similar to that just described and to a detector used with the bubble lattice.⁸

Bubble expansion and contraction is accomplished in the structure of Fig. 45. The analogy to a chevron expander structure is apparent. Factors relating J and \bar{H}_z for an array of parallel conductor strips were introduced in Section III and are detailed further in Section X. In typical geometries, a current density $J = 4 \text{ mA}/\mu\text{m}$ generates a bias field decrement of nearly 20 Oe. In other words, stripout in the detector will be sustained over a 20-Oe range of the bias field. Since all the dual-conductor functions operate down to the strip-to-bubble transition field, we can expect to realize most of that bias field range.

Design curves to estimate the in-plane field H_x due to (i) the strip domain and (ii) the drive conductor strips are given in Figs. 46 and 47, respectively. At 8- μm period, assuming $4\pi M_s = 500G$ and $J = 4\text{mA}/\mu\text{m}$, we expect $H_x = 100 \text{ Oe}$ for a typical strip domain and a maximum field of 20 Oe (depends on detector-to-conductor strip positioning) from the drive conductors. The magnitude of these fields suggests that a thin-permalloy magnetoresistive sensor can be used. A photograph

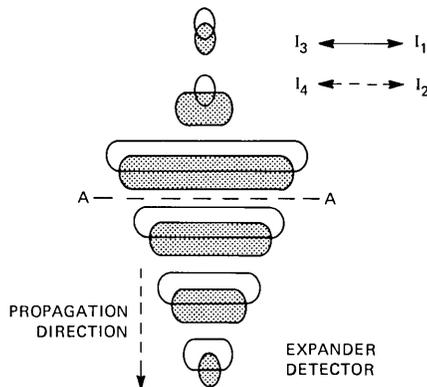


Fig. 45—Details of a dual-level expander detector. Bubbles, entering from the top, are expanded for either magnetoresistive or inductive sensing at A-A.

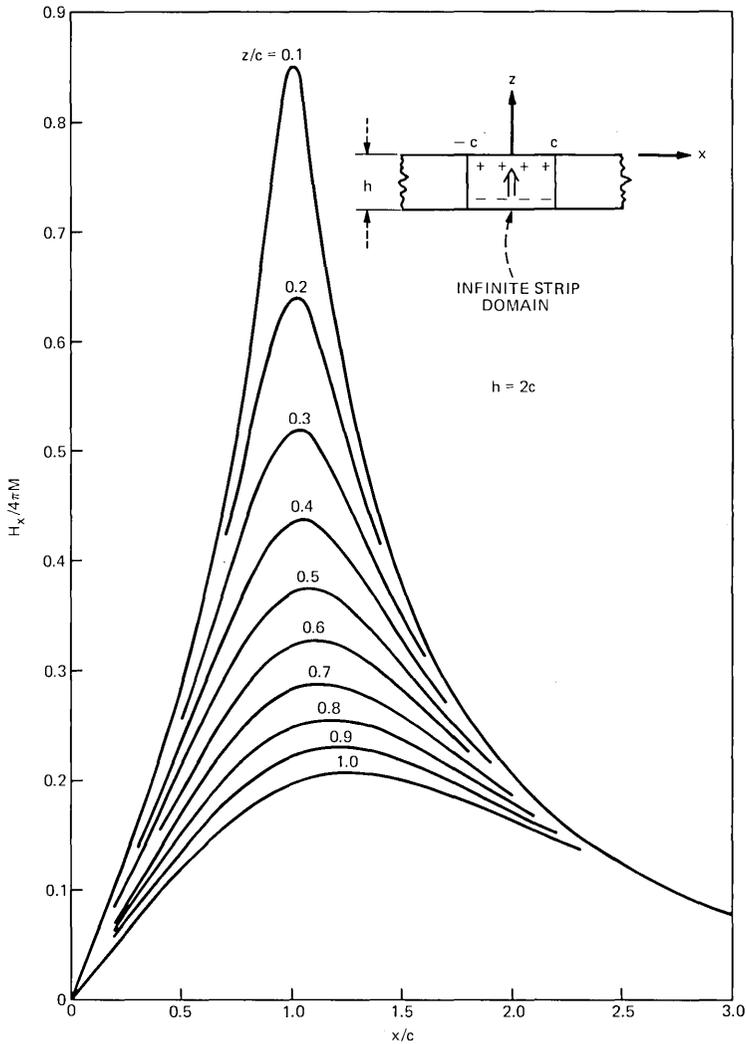


Fig. 46—Calculated in-plane field from a strip domain effective on a magnetoresistive sensor.

of the conductor-expander section complete with a “Chinese character” sensor is seen in Fig. 48. The sensor is processed on top of the pair of slotted conductor sheets.

In current-access devices, there is the opportunity to design a chip without permalloy. For example, if a strip domain is driven by an RF component of the bias field, the variation in the strip’s flux can be sensed by a conductor pickup loop. This scheme is being pursued by J. M. Geary, and he has allowed us to present some of his preliminary

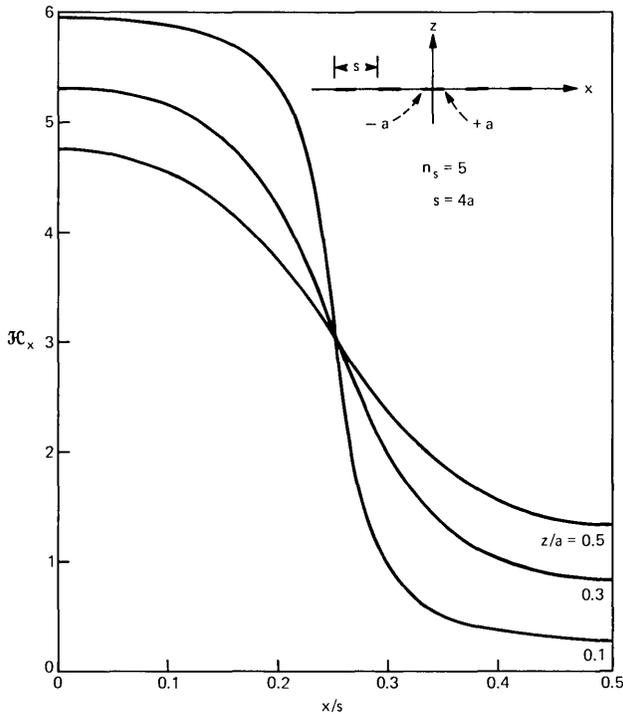


Fig. 47—Calculated in-plane field from the parallel drive strips of an expander detector.

results. A single-turn pickup loop $20\ \mu\text{m}$ in length was patterned in a conducting sublayer at “A...A” of Fig. 45. An output signal of $5\ \mu\text{V}$ was obtained when the width of a strip domain was modulated at 8 MHz (Fig. 49). RF detection in magnetic devices is not new and, in particular, we refer to an excellent article by Benrud et al., which deals with RF detection in thin permalloy-film memories.¹⁶

7.6 Transfer

Transfer is a particularly useful function since, combined with propagation, generation, and detection, it completes one set of functions that can be used to design a major-minor chip. This is significant, as it is well known that the major-minor organizations give improved performance over single loop shift registers. More important is that a dual-conductor major-minor chip can be partitioned into individually accessed regions with a resultant reduction in chip power dissipation.

One obvious way to implement a transfer is to use a “transfer” conductor patterned as part of a third conductor level to gate bubbles from minor storage loops to a major track and back again in much the

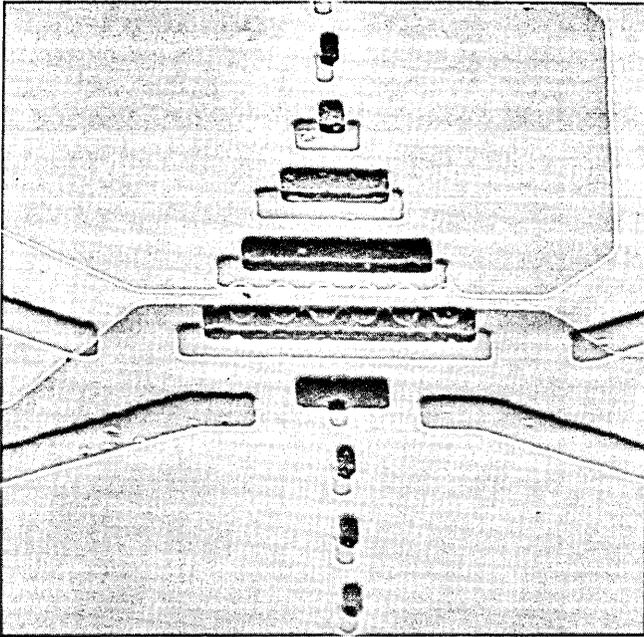


Fig. 48—An SEM photograph of an expander detector with a thin permalloy magneto-resistive sensor. The circuit period is $8\ \mu\text{m}$.

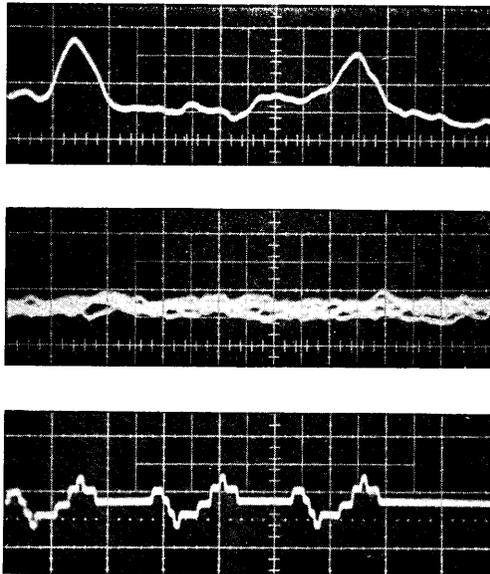


Fig. 49—Strip domains, excited by an 8-MHz RF signal, are detected in a single-loop. The raw signal, approximately $5\ \mu\text{v}$ in amplitude, amplified and demodulated is shown in the upper trace. The bias field was taken above bubble collapse for the middle trace. A composite of the drive current is shown in the lower trace. Horizontal scale is $20\ \mu\text{s}/\text{div}$.

spirit of “field access” permalloy structures. However, perhaps one less processing level will be needed if the transfer can be accomplished within the framework of the propagate conductor levels themselves.

Such a compatible transfer gate can be realized by nesting apertures within one another as per the upper turn of the closed loop in Fig. 50. When this arrangement is used, it is only necessary to modify the propagation pulse sequence to cause bubbles in the transfer position to move into exit paths, whereas bubbles not in the transfer position are only idled backward and forward and always in a manner consistent with the retention of data when normal propagation resumes. Returning to Fig. 50, bubbles in the nine-step loop propagate clockwise with the standard pulse sequence [...12341234...]. However, a bubble located at the transfer position “1a” is transferred to position “4b” by the sequence [123412141234] in which the steps associated with the

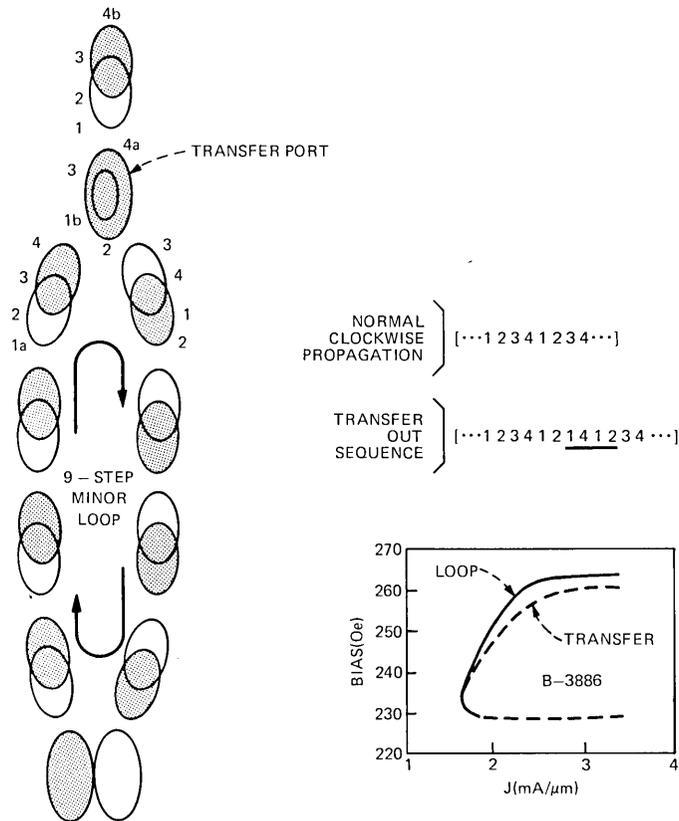


Fig. 50—Design and performance of a transfer port. In the transfer itself, the first level slot is nested within the slot of the second level. The propagation sequence is reprogrammed to transfer.

actual transfer have been underlined. It should be remembered that if data are to be preserved elsewhere, it is important to use only next number transitions in the transfer sequence.

Margins of propagation and transfer taken at 400 kHz are also included in Fig. 50. It can be seen that this transfer has very good margins indeed. More recent data have shown that, when the pulse which moves a bubble from 1b to 4a (4 in the underlined sequence) is lengthened by $0.25 \mu\text{s}$, then an otherwise 1-MHz operation can be maintained. That is, only even-odd number transitions are permitted.

7.7 Dual-conductor chip design considerations

There are two distinct aspects to the minimization of dissipation in dual-conductor devices. We discuss the minimization of power consumption on a local basis in Section 7.7.1 and on a chip basis in Section 7.7.2.

7.7.1 Optimization of conductor thickness

Power dissipation will always be a concern in current-access devices. For example, for $J = 1.5 \text{ mA}/\mu\text{m}$, $\lambda = 8 \mu\text{m}$ and $R = 0.1 \Omega/\square$, the continuous power dissipation is $14.4 \mu\text{W}/\text{b}$. Even at $J = 1 \text{ mA}/\mu\text{m}$ and $\lambda = 4 \mu\text{m}$, the power dissipation is $1.6 \mu\text{W}/\text{b}$. In this section, we turn to the problem of finding the correct thickness for conductors 1 and 2 under the constraint of an intervening insulator of thicknesses. From a practical standpoint, the insulator must withstand any voltage that exists between the two levels. The drive conductors are usually commoned at one end so that the maximum voltage encountered is the sum of the voltage drops of the layers. For the chip design discussed in the next section, the maximum inter-conductor voltage is 7.5 V.

For the purpose of calculating relative power consumption in a 2-conductor device, regard the fields produced by each level as those from an infinitesimally thin conducting sheet at the median plane of the actual conductor. For numerical work, the fields computed from eq. (25) will be used. When the two-conductor levels produce the same field, the relative power per unit area, normalized to the hole diameter $2a$, is

$$P = 2a/t_1 \bar{\mathcal{H}}_1^2 + 2a/t_2 \bar{\mathcal{H}}_2^2, \quad (10)$$

where t_1 and t_2 are the conductor thicknesses. The conductor of thickness t_1 is nearer the garnet. The two-conductor levels are separated by an insulator s units thick, as shown in Fig. 51. Values for P were computed at the maximum of $\bar{\mathcal{H}}$ near $r \approx a$ for $h/a = 0.8$ subject to the height constraint

$$a\tau = t_1 + s + t_2. \quad (11)$$

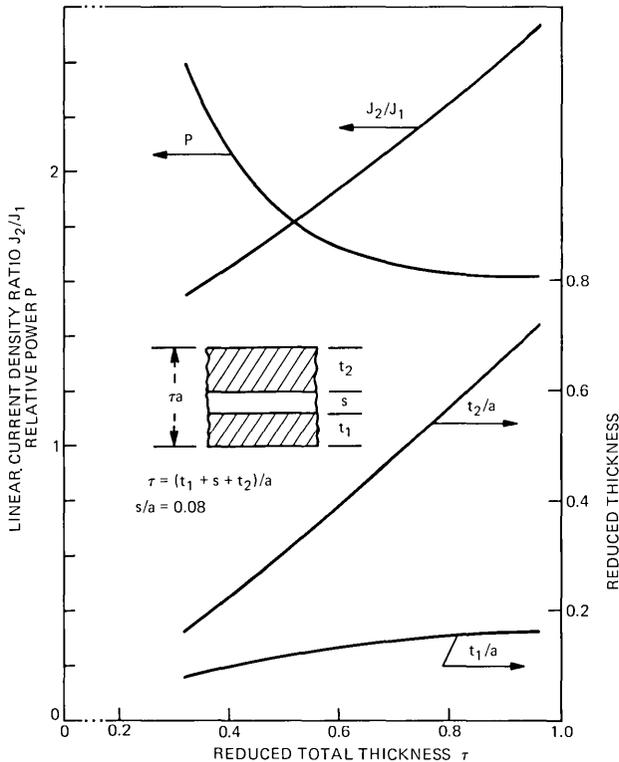


Fig. 51—Nomograph to determine the optimum thicknesses of levels 1 and 2 to minimize power dissipation in a dual-conductor circuit. Analysis based on a circular aperture of diameter $2a$ and garnet thickness $h = a$.

For each of several τ values, the minimum of eq. (10) was found. The results are shown in Fig. 51 together with the ratio of the current densities J_2/J_1 for the two levels. This ratio follows from the requirement of equal fields from the levels: $J_2/J_1 = \bar{\mathcal{H}}_1/\bar{\mathcal{H}}_2$.

The behavior of P as a function of τ represents a trade-off between efficient field production— $\bar{\mathcal{H}}$ large—and efficient power transmission in the conductor levels— t large. The best compromise between these conflicting requirements occurs for $\tau \approx 1$. The broad minimum of P about $\tau \approx 1$ permits a wide choice of τ with only minor changes in power consumption. For example, a 50-percent reduction of τ from 0.96 to 0.48 causes only a 16-percent increase in P . For current experimental devices, typical dimensions are $t_1 = 0.25 \mu\text{m}$, $s = 0.15 \mu\text{m}$, and $t_2 = 0.35 \mu\text{m}$. Using $a = 2 \mu\text{m}$, one finds $\tau = 0.38$. Although a larger value does reduce power consumption, it also increases processing complexity.

7.7.2 Design of a major-minor chip with partitioning

We now introduce partitioning of the conducting layers in a dual-conductor shift-register or major-minor chip design. In a partitioned chip, the main storage area and the major track are independently driven with a resultant reduction in chip dissipation. The shift-register and the major-minor chip layouts are very similar except that the shift-register uses parallel tracks rather than minor loops. Both of these designs allow for redundancy in the storage area. Given the functions of propagation, generation, annihilation, and transfer, the design of dual-conductor chips proceeds much as for field-access chips. There are, however, considerations peculiar to these current-accessed chip designs that we shall bring to your attention in this section.

The design and performance characteristics of a hypothetical major-minor chip with an 8- μm period and 262,144-b capacity is outlined in Fig. 52. Similar designs have been proposed by J. L. Smith and P. I. Bonyhard. Bubbles are assumed to move in bursts at a 1-MHz stepping rate. The chip is partitioned into independently accessed storage and detector-major track zones. The resistance everywhere is assumed to be 0.1 Ω /square. Each zone is driven at the current density indicated in the figure. If the entire chip was run continuously at 1 MHz, it would dissipate over 4 W. With partitioning, these zones can be independently accessed, reducing the chip dissipation substantially with no sacrifice in performance. In a partitioned chip it is, of course,

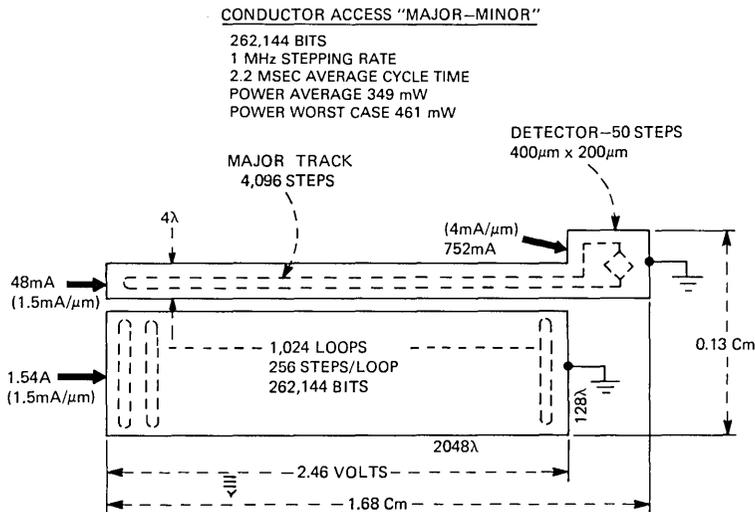


Fig. 52—Proposed design of a partitioned "major-minor" dual-conductor chip. The chip is made rectangular to increase the drive impedance and partitioned to permit selective drive to reduce the on-chip power dissipation.

necessary to move bubbles between zones. Propagation across slots was discussed with generators in Section 7.4.

It is desirable to make the impedance of the storage area compatible with driver circuitry. If the geometry of the storage area is laid out as a square, then it will present about 0.1Ω resistance to its drivers—completely independent of the storage capacity. However, if the conductor sheet is reconfigured as a rectangle, its resistance can be “transformed” into a better match for semi-conductor drivers.

If the resulting “stick-like” chip is too awkward to handle, it can be folded, since the current (and bubbles) will have no difficulty negotiating the turns. A rectangular chip does have the advantage that it can dissipate heat more readily than a square chip and that it is well configured for a coilless package.

The number of minor loops is determined by noting that 1024 loops are, in binary multiples, the most that can be driven from a 5 V supply. This storage area would dissipate 3.77 W if driven continuously, but its duty factor is very low (0.03) so the average power dissipation is just 110 mW. The detector and major track do run essentially continuously; however, they together represent a very small fraction of the chip area and, in total, dissipate only 239 mW. As a result, the total chip dissipation is but 349 mW. Further information on cycle time, duty factor, and power dissipation is summarized in Table II. If the chip capacity is quadrupled (1024 b/loop), the chip dissipation would only double and the cycle times would remain about the same.

VIII. EPITAXIAL GARNET MATERIAL

In this section, we examine the material requirements for current-access bubble devices. The specific material parameters of particular importance to this technology are discussed and a garnet system proposed which allows relatively high-domain wall mobility along with adequate uniaxial magnetic anisotropy.

8.1 Growth and characterization

All compositions grown for the current-access high-density bubble devices were grown on $Gd_3Ga_5O_{12}$ substrates by the LPE dipping

Table II—Characteristics of partitioned 262-kbit dual-conductor chip

Zone	Steps	Transit Time (ms)	Average Case			Worst Case		
			Cycle Time	Duty Factor	Power (mW)	Cycle Time	Duty Factor	Power (mW)
Minor loops	256	0.256	0.128	0.03	110	0.256	0.06	230
Major track and detector	4146	4.146	4.146	0.97	239	4.146	0.94	231
			4.274	1.00	349	4.402	1.00	461

technique using supercooled melts.¹⁷ Axial rotation was maintained constant at approximately 100 rpm during the growth period. The details of the apparatus and growth techniques have been discussed elsewhere.¹⁸⁻²¹ Growth temperatures were restricted to $875 \pm 60^\circ\text{C}$. Lattice parameter measurements were obtained using the HPM method.²² Magnetic property measurements were obtained using the standard techniques.²³⁻²⁵

8.2 Material properties

Table III lists the parameters which must be considered when designing a material to be used in dual conductor-current access devices. Items 1 to 11 in Table III are important parameters to be considered when designing a bubble material for use in most bubble circuit technologies. Items 11 to 13 are of particular importance for dual conductor-current access devices. Because of the high-speed performance inherent in this technology, the mobility of the bubble material is of primary importance. Dynamic coercivity and bubble propagation angle are two other parameters that become increasingly important in small bubble devices. Low dynamic coercivities allow lower drive currents to be used for bubble propagation. A discussion of the importance of the bubble propagation angle is presented in Section 7.3. An approach to increasing the mobility while maintaining adequate q values as well as an approach to lowering the dynamic coercivity is presented later in this section.

Epitaxial garnet films of nominal composition $\text{Y}_{1.2}\text{Sm}_{0.4}\text{Lu}_{0.5}\text{Ca}_{0.9}\text{Ge}_{0.9}\text{Fe}_{4.1}\text{O}_{12}$ have been shown to be useful in $16\text{-}\mu\text{m}$ period field-access bubble devices.²⁶ For $3\text{-}\mu\text{m}$ diameter bubbles, this composition offers the advantages of domain wall mobility slightly in excess of $250\text{ cm s}^{-1}\text{ Oe}^{-1}$ along the q values of 5 and precise control of the temperature dependence of the magnetic properties by the appropriate adjustment of melt chemistry. Figure 53 shows the material length

Table III—Material parameters for current-access devices

1. Saturation magnetization, $4\pi M_s$, (Gauss)
2. Exchange constant, A (ergs cm^{-1})
3. Uniaxial anisotropy, K_u (ergs cm^{-3})
4. Material length parameter, l (μm)
5. Anisotropy field, H_K (Oe)
6. Quality factor, q (dimensionless)
7. Bubble collapse field, H_{col} (Oe)
8. Lattice parameter, L.P. (\AA)
9. Magnetostriction coefficient, λ_{111} , λ_{110}
10. Temperature dependence of H_{col} , ($\% \text{ }^\circ\text{C}^{-1}$)
11. Bubble mobility, μ (cm/sec-Oe)
12. Dynamic coercivity, ΔH_c (Oe)
13. Bubble propagation angle, ψ (degrees)

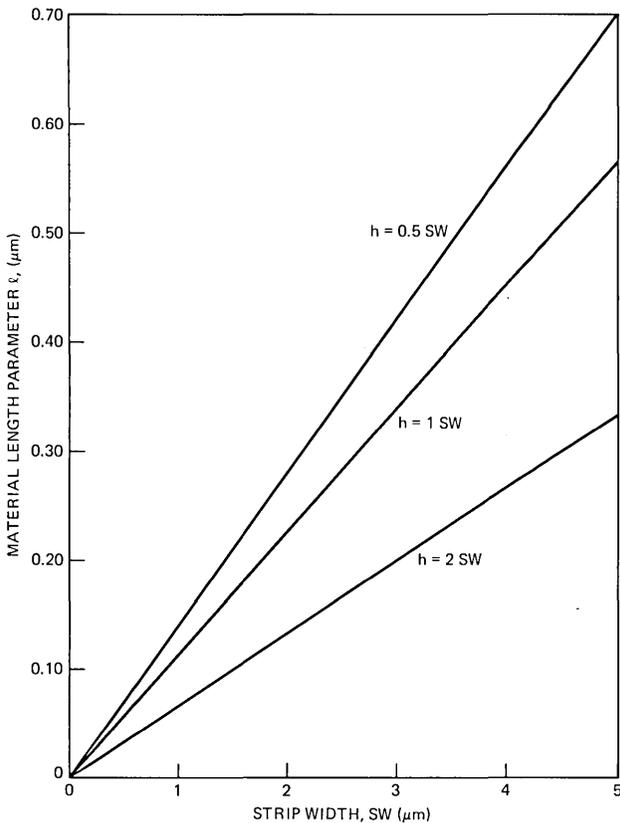


Fig. 53—Material length parameter (l) vs strip width (sw) with the thickness-to-strip-width ratio as a parameter.

parameter (l) as a function of domain strip width (sw) with the thickness-to-strip-width (h/sw) ratio as a parameter. Depending upon the h/sw ratio, the l parameter needed for a strip width of $1.7 \mu\text{m}$ ranges from 0.115 to $0.235 \mu\text{m}$. For the case of $h/sw = 1$, the material length parameter must be $0.191 \mu\text{m}$.

Extending the $(\text{YSmLuCa})_3(\text{FeGe})_5\text{O}_{12}$ system to smaller bubbles results in the nominal composition $\text{Y}_{1.0}\text{Sm}_{0.5}\text{Lu}_{0.7}\text{Ca}_{0.8}\text{Ge}_{0.8}\text{Fe}_{4.2}\text{O}_{12}$ which exhibits $1.7\text{-}\mu\text{m}$ diameter bubbles and the properties listed in Table IV. Materials grown from the general system $(\text{YSmLuCa})_3(\text{FeGe})_5\text{O}_{12}$ have exhibited dynamic coercivities ranging from 1.4 Oe to 3.2 Oe . The higher coercivities being measured are on the larger anisotropy materials, supporting smaller diameter bubbles. Because of the particular importance of bubble mobility in the current-access technology, a technique was needed to maintain a q value of

approximately 3 to 4 for materials exhibiting an l value of 0.16 to 0.19 μm while also increasing the domain wall mobility.

It would be advantageous to be able to decrease the samarium content of the composition listed in Table IV while maintaining the q value, which would result in an increased mobility. One approach to achieving this is to increase the concentration of the most "anisotropy active" constituents. Figure 54 shows the calculated uniaxial anisotropy as a function of the lutetium content of garnet compositions in the three systems $(\text{YLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$, $(\text{LaLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$, and $(\text{LaLuSm})_3(\text{FeGa})_5\text{O}_{12}$. The samarium content and the Curie temperature are used as parameters. Curves A through C are calculated for Ca-Ge substitution of ~ 0.8 moles per garnet formula unit and a Curie temperature of $\sim 480^\circ\text{K}$. Curves D and E represent gallium substitution leading to a Curie temperature of $\sim 440^\circ\text{K}$. We point out that the curves shown in Fig. 54 do not take into account the lattice constant constraint usually imposed on epitaxial garnet films. The compositions required to achieve a lattice constant match with $\text{Gd}_3\text{Ga}_5\text{O}_{12}$ (GGG) substrates (12.383\AA) are indicated by the point on each curve marked L.M. Comparison of curves A and B in Fig. 54 reveals that, to decrease the samarium content from 0.5 to 0.3 in the garnet formula and maintain both the anisotropy of ~ 35000 ergs- cm^{-3} and a lattice parameter match with the GGG substrate, compositions outside the $(\text{YLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$ system must be considered. Curve C indicates that the $(\text{LaLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$ system allows both the uniaxial anisotropy of ~ 35000 ergs- cm^{-3} and a lattice match to be achieved. A discussion of the advantages of lanthanum have been presented elsewhere.²⁷

It is well known that materials with high $4\pi M_s$ require relatively little diamagnetic ion substitution. The advantages realized in the divalent-tetravalent ion substitution system can be obtained with less complicated crystal compositions when high moment material is required.²⁸ For materials exhibiting a high $4\pi M_s$, Ca^{2+} - Ge^{4+} substitution could result in too high a Curie temperature making it difficult to

Table IV—Composition:
 $\text{Y}_{1.0}\text{Sm}_{0.5}\text{Lu}_{0.7}\text{Ca}_{0.8}\text{Ge}_{0.8}\text{Fe}_{4.2}\text{O}_{12}$

Lattice parameter	12.383 \AA
$4\pi M_s$	510 G
K_u	36,600 ergs cm^{-3}
l	0.19 μm
A	2.7×10^{-7} ergs cm^{-1}
H_k	1800 Oe
q	3.5
μ	~ 300 cm/S-Oe
ΔH_c (dynamic coercivity)	3.0 Oe

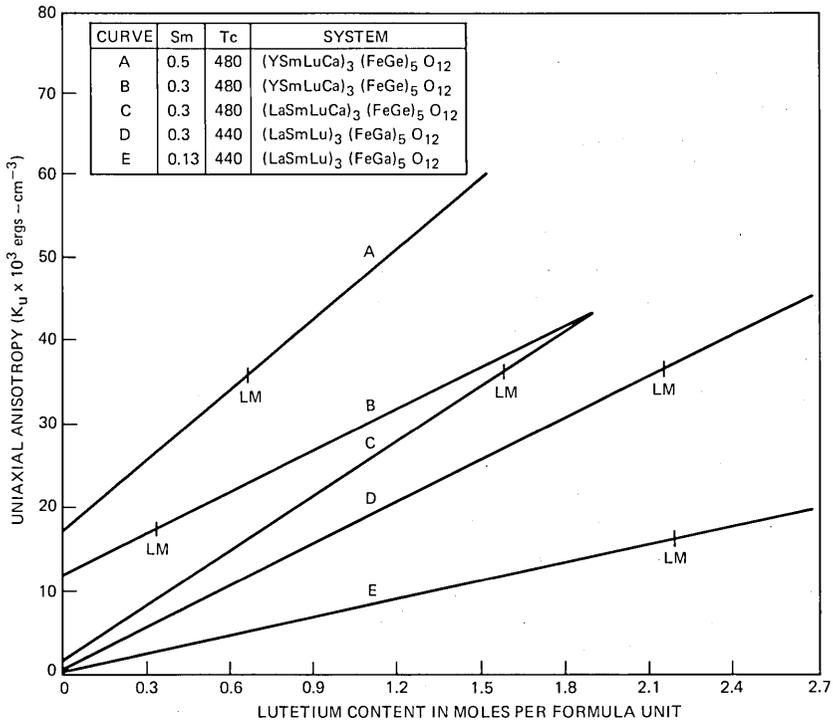


Fig. 54—Calculated uniaxial magnetic anisotropy vs lutetium content for the systems (YSmLuCa)₃(FeGe)₅O₁₂, (LaSmLuCa)₃(FeGe)₅O₁₂, and (LaLuSm)₃(FeGa)₅O₁₂. The samarium content and the Curie temperature are varied. The GGG lattice matching compositions are indicated by the symbol LM on each curve.

control the temperature dependence of the magnetic properties to match the barium ferrite bias magnets.

A slight imbalance of the divalent-tetravalent ion ratio could also result in increased coercivity. The immersion of a gallium garnet substrate into a melt supersaturated with an iron garnet containing Ca²⁺-Ge⁴⁺ might be expected to show increased substrate dissolution prior to initial film growth as compared to the same substrate immersed in a melt supersaturated with a gallium containing iron garnet. Considering the complications in melt chemistry and the careful control of growth conditions necessary to minimize compositional fluctuations in Ca²⁺-Ge⁴⁺ substituted garnets, a move to simpler film compositions would be advantageous.

Curve D illustrates the anisotropy calculated for the (LaLuSm)₃(FeGa)₅O₁₂ system. Note that the lattice match composition occurs at a lutetium concentration of ~2.15 moles per garnet formula unit and that a uniaxial anisotropy of ~35000 ergs-cm⁻³ can be obtained. Table V lists a nominal composition in the

(LaLuSm)₃(FeGa)₅O₁₂ system along with some measured magnetic properties. Curve E shows that lowering the samarium concentration to 0.13 moles per garnet formula unit results in a significantly lower anisotropy for the GGG lattice matching composition.

8.3 Dynamic coercivity

The dynamic coercivity of epitaxial bubble garnet films supporting ~1.7- μ m diameter bubbles has been measured as a function of film thickness.²⁹ Figure 55 shows the dynamic coercivity measured by a bubble translation technique as a function of inverse film thickness. All curves are for material exhibiting 1.7- μ m diameter bubbles. Curve C is for a sample in the (YLuSmCa)₃(FeGe)₅O₁₂ system, while curves A and B represent two compositions in the (LaLuSm)₃(FeGa)₅O₁₂ system. Curve A is for a samarium concentration of ~0.13 moles per garnet formula unit, while curve B represents data taken on a sample containing a samarium concentration of 0.3 moles. Note that the gallium material exhibits significantly lower dynamic coercivities than does the Ca²⁺-Ge⁴⁺ substituted material.

8.4 Melt chemistry

The phase equilibria observed in the (YSmLuCa)₃(FeGe)₅O₁₂ system have been presented previously.¹⁹ Lanthanum does not form an iron garnet, and lutetium iron garnet has been reported being grown only once in bulk form,³⁰ possibly because of phase equilibria difficulties. As the concentration of lanthanum increases in the pseudo-ternary Flux—Fe₂O₃— Σ Ln₂O₃, the garnet phase field would be expected to narrow considerably. Figure 56 shows a section of the pseudo-ternary Flux—Fe₂O₃— Σ Ln₂O₃, where Σ Ln₂O₃ corresponds to the sum of La₂O₃+Lu₂O₃. The ratio of La₂O₃/Lu₂O₃ was fixed at 0.332 to allow stress-free epitaxial film growth on GGG substrates. Note that all data shown in the pseudo-ternary are for unsubstituted iron garnets. Substitution of diamagnetic ions could be expected to shift the location of the phase boundaries. Addition of gallium results in the garnet-perovskite boundary being displaced towards the Flux—Fe₂O₃ binary. Depending on the overall garnet oxide concentration in the melt, the

Table V—Composition:
La_{0.6}Lu_{2.1}Sm_{0.3}Ga_{0.9}Fe_{4.1}O₁₂

$4\pi M_s$	466G
K_u	35,500 ergs cm ⁻³
A	2.0×10^{-7} ergs cm ⁻¹
l	0.18 μ m
H_K	1915 Oe
Q	4.1
μ	500-1000 cm/S-Oe
ΔH_c (dynamic coercivity)	~2 Oe

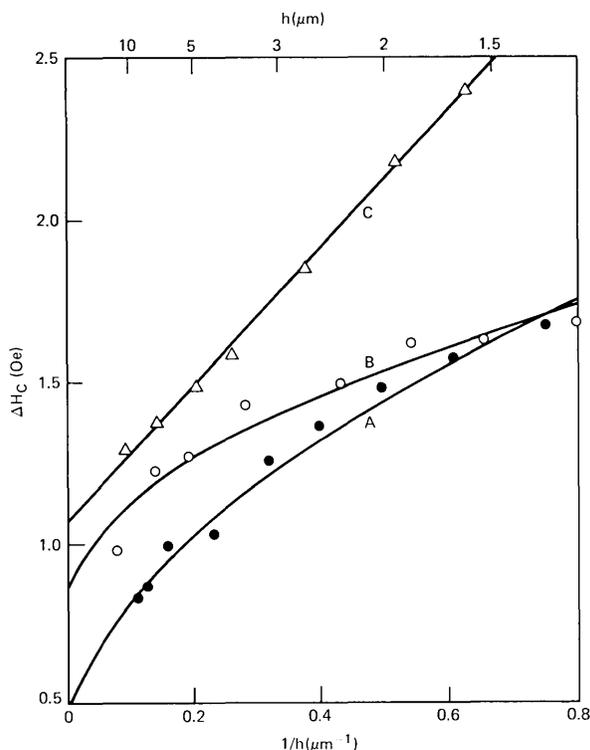


Fig. 55—Bubble dynamic coercivity (ΔH_c) as a function of inverse film thickness for 1.7- μm bubble material. Curve C is for $(\text{YLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$. Curves A and B represent two compositions in the system $(\text{LaLuSm})_3(\text{FeGa})_5\text{O}_{12}$. Curve A is for a samarium concentration of ~ 0.13 moles per garnet formula unit. Curve B represents data on a sample containing 0.3 moles of samarium.

range of $\text{Fe}_2\text{O}_3/\sum\text{Ln}_2\text{O}_3$ over which garnet is found to be the primary phase field is from 6 to 9. For comparison, the $\text{Fe}_2\text{O}_3/\text{Y}_2\text{O}_3$ ratio range for garnet in the Flux— Fe_2O_3 — Y_2O_3 pseudo-ternary is from 12 to 40.

Because of the large size of La^{3+} , one would expect a relatively low value for the La^{3+} distribution coefficient. The definition for the distribution coefficient of lanthanum is:

$$k^{\text{La}} = \frac{\left(\frac{\text{La}}{\text{La} + \text{Lu} + \text{Sm}} \right)_{\text{crystal}}}{\left(\frac{\text{La}}{\text{La} + \text{Lu} + \text{Sm}} \right)_{\text{melt}}},$$

where the symbols in the fractions represent the number of moles present. Although the value for the effective coefficient will be rate-dependent, the value for k^{La} under average growth conditions is found to be ~ 0.42 .

The use of compositions in the $(\text{LaLuSm})_3(\text{FeGa})_5\text{O}_{12}$ system allows less complicated crystal compositions to be used. Lower dynamic coercivities and increased domain wall mobility can be obtained.

IX. PROCESSING

There are two major concerns in processing dual conductor structures. First, the two conductors should both be as close as possible to the epitaxial film supporting the bubbles. Hence, a prespacer (analogous to the prespacer used between conductor and epitaxial layer in standard field access devices) should be as thin as possible or completely omitted. In addition, it is immediately obvious that the dielectric layer between the conductors must be as thin as the limits of dielectric breakdown permit. Second, the very small period devices possible within the material constraints of this technology will require increasingly precise control for etching the fine-aperture metal film pattern. What is required is an etching method that will produce metal patterns dimensionally identical to the etch mask.

9.1 Material requirements

9.1.1 Insulating films

If a prespacer is used, the material and/or process selection is quite unrestrained as the surface is completely planar and the layer thickness

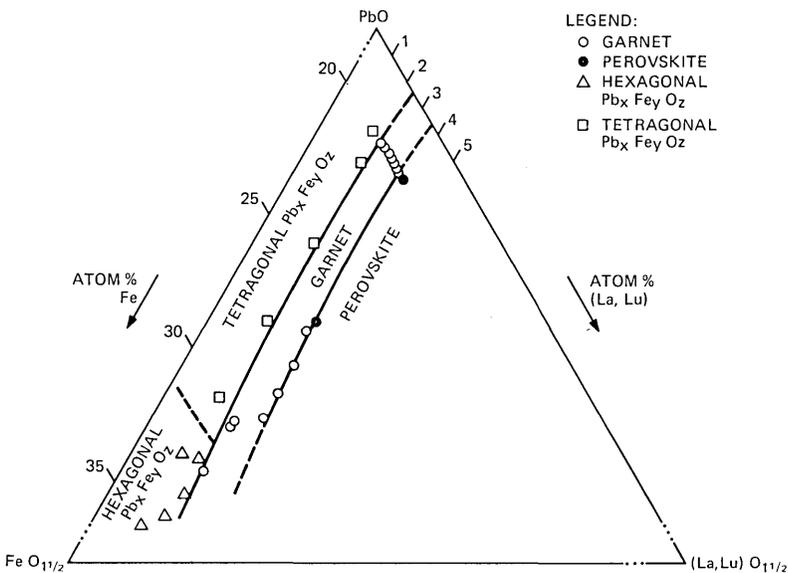


Fig. 56—A section of the pseudoternary Flux— Fe_2O_3 — $\Sigma\text{Ln}_2\text{O}_3$, where $\Sigma\text{Ln}_2\text{O}_3$ represents $\text{La}_2\text{O}_3 + \text{Lu}_2\text{O}_3$.

required is very small. Thus, several avenues can probably be used: RF diode-sputtered SiO_2 or SiN ; RF magnetron or S-gun deposited SiO_2 or SiN ; plasma-deposited or CVD-deposited SiO_2 or SiN . In all the work reported, prepacers were SiO_2 deposited by RF plasma deposition using an Applied Materials Plasma I radial flow system.

The insulating layer between conductors has more stringent requirements due to the multiplicity of crossovers and the need for minimum thickness. Thus, very good dielectric properties are required and a process that deposits a conformal coating is mandatory, i.e., equivalent deposition rates on vertical and horizontal surfaces are required (Fig. 57). Higher-pressure processes such as plasma-deposition or CVD are clearly preferred over line-of-sight deposition methods such as evaporation or sputtering. A relatively low-temperature (c.f. 250 to 280°C) plasma SiO_2 deposition process has been developed using the Applied Materials Plasma I apparatus.³¹ Dual conductor circuits fabricated using the plasma- SiO_2 insulator exhibit breakdown potentials equivalent to those expected for high-quality, thermally grown SiO_2 films of an equivalent thickness. We thus conclude that the step coverage over the dual conductor features is optimal.

9.1.2 Conductor materials

Though any highly conductive metal alloy system may be used, Al or Au alloys appear quite suitable, and most of the work reported has

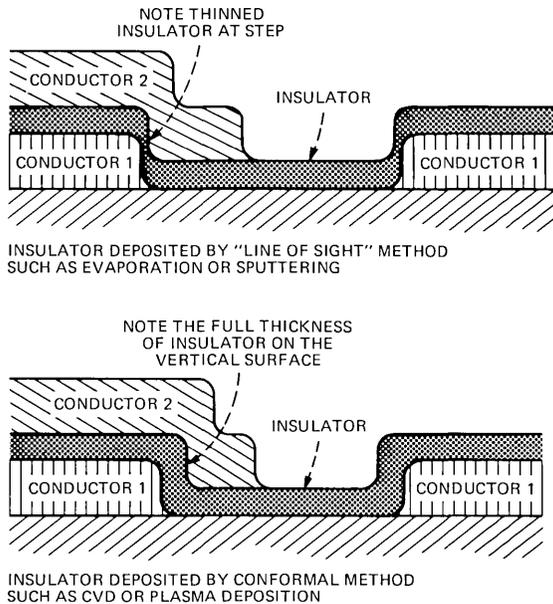


Fig. 57—Effect of conformal vs line-of-sight deposition on insulating film thickness at crossovers.

used Al alloys with low copper content (0.7 to 4 percent). The AlCu alloys are particularly suited because the relatively high current density required for some functions in the dual conductor circuits ($>10^5$ A/cm²) indicates the use of conductor materials with low electromigration rates. The AlCu alloys have been evaporated using a standard e-gun source and planetary wafer fixturing to maximize step coverage. It is advantageous to carry out this deposition at as low a substrate temperature as possible to prevent stress buildup in the AlCu films.³²

9.2 Metal patterning

9.2.1 Photolithography

Standard contact-print photolithography was used on devices reported here.³³ We anticipate that other high-resolution methods such as projection printing, X-ray exposure, etc. may be used equally well.

9.2.2 Metal patterning

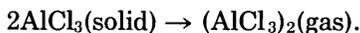
In standard rotating magnetic field-access bubble devices, the Al conductor layer is relatively simple, and at least for 16- μ m period circuits can be etched with standard wet chemical methods. In dual conductor-accessed circuits, however, wet-etch processes are unusable in small period circuits due to the undercutting of the photoresist pattern by the isotropic etch process. This undercutting severely limits the dimensions achievable in conductor patterning. In fact, since these circuits will invariably require that near state-of-the-art resolution be reproduced in the resist mask, no isotropic process or process which biases the conductor line or space dimensions with respect to the photoresist pattern should be used. In the following section, we describe a plasma-etching process that does indeed exactly reproduce the etch mask in the Al film. It is conceivable that other methods may be used, such as the planar anodization method described by D. K. Rose.³⁴

For an etch patterning process, very high material etch selectivity will be required. If no prespacer is used, the first-level etch process cannot be allowed to cut into the epitaxial garnet film and thin it selectively. Similar constraints operate when etching the second conductor, since any thinning of the underlying insulator layer will obviously reduce the dielectric breakdown strength between the two conductors. Plasma etching is an obvious choice to satisfy the requirements for etching Al alloy films for high-density dual conductor circuits for the following reasons: (i) plasma etching uses a chemical process to remove metal and very high etch selectivity can be achieved for different materials, (ii) careful selection and control of plasma process parameters can replicate photoresist patterns in aluminum films without measurable ($<0.1 \mu\text{m}$) dimensional variation. (A good survey of plasma etch processes can be found in Refs. 35 and 36.)

The particular RF (13.56-MHz) plasma process adapted is based on one developed by D. Wang, et al.³⁷ The apparatus used was designed by C. J. Mogab and F. B. Alexander,³⁸ and the etch chamber is shown schematically in Fig. 58. The basic etch process involves the following reaction:



followed by volatilization of the AlCl_3 solid



There is evidence that the anisotropic etch process is promoted by the impact of charged particles on the horizontal surface (Fig. 59).

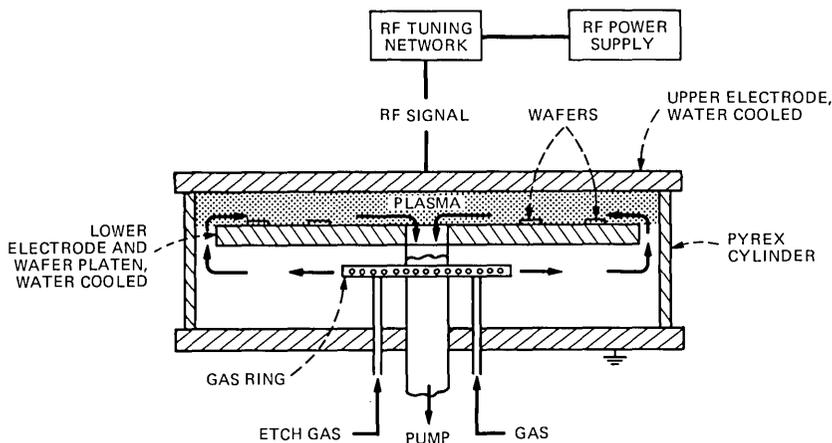


Fig. 58—Radial flow plasma etch apparatus.

CI - CONTAINING PLASMA

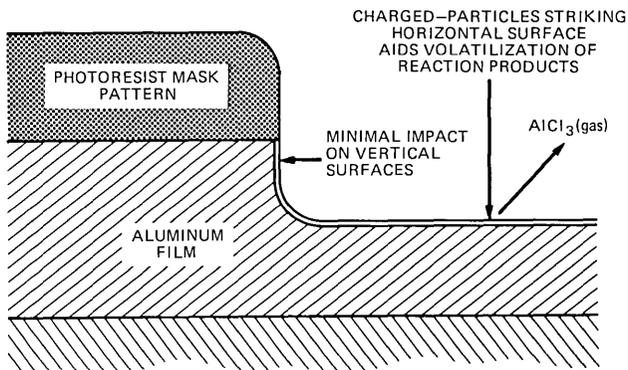


Fig. 59—Schematic of anisotropic etch process.

Obviously, a bias-free etching process will produce near-vertical walls on the metal patterns, which might pose problems in the cross-over features in the dual conductor circuits. To date, the combination of conformal plasma SiO_2 and good step coverage in the Al evaporation process have minimized the potential problems, both with respect to dielectric breakdown and also propagate drive margin reduction, which might be caused by high-resistance crossovers. Figures 60 to 63 are SEM photographs of plasma-etched, dual-conductor patterns. Figure 60 shows a horizontal, straight-line propagate path, while Figure 61 shows several vertical loops. Figure 62 shows one of the generator designs tested. Figure 63 shows, at the same magnification, three data storage loops designed and fabricated within identical linewidth and alignment tolerances. The three loops are of 8-, 6-, and 4- μm period.

X. MAGNETIC FIELDS AND BUBBLE MOTION

Early conductor-drive devices^{39,40} confined currents to discrete conductor lines narrower than a bubble diameter. The present approach to conductor propagation distributes current over a conducting sheet with openings which provide the necessary spatial variations of magnetic field. To furnish some insights and a design aid for conductor-access propagation, two simple geometries—parallel conducting stripes

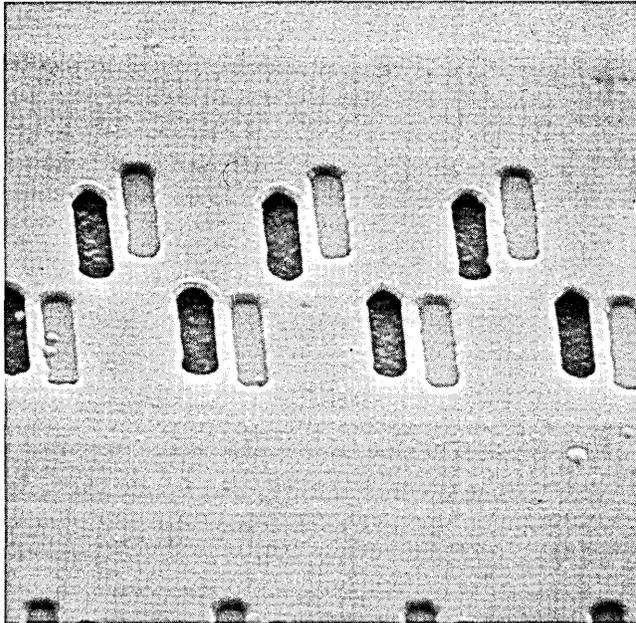


Fig. 60—An SEM of an 8- μm period dual-conductor propagate pattern. The darker apertures are in the first conductor and the lighter in the second conductor pattern. These apertures were etched in the AlCu alloy by a plasma etch technique using a Cl-containing gas (2860X).

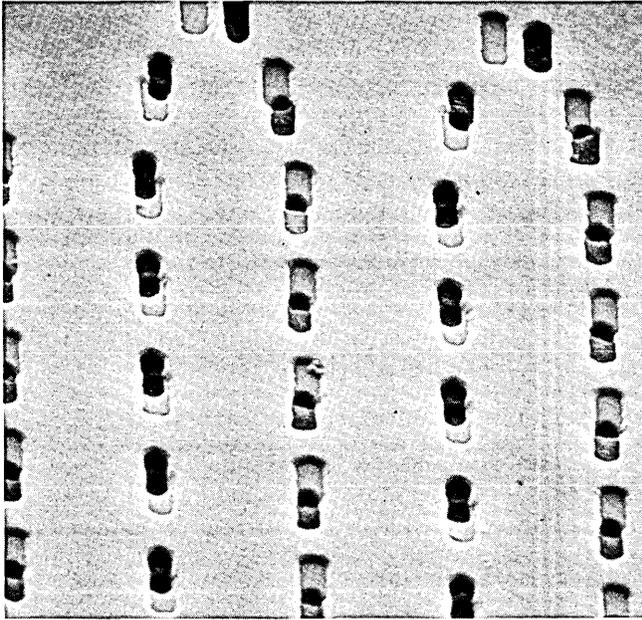


Fig. 61—An SEM of 8- μm memory storage loops (2200X).

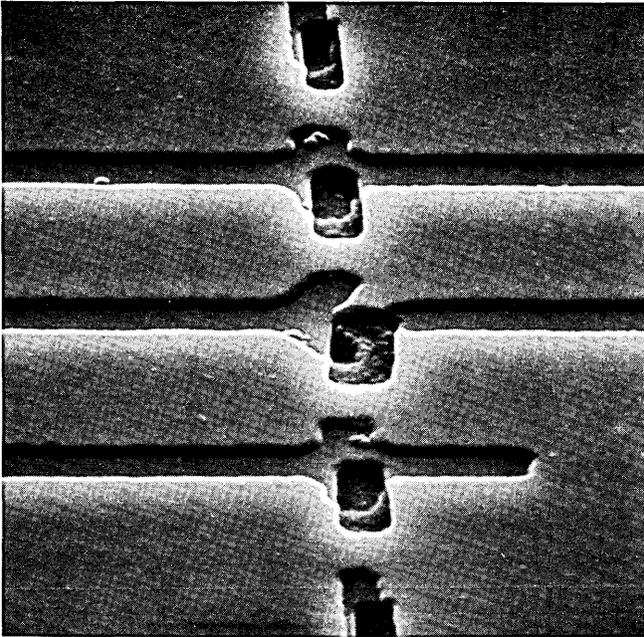


Fig. 62—An SEM of an experimental generator pattern plasma etched in AlCu (3160X).

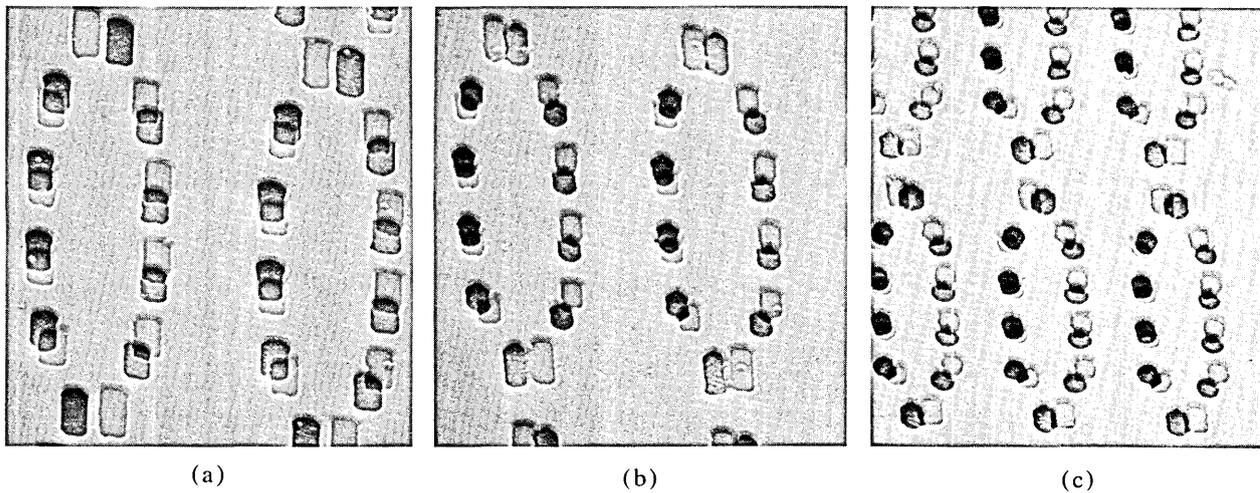


Fig. 63—Data storage loops plasma etched in AlCu films. The three patterns follow identical dimensional design rules (2870X). (a) 8- μ m period; (b) 6- μ m period; (c) 4- μ m period.

and a circular hole in an infinite conducting sheet—were investigated. Bubble motion in a gradient field was also studied.

10.1 Stripe conductors

Consider an infinitesimally thin conducting sheet in the plane $z = 0$. From this sheet, infinite in the y direction, remove material to form conducting stripes of width $2a$ and spacing $s > 2a$; see Fig. 64. Each stripe carries a current $I = -2aJ_o$ in the y direction, where J_o is the constant linear current density of each stripe. In MKS units, the contribution⁴¹

$$H_z(n) = (J_o/2\pi) \ln[r_2(n)/r_1(n)] \quad (12)$$

of the n th stripe to the z component of the field intensity may be expressed in dimensionless form by the quantity

$$\mathcal{H}_s(n) = \frac{4\pi H_z(n)}{J_o} = \ln \left[\frac{(\xi + \alpha - n\sigma)^2 + \zeta^2}{(\xi - \alpha - n\sigma)^2 + \zeta^2} \right], \quad (13)$$

where the lengths $r_1(n)$ and $r_2(n)$ are defined in Fig. 64, $\xi = x/h$, $\zeta = z/h$, $\alpha = a/h$, and $\sigma = s/h$. For bubble propagation, one is interested in the average intensity

$$\bar{\mathcal{H}}_s(n) = h^{-1} \int_{z_s}^{z_s+h} H_z(n) dz, \quad (14)$$

where z_s is the garnet-conductor separation and h the garnet-film thickness. After integrating eq. (14) and summing over the stripes, one finds

$$\bar{\mathcal{H}}_s = \sum_n \sum_{j=1}^2 (-1)^j [\zeta_j \ln A_{nj} + 2\beta_j(n) \text{Arctan } C_{nj}], \quad (15)$$

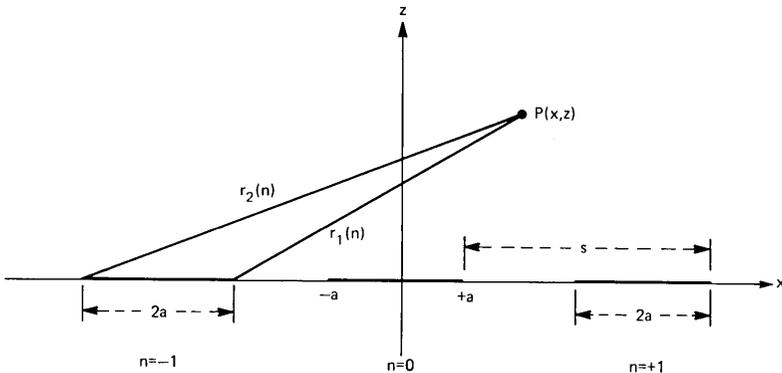


Fig. 64—Coordinate system for an array of infinite conducting stripes. The conductors, $2a$ units wide and spaced s units apart, occupy the plane $z = 0$. Each stripe extends to infinity in the $\pm y$ directions, carries a current I , and bears an index n .

where

$$\begin{aligned}\bar{\mathcal{H}}_s &= 4\pi\bar{H}_z/J_o = \text{dimensionless average intensity,} \\ \bar{H}_z &= \sum_n \bar{H}_z(n) = \text{average intensity,} \\ \zeta_1 &= z_s/h, \quad \zeta_2 = \zeta_1 + 1, \quad \beta_j(n) = \xi + (-1)^j\alpha - n\sigma, \\ A_{nj} &= [\zeta_j^2 + \beta_j^2(n)]/[\zeta_j^2 + \beta_1^2(n)], \\ C_{nj} &= \beta_j(n)/[\beta_j^2(n) + \zeta_1\zeta_2].\end{aligned}$$

When the number of stripes is small, say, less than 100, the best way to evaluate eq. (15) is by programming the right-hand side as it stands. As the number of stripes n_s is increased, this procedure becomes less attractive because $2n_s$ evaluations of the logarithm and arctangent functions are needed for each value of x . Note that, for $n_s < \infty$, $\bar{\mathcal{H}}_s$ is not periodic; for a complete numerical description of eq. (15) in the vicinity of the stripes, the number of function evaluations is $2n_x n_\alpha n_s^2$, where n_x is the number of points per period s and n_α the number of α values of interest.

When n_s is large, say, $n_s > 100$, the behavior of $\bar{\mathcal{H}}_s$ near the array center may be approximated by an infinite number of stripes. This choice permits a significant reduction in computation time but fails to give information about the behavior near the array edges—a matter that will be investigated separately. As $n_s \rightarrow \infty$, the dimensionless average intensity, regarded as a function of ξ , has the properties

$$\bar{\mathcal{H}}_s(\xi) = \bar{\mathcal{H}}_s(\xi + \sigma) \quad \text{and} \quad \bar{\mathcal{H}}_s(-\xi) = -\bar{\mathcal{H}}_s(+\xi). \quad (16)$$

Consequently, one can restrict evaluation of $\bar{\mathcal{H}}_s$ to the points $0 \leq \xi \leq \sigma/2$. Equation (16) implies $\bar{\mathcal{H}}_s(0) = \bar{\mathcal{H}}_s(\sigma/2) = 0$. This result may also be deduced from symmetry considerations. For an origin at the center of an opening between stripes, the same symmetry arguments require $\bar{\mathcal{H}}_s$ to be antisymmetric about this origin. For numerical evaluation of $\bar{\mathcal{H}}_s$ in the limit $n_s \rightarrow \infty$, eq. (15) was rewritten as

$$\bar{\mathcal{H}}_s = \mathcal{H}_1 + \mathcal{H}_2, \quad (17)$$

where

$$\begin{aligned}\mathcal{H}_1 &= \sum_{j=1}^2 (-1)^j \zeta_j \left[\ln A_{oj} + \sum_{n=1}^N [\ln A_{nj} A_{-nj}] \right. \\ &\quad \left. - (8\alpha\xi/\sigma^2) \left[\zeta(2) - \sum_{n=1}^N n^{-2} \right] \right. \\ &\quad \left. - [8\alpha\xi(\xi^2 + \alpha^2 - 3\zeta_j)/\sigma^4] \left[\zeta(4) - \sum_{n=1}^N n^{-4} \right] \right],\end{aligned}$$

$$\mathcal{H}_2 = 2 \sum_{j=1}^2 \sum_{n=-N}^{+N} [(-1)^j \beta_j(n) \text{Arctan } C_{nj}]$$

$$- [16\alpha\xi(\zeta_2^3 - \zeta_1^3)/\sigma^4] \left[\zeta(4) - \sum_{n=1}^N n^{-4} \right],$$

$\zeta(k)$ = Riemann zeta function.⁴²

Equation (17) adds the contributions from the $2N + 1$ stripes nearest the origin as they stand in eq. (15) to the contribution from the remaining stripes approximated by a truncated series in the inverse powers of the stripe index n . Since the leading term in the power series

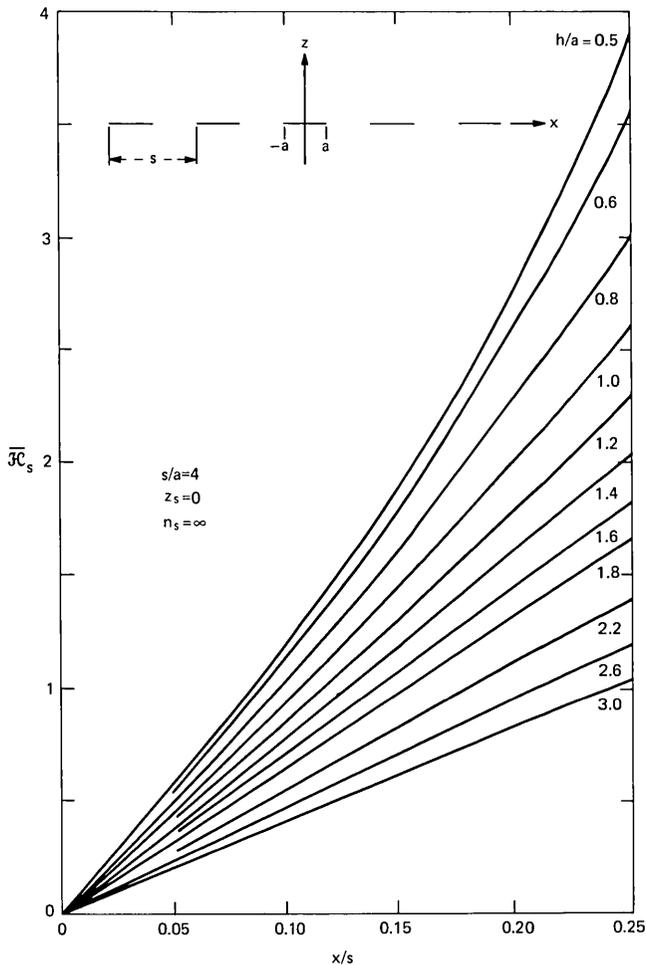


Fig. 65—Dimensionless average value $\overline{\mathcal{H}}_s$ of the z component of magnetic field as a function of x/s . The curves are graphs of eq. (17) for $s = 4a$ and $z_s = 0$. The ordinate $\overline{\mathcal{H}}_s$ is numerically equal to the field in oersteds when the linear current density of each stripe is $1 \text{ mA}/\mu\text{m}$.

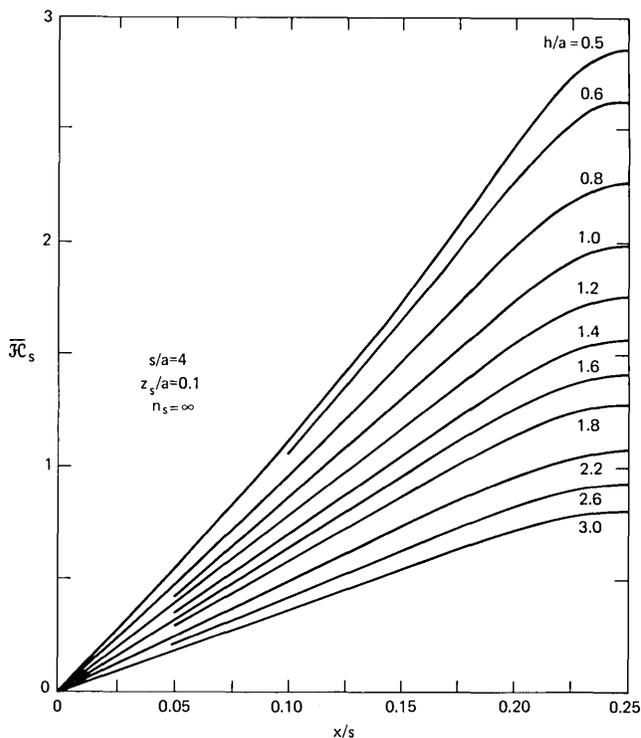


Fig. 66—Same as Fig. 65, except $x_s = 0.1a$.

is n^{-2} , the field at the origin due to a pair of stripes, remote from and symmetric about the origin, decreases as the inverse square of the distance of either stripe from the origin. Accuracy requirements determine N . For the choice $N = 10$, used in the computations presented here, the estimated error in $\overline{\mathcal{H}}_s$ is less than 10^{-5} . Figures 65 to 69 show graphs of $\overline{\mathcal{H}}_s$ vs x/s for parameter values useful to the device designer. Curves of $\overline{\mathcal{H}}_s$ for $z_s = 0$ are unique in that each has a cusp at $x/s = 0.25$, where the slope $\partial\overline{\mathcal{H}}_s/\partial x$ is not continuous. The choice $s = 4a$ requires calculation only for $0 < \xi \leq \sigma/4$ because $\overline{\mathcal{H}}_s$ is symmetric about $\xi = \pm\sigma/4$. The origin of this symmetry may be understood by superimposing another current sheet with current density $\mathbf{J} = (J_o/2)\hat{j}$ on the existing stripe pattern of Fig. 64. This addition leaves the z component of field unaltered. It creates, however, a system of adjacent conducting stripes of width $2a$, carrying the current $I = J_o a$ alternately in the positive and negative y direction. Such an array of stripes enjoys symmetry about the points of discontinuity. Contributions to the z component of field from any pair of stripes symmetrically disposed about a selected discontinuity will add; $\overline{\mathcal{H}}_s$ will, therefore, have extrema at points of current-density discontinuity. To illustrate how the average field intensity decreases with increasing distance from the conduc-

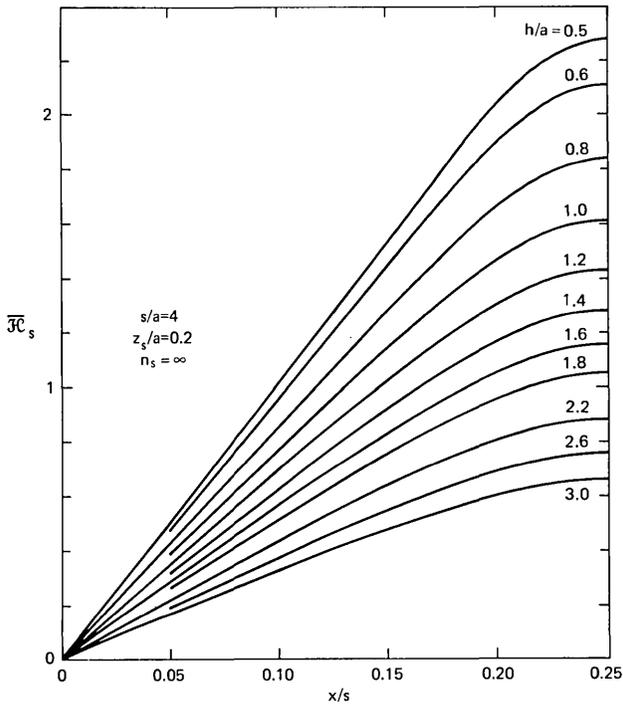


Fig. 67—Same as Fig. 65, except $z_s = 0.2a$.

tor array, values of \bar{H}_s were computed for several stand-off distances z_s . Since \bar{H}_s assumes its maximum value at $x/s = 0.25$ for $s/a = 4$, the stripe edge was chosen for the calculations. See Fig. 70.

To convert to cgs units, write $\bar{H}_z = \bar{H}_s J_o$ and express J_o in A/mm or mA/ μ m. In other words, the dimensionless quantity \bar{H}_s gives the average field intensity in oersteds when the linear current density J_o is 1 mA/ μ m.

The gradient of the average field $\partial \bar{H}_z / \partial x$ can be obtained with only a slight increase in computing effort because the needed function values are available. The dimensionless gradient is

$$\bar{G} = \left(\frac{4\pi h}{J_o} \right) \frac{\partial \bar{H}_z}{\partial x} = 2 \sum_n \sum_{j=1}^2 (-1)^j \text{Arctan } C_{nj}. \quad (18)$$

When the number of stripes n_s is large, one may use the approximation

$$\begin{aligned} \bar{G} = 2 \sum_{-N}^{+N} \sum_{j=1}^2 [(-1)^j \text{Arctan } C_{nj}] - (8\alpha/\sigma^2) \left[\zeta(2) - \sum_1^N n^{-2} \right] \\ - (8\alpha/\sigma^4) [3\zeta^2 + \alpha^2 - \zeta_2^2 - \zeta_2 \zeta_1 - \zeta_1^2] \left[\zeta(4) - \sum_1^N n^{-4} \right], \quad (19) \end{aligned}$$

an expression correct to the 4th order in the quantity $1/n$. Equation

(19) adds the contribution from the $2N + 1$ stripes nearest the origin to the contribution, approximated by a truncated series in powers of $1/n$, of the infinitely many remaining stripes. Accuracy requirements determine the value of N .

10.2 Infinite conducting sheet with circular hole

Consider again an infinitely thin conducting sheet in the plane $z = 0$. Let the sheet contain a circular hole, extend to infinity, and support

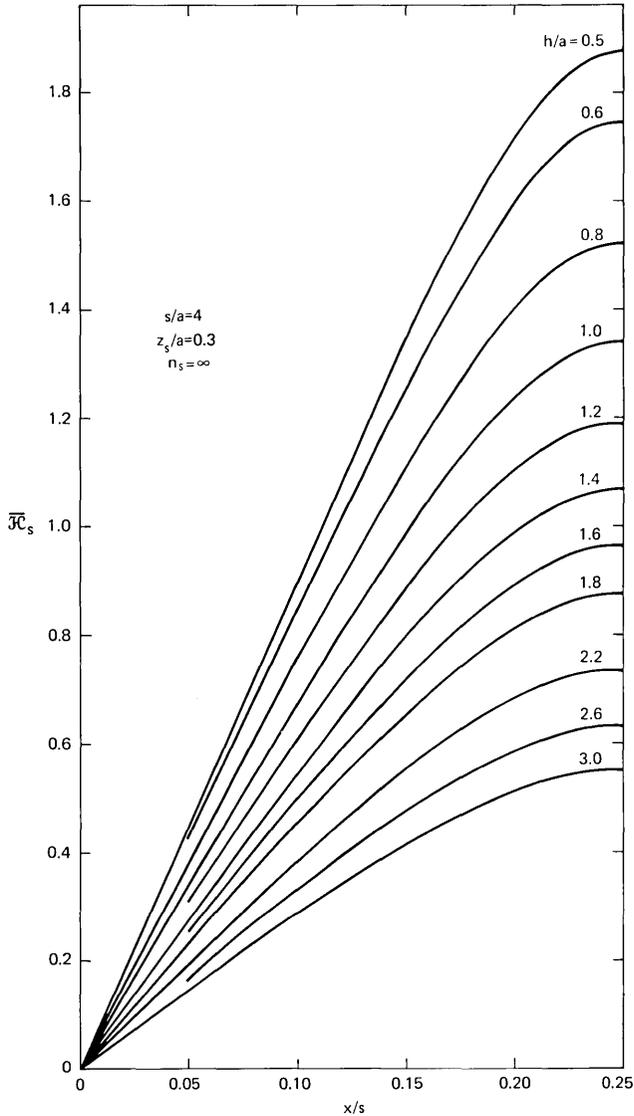


Fig. 68—Same as Fig. 65, except $z_s = 0.3a$.

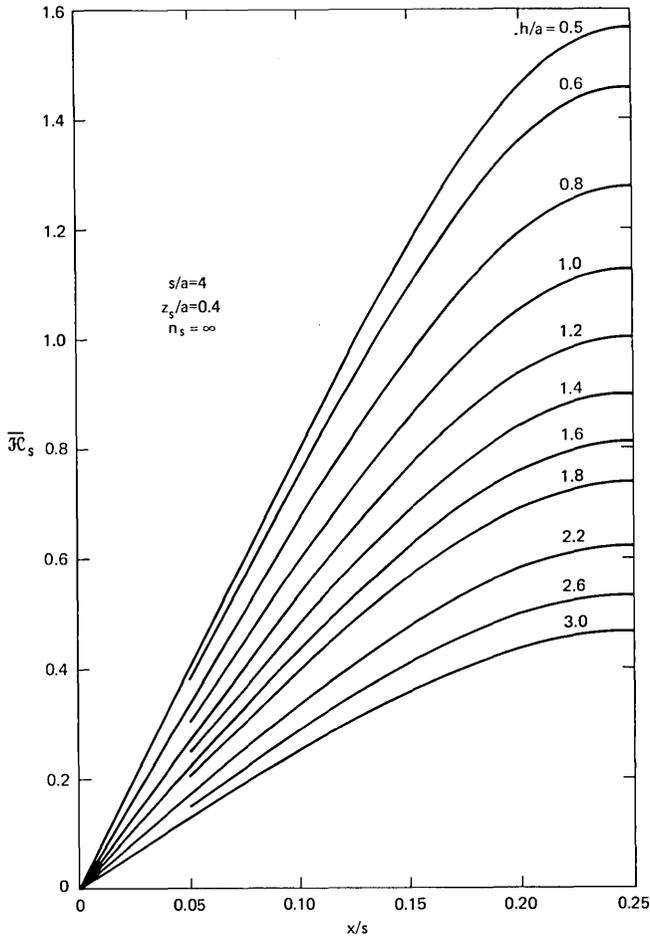


Fig. 69—Same as Fig. 65, except $z_s = 0.4a$.

the constant current density $\mathbf{J} = J_o \hat{i}$ at points infinitely removed from the hole. The origin of coordinates is chosen at the symmetry point of the geometry as shown in Fig. 71. The problem to be solved—a means of obtaining the z component of the magnetic field at the point $P(r, \theta)$ —falls into two parts: (i) solution of a boundary-value problem to yield the linear current density \mathbf{J} in the plane $z = 0$, and (ii) integration over the source distribution to obtain the field.

The first of these is discussed by Milne-Thompson⁴³ in connection with two-dimensional fluid flow around a cylindrical obstacle. The current density is

$$\mathbf{J} = J_o [1 - (a/\rho)^2 \cos 2\phi] \hat{i} - J_o (a/\rho)^2 \sin 2\phi \hat{j}, \quad (20)$$

where

J_o = constant linear current density at infinity,

a = hole radius,

ρ = radial source coordinate,

ϕ = angular source coordinate.

The reader may quickly verify that eq. (20) satisfies the requisite boundary conditions. The current density is tangential at $\rho = a$ and $\mathbf{J} \rightarrow J_o \hat{i}$ as $\rho \rightarrow \infty$. One can now obtain the magnetic field in MKS units from

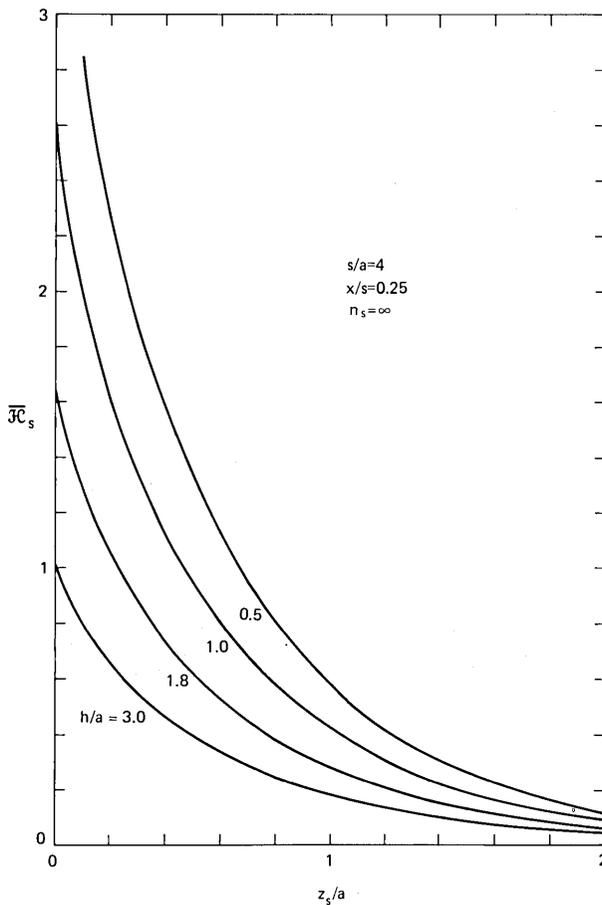


Fig. 70—Dimensionless average value $\bar{\mathcal{H}}_z$ of the z component of magnetic field as a function normalized stand-off distance z_s/a . The curves are graphs of eq. (17) for $s = 4a$ and $x = 0.25s$. The ordinate $\bar{\mathcal{H}}_z$ is numerically equal to the field in oersteds when the linear current density of each stripe is $1 \text{ mA}/\mu\text{m}$.

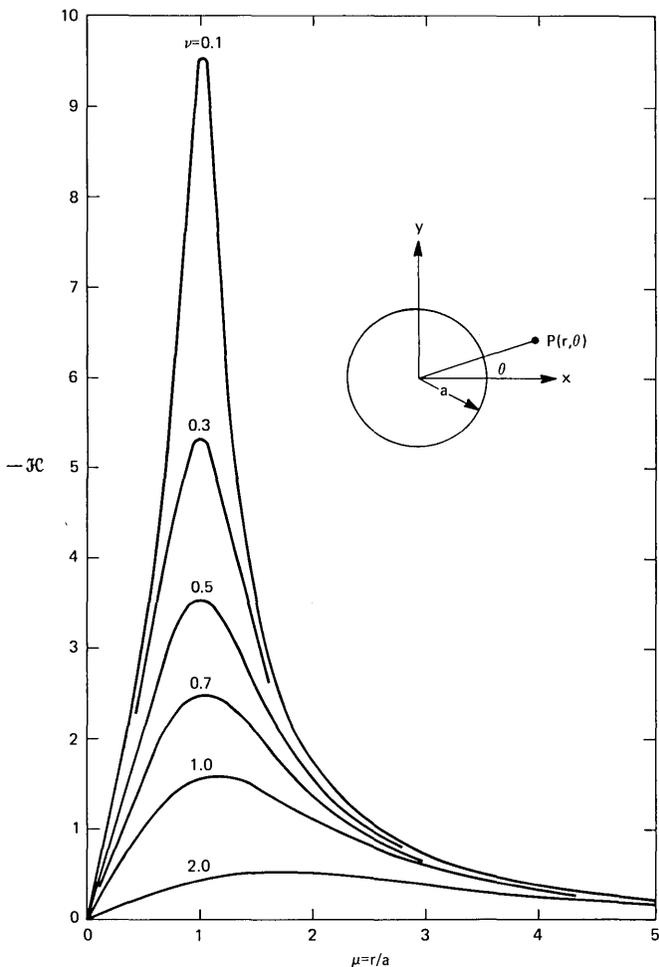


Fig. 71—Current sheet with circular hole. The dimensionless z component of the magnetic field is shown as a function of the reduced distance $\mu = r/a$ for several values of $\nu = z/a$. The curves are graphs of \mathcal{H} defined by the integral below eq. (23). The ordinate \mathcal{H} is numerically equal to the field in oersteds when $\theta = 90^\circ$ and the linear current density approaches 1 mA/ μ m as $r \rightarrow \infty$.

$$4\pi\mathbf{H} = \int_s \frac{\mathbf{J} \times \mathbf{R}}{R^3} d\xi d\eta, \quad (21)$$

where

$$\mathbf{R} = (x - \xi)\mathbf{i} + (y - \eta)\mathbf{j} + z\mathbf{k} \quad \text{and} \quad R = |\mathbf{R}|.$$

The surface integral, with respect to the source coordinates ξ and η , extends over the entire $\xi - \eta$ plane apart from the points $\rho = (\xi^2 + \eta^2)^{1/2} < a$, where the current density vanishes. After projecting

eq. (21) on the z -axis and expressing all components in cylindrical coordinates, one obtains a double integral of the form

$$4\pi H_z = J_o \int_0^{2\pi} d\phi \int_a^\infty \Phi(r, \theta, z, \rho, \phi) d\rho. \quad (22)$$

All radial integrals are tabulated; see, for example, Gradshteyn and Ryzhik.⁴⁴ The remaining angular integration, however, must be done numerically. After appropriate changes of the variables and collecting terms, eq. (22) reduces to

$$4\pi H_z = J_o \mathcal{H} \sin \theta, \quad (23)$$

where

$$\mathcal{H} = 2 \int_0^\pi (F_1 + F_2 + F_3) d\psi = \mathcal{H}(\mu, \nu),$$

$$F_1 = \frac{C_0 + C_1 C + C_2 C^2 + C_3 C^3 + C_4 C^4}{\sigma^2 (\sigma^2 - \mu^2 C^2) (\sigma^2 + 1 - 2\mu C)^{1/2}},$$

$$F_2 = C \ln[(\sigma^2 + 1 - 2\mu C)^{1/2} + 1 - \mu C],$$

$$F_3 = \mu(2C^2 - 1)\sigma^{-3} \ln \left[\frac{\sigma(\sigma^2 + 1 - 2\mu C)^{1/2} + \sigma^2 - \mu C}{\sigma - \mu C} \right],$$

$$\mu = r/a, \quad \nu = z/a,$$

$$\sigma = (\mu^2 + \nu^2)^{1/2}, \quad C = \cos \psi,$$

$$C_0 = \mu\sigma^2(1 + \sigma^2),$$

$$C_1 = (\sigma^2 + \mu^2)(1 - \sigma^2),$$

$$C_2 = -\mu[\sigma^2(\sigma^2 + 3) + 2\mu^2],$$

$$C_3 = 2\mu^2(\sigma^2 - 1),$$

$$C_4 = 4\mu^3.$$

Numerical integration over the angular variable of eq. (22) employed the automatic numerical quadrature routine QUAD.⁴⁵ The dependence of \mathcal{H} on the radius and altitude is shown in Fig. 71. Using this figure and eq. (23), one may now estimate H_z in MKS units. In the hybrid units described earlier, the z component of magnetic field in oersteds is numerically equal to \mathcal{H} when $J_o = 1$ mA/ μ m and $\theta = \pi/2$.

To obtain information about the average value,

$$\bar{H}_z = \frac{1}{h} \int_{z_s}^{z_s+h} H_z dz = \frac{J_o \sin \theta}{4\pi h} \int_{z_s}^{z_s+h} \mathcal{H}(\mu, \nu) dz, \quad (24)$$

the dimensionless quantity

$$\bar{\mathcal{H}} = \bar{\mathcal{H}}(\mu, \gamma) = h^{-1} \int_{z_s}^{z_s+h} \mathcal{H} dz = \gamma^{-1} \int_{\nu_s}^{\nu_s+\gamma} \mathcal{H}(\mu, \nu) d\nu \quad (25)$$

was computed. Here $\nu_s = z_s/a$ and $\gamma = h/a$. Since several γ values are of interest, one may take advantage of the linear property of integration to reduce the number of function evaluations. Index the γ values of interest with j such that $\gamma_{j+1} > \gamma_j$. To obtain a uniform formula and accommodate Fortran requirements, one must choose $\gamma_1 = 0$ and the smallest value of interest γ_2 . Then, with the notation $\bar{\mathcal{H}}_j = \bar{\mathcal{H}}(\mu, \gamma_{j+1})$, one finds

$$\bar{\mathcal{H}}_j = \frac{1}{\gamma_{j+1}} \sum_{i=1}^j g_i, \quad (26)$$

where

$$g_i = \int_{\nu_s+\gamma_i}^{\nu_s+\gamma_{i+1}} \mathcal{H}(\mu, \nu) d\nu.$$

Each g_i was evaluated with a 3-point Gaussian rule; for the γ_j used in the computation the estimated error is of order 10^{-4} . The radial slope $\bar{\mathcal{P}} = \partial\bar{\mathcal{H}}/\partial\mu$ was also computed using the approximation

$$\bar{\mathcal{P}}(\mu_k, \gamma_j) \approx [\bar{\mathcal{H}}(\mu_{k+1}, \gamma_j) - \bar{\mathcal{H}}(\mu_{k-1}, \gamma_j)]/2\Delta u, \quad (27)$$

where Δu is the constant difference between adjacent values of μ . The dependence of $-\bar{\mathcal{H}}$ and $-\bar{\mathcal{P}}$ on r/a is shown in Figs. 72 and 73, respectively. Contour maps of constant \bar{H}_z demonstrate the overall behavior of the field. Figure 7 shows contours of $\bar{H}_z = \text{constant}$ for the reduced stand-off distances $z_s/a = 0.06$ and 0.3 . For both, $h/a = 0.8$.

10.3 Linear chain of oval holes

Current experimental devices use quasi-elliptic openings rather than circular holes; see, for example, Fig. 29. The work of Richmond⁴⁶ with electric fields about a linear grating permits more realistic magnetic field calculations. Consider an infinite chain of identical oval holes on the y -axis, spaced $2b$ units apart, each with semiaxes $a_2 > a_1$ parallel to the y and x axes, respectively. See Fig. 74 for the oval at the origin. Because of symmetry and the periodicity $2b$ in the y direction, Laplace's equation need be solved only in the region of Fig. 74. Richmond⁴⁶ gives the conformal transformation

$$z = (a_1/c)w + (a_2/\gamma)\cosh^{-1}(\cos \gamma \cosh w) \quad (28)$$

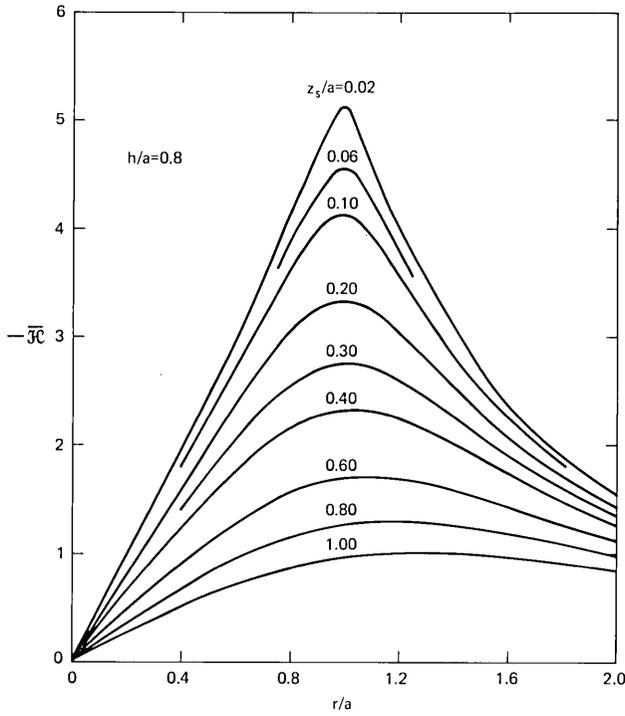


Fig. 72—Dimensionless average value \bar{H} of z component of magnetic field as a function of r/a for an infinite current sheet with a circular hole. The curves are graphs of eq. (25) for $h = 0.8a$, where a is the hole radius. The ordinate \bar{H} is numerically equal to the field in oersteds when $\theta = 90^\circ$ and the linear current density approaches 1 mA/ μm as $r \rightarrow \infty$.

from which one finds the complex linear current density

$$J = \frac{2bJ_o}{\pi} \left[\frac{dw}{dz} \right]^* = \left[\frac{\alpha_2 \gamma c J_o (\cos^2 \gamma \cosh^2 w - 1)^{1/2}}{\alpha_1 \gamma (\cos^2 \gamma \cosh^2 w - 1)^{1/2} + c \cos \gamma \sinh w} \right]^* \quad (29)$$

outside the oval. Here J_o , a real constant, is the linear current density as $x \rightarrow \infty$, $\alpha_1 = a_1/a_2$, $\alpha_2 = 2b/\pi a_2$, $z = x + iy$, $w = u + iv$, and * means complex conjugate. Inside the oval, $J = 0$. The real parameters c and γ satisfy the nonlinear equations $\cos \gamma \cosh c = 1$ and $c(\gamma \alpha_2 - 1) = \alpha_1 \gamma$. Elimination of c from the equations leads to an efficient algorithm, based on Newton's method, for machine computation of γ for practical values of $\alpha_1 \leq 1$ and α_2 . The iteration scheme

$$\gamma_{i+1} = \gamma_i - \frac{f_i}{f'_i} = \gamma_i + \frac{\gamma_i g_i [\alpha_1 \gamma_i + g_i (1 - \alpha_2 \gamma_i)] \cos \gamma_i}{\alpha_1 \gamma_i^2 + g_i^2 \cos \gamma_i}, \quad (30)$$

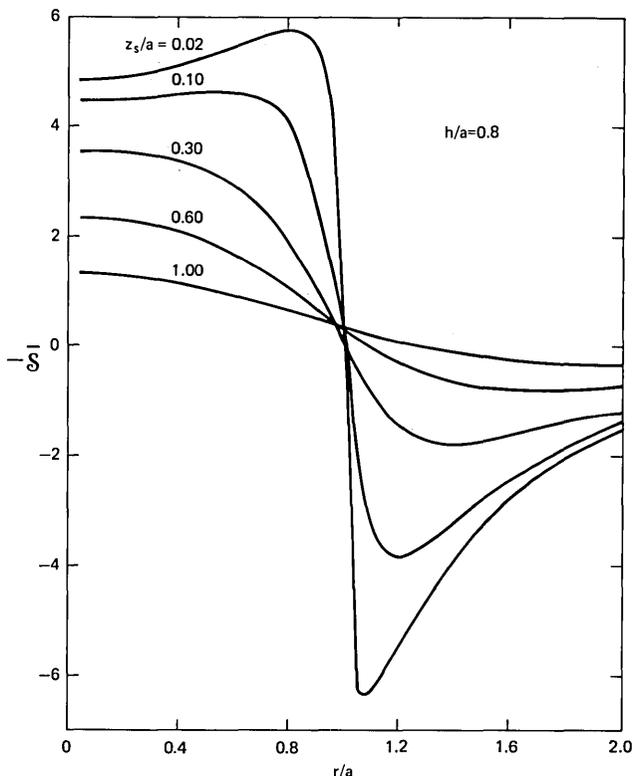


Fig. 73—Dimensionless slope $\bar{\mathcal{S}} = \partial\bar{\mathcal{H}}/\partial\mu$ of the z component of magnetic field as a function of r/a for an infinite current sheet with a circular hole. The curves are graphs of eq. (27) for $h = 0.8a$, where a is the hole radius.

where

$$f = \alpha_1 g^{-1} + \gamma^{-1} - \alpha_2,$$

$$g = \ln[(1 + \sin \gamma)/\cos \gamma],$$

$$f' = df/d\gamma,$$

$$f_i = f(\gamma_i),$$

$$f'_i = df/d\gamma \text{ evaluated at } \gamma = \gamma_i,$$

may be started with $\gamma_o = 1/\alpha_2$. If the iteration process ends with γ_j , the parameter c follows from $c = g_j = \ln[(1 + \sin \gamma_j)/\cos \gamma_j]$.

By setting u or v constant in eq. (28), one obtains equipotential or flow lines, respectively. Several flow lines, indexed with the angle variable $\Psi = 180v/\pi$, are shown in Fig. 74. The oval satisfies the parametric equations

$$x/a_2 = (\alpha_1/c) \cosh^{-1}(q/\cos \gamma) \tag{31a}$$

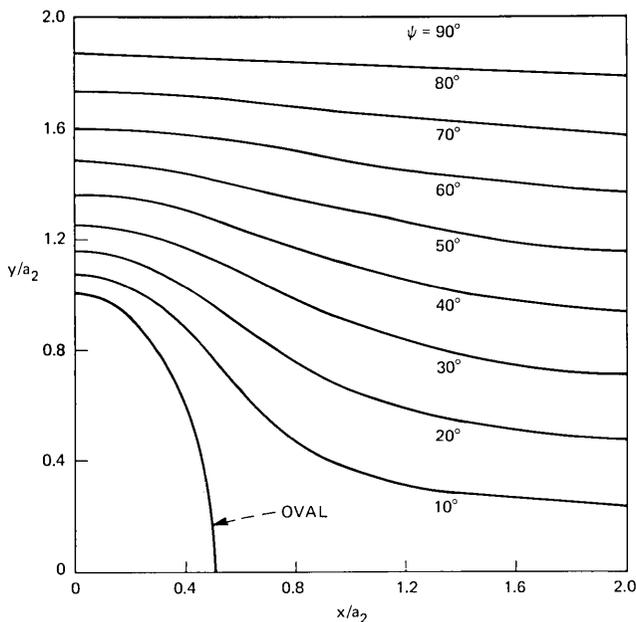


Fig. 74—Current flow near an infinite chain of oval holes. Identical oval holes with semiaxes $a_2 = 2a_1$ and spacing $2b = 4a_2$ occupy the y -axis. The flow lines, indexed with $\psi = 180v/\pi$, were obtained from eq. (28) with $v = \text{constant}$. Note the almost elliptical shape of the oval for the parameter ratios $a_1/a_2 = 0.5$ and $b/a_2 = 2$ used here.

and

$$y/a_2 = (1/\gamma)\cos^{-1}q, \quad (31b)$$

where $0 \leq \cos \gamma \leq q \leq 1$. Richmond⁴⁶ discusses the departure from ellipticity. For geometries of practical interest, say, $a_1/a_2 \approx 0.5$ and $b/a_2 \approx 2$, the oval may be regarded as an ellipse. The distribution of current along the line of minimum separation between ovals satisfies eq. (29) with $u = 0$. It was computed from

$$\frac{J}{J_0} = \frac{\alpha_2 \gamma c (1 - \cos^2 \gamma \cos^2 v)^{1/2}}{\alpha_1 \gamma (1 - \cos^2 \gamma \cos^2 v)^{1/2} + c \cos \gamma \sin v}, \quad 0 \leq v \leq \pi/2 \quad (32)$$

and is shown in Fig. 75. The coordinate value associated with v follows from eq. (28).

To complete the field computations, an efficient method for the numerical double integration of the components of eq. (21) must be constructed. This work is in progress.

10.4 Bubble motion

Experimental evidence, Fig. 34, for example, strongly suggests bubbles with different hardness parameters. A study of bubble motion will

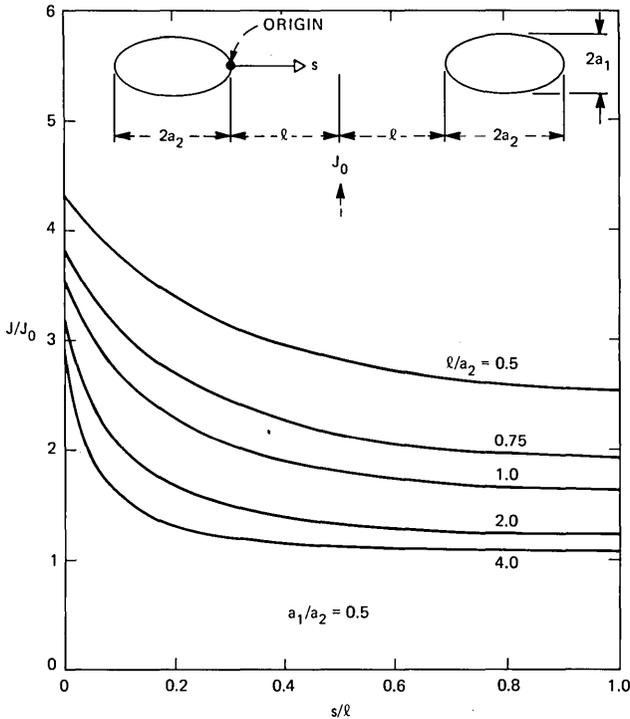


Fig. 75—Distribution of linear current density along the line of minimum separation between ovals of an infinite chain. The current density is normalized to J_0 , its value infinitely far from the ovals. The graphs were obtained from eq. (32).

aid the device designer by prediction of trajectories and transit times.

With the assumption that the bubble speed,⁴⁷

$$V = \mu_b [d_b |\nabla \bar{H}_z| - 8H_c/\pi] \geq 0, \quad (33)$$

is independent of the hardness or S number, the equations of motion in rectangular coordinates are

$$\dot{x}_i = -VT_{ij}e_j, \quad (34)$$

where V is a scalar point function, μ_b is the bubble mobility equal to one-half the wall mobility, d_b is the bubble diameter, \bar{H}_z is the z component of field averaged over the bubble height, H_c is the coercivity field, $T_{11} = T_{22} = \cos \phi$, $T_{12} = -T_{21} = \sin \phi$, $\hat{e} = \nabla \bar{H}_z / |\nabla \bar{H}_z|$, e_j is the projection of \hat{e} on the j th coordinate axis, and ϕ is the angle between $-\hat{e}$ and the local velocity \mathbf{V} . The angle ϕ depends on S according to $\sin \phi = 8SV/\gamma d_b^2 |\nabla \bar{H}_z|$ and is regarded as a fixed parameter in eq. (34). Here γ is the gyromagnetic ratio of the material.⁴⁸ See also Fig. 76. When applied to bubble motion due to fields from a hole of radius a in a current sheet, eqs. (33) and (34) assume the form

$$\dot{\xi}_1 = -(F_c/\mu^3)[\xi_1\xi_2(\mu\dot{\mathcal{P}} - \dot{\mathcal{H}})\cos\phi + (\mu\xi_2^2\dot{\mathcal{P}} + \xi_1^2\dot{\mathcal{H}})\sin\phi], \quad (35a)$$

$$\dot{\xi}_2 = +(F_c/\mu^3)[\xi_1\xi_2(\mu\dot{\mathcal{P}} - \dot{\mathcal{H}})\sin\phi - (\mu\xi_2^2\dot{\mathcal{P}} + \xi_1^2\dot{\mathcal{H}})\cos\phi], \quad (35b)$$

where

$$F_c = 1 - 8H_c/\pi d_b |\nabla\bar{H}_z|,$$

$$|\nabla\bar{H}_z| = (J_o/a)[\dot{\mathcal{P}}^2\sin^2\theta + (\dot{\mathcal{H}}/\mu)^2\cos^2\theta]^{1/2},$$

$$\sin\theta = \xi_2/\mu,$$

$$\cos\theta = \xi_1/\mu,$$

$$\mu = (\xi_1^2 + \xi_2^2)^{1/2},$$

$$\xi_1 = x/a,$$

$$\xi_2 = y/a,$$

$$\dot{\mathcal{P}} = \partial\dot{\mathcal{H}}/\partial\mu.$$

Hybrid units were used for eq. (35); this allows substitution for H_c in oersteds. The current density J_o , however, must be in units of mA/ μ m. The dot over the variables on the left-hand side of eq. (35) means differentiation with respect to the normalized time $T = t/\tau$, where $\tau = a^2/\mu_b d_b J_o$. Solutions to these coupled, nonlinear differential equations must, in general, be obtained by numerical techniques. Some special cases, amenable to an analytical approach, are now described.

The equations of motion, eq. (34), assume a simple form when written in polar coordinates:

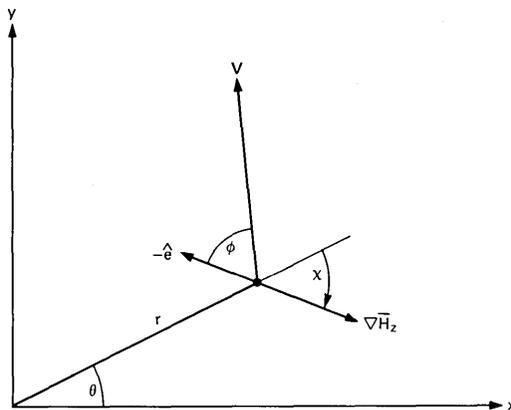


Fig. 76—Coordinate system and symbols used in the equations of bubble motion. Note that the angle χ , as shown here, is a negative quantity because the projection of $\nabla\bar{H}_z$ on the unit vector $\hat{\theta}$ of the polar coordinates (r, θ) is negative.

$$\dot{r} = dr/dt = -V\cos(\phi - \chi), \quad (36a)$$

$$\dot{\theta} = d\theta/dt = +(V/r)\sin(\phi - \chi), \quad (36b)$$

where χ is the angle between $\nabla\bar{H}_z$ and \hat{r} a unit vector in the radial direction. Division of eq. (36b) by (36a) yields the velocity ratio

$$r\dot{\theta}/\dot{r} = v_\theta/v_r = \tan(\chi - \phi). \quad (37)$$

Consider now the special case $\chi = 0$. This implies $\partial\bar{H}_z/\partial\theta$ vanishes. Since ϕ is a constant, eq. (37) may be integrated to furnish the trajectory

$$\theta = \theta_o + \tan \phi \ln(r_o/r), \quad (38)$$

where r_o and θ_o specify the position at some reference time, $t = t_o$. The path is a logarithmic spiral; it is unique in that it is independent of the radial behavior of \bar{H}_z . The dependence of \bar{H}_z on r will, however, determine the position and velocity along the spiral as a function of the time by inversion of the quadrature

$$t = t_o + \sec \phi \int_r^{r_o} \frac{ds}{V(s)}. \quad (39)$$

Equation (39) follows from eq. (36a). When the functional dependence of \bar{H}_z on r is sufficiently simple, eq. (39) may be integrated and inverted. For example, when $\bar{H}_z = H_o(r/d_b)^2$, $V = v(r) = (r - \alpha d_b)/\tau_o$. Here H_o is a constant, $\alpha = 4H_c/\pi H_o$, $\tau_o = d_b/\mu_w H_o$, and $\mu_w = 2\mu_b$ is the wall mobility. Integration of eq. (39) furnishes

$$r = \alpha d_b + (r_o - \alpha d_b)\exp(-\lambda t), \quad (40)$$

where $\lambda = \tau_o^{-1} \cos \phi$ and $t_o = 0$. Figure 77 shows several trajectories. Bubble motion starts at a point $r = r_o$ marked $t/\tau_o = 0$. From there, bubbles travel along logarithmic spiral paths terminating at the origin. The numerals along the various paths give bubble position in units of t/τ_o for $\alpha d_b/r_o = 0.5$. When $H_c = 0$ and $0 \leq \phi < 90^\circ$, all bubbles spiral into the origin arriving there as $t \rightarrow \infty$.

Two special cases of bubble motion near a circular opening allow integration of eq. (35) by quadrature. For both, $S = \phi = 0$. Consider a bubble located on the y axis at $t = 0$. Using polar coordinates, one finds immediately $\theta = \text{constant}$ since $\phi = 0$ and $\chi = 0$ or π ; see eq. (36). The radial equation is solved by the quadrature

$$t = \tau \int_0^{r/a} \frac{d\sigma}{|\mathcal{P}(\sigma)| - \beta} \quad (41)$$

for motion starting at the origin. Here, $\tau = a^2/\mu_b d_b J_o$ and $\beta = 8H_c a/\pi d_b J_o$ with H_c in oersteds and J_o in mA/ μm . Numerical evalua-

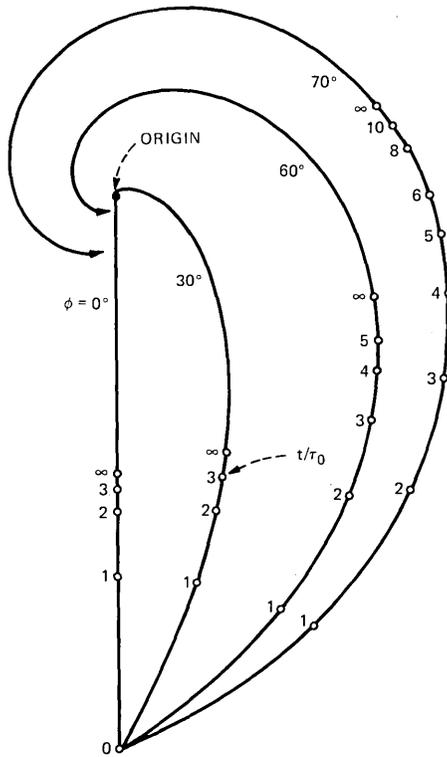


Fig. 77—Logarithmic spiral trajectories described by eq. (38). Bubbles spiral into the origin along these paths when $H_c = 0$. For $H_c > 0$, motion is along the same paths but bubbles do not reach the origin. The timing markers, in units of t/τ_0 , show positions for $ad_b/r_0 = 0.5$.

tion of eq. (41) for $a = d_b = 2\mu\text{m}$, $H_c = 1\text{ Oe}$, $\mu_b = 1\mu\text{m}/\mu\text{s-Oe}$, $J_0 = 1\text{ mA}/\mu\text{m}$, $z_s/a = 0.06$, and $h/a = 0.8$ shows that $0.70\mu\text{s}$ are required for motion from $r = 0$ to $r = 0.8a$. For these parameters, the characteristic time τ equals $2\mu\text{s}$. Arrival at $r/a = 0.8$ completes approximately 90 percent of the possible radial motion.

The second special case occurs when a bubble resides at the radial minimum of \mathcal{H} at $t = 0$. Enter eq. (36) with $\phi = 0$ and $\chi = -\pi/2$. Thus, $r = r_e = \text{constant}$, and the equilibrium value r_e is found from the solution of $\mathcal{P}(\mu) = \partial\mathcal{H}/\partial\mu = 0$ with $\mu_e = r_e/a$. As may be seen from Fig. 72, $r_e/a \approx 1$ for small z_s/a . Equation (36b),

$$\dot{\theta} = V/r_e = \tau_1(\cos\theta - \beta_1), \quad (42)$$

where $\tau_1 = a^2\mu_e^2/\mu_b d_b J_0 |\mathcal{H}_e|$, $\beta_1 = 8H_c a \mu_e / \pi d_b J_0 |\mathcal{H}_e|$, and $\mathcal{H}_e = \mathcal{H}(\mu_e)$ is the value of \mathcal{H} at the equilibrium radius, has the solution

$$t = \tau_1 \csc\theta_m \ln \left[\frac{\tan(\theta_m/2) + \tan(\theta/2)}{\tan(\theta_m/2) - \tan(\theta/2)} \right] \quad (43)$$

for a bubble at $\theta = 0$ when $t = 0$. The angle θ_m is attained as $t \rightarrow \infty$ and satisfies the condition $\cos \theta_m = \beta_1$. For the numerical values used in the previous example together with $\mu_e = 1.00$ and $|\mathcal{H}_e| = 4.56$, one finds $\theta_m = 56.0^\circ$ and the time needed for travel from $\theta = 0$ to $\theta = 0.9\theta_m$ is $1.48 \mu s$.

10.5 Discussion

The graphs of Figs. 65 to 69 apply to an infinite number of conducting stripes. As shown in these figures, $\mathcal{H}_s = 0$ at the origin, a conductor center, and rises to a maximum at the right conductor edge. The symmetry of \mathcal{H}_s about a conductor edge for $s = 4a$ together with the periodicity and antisymmetry expressed by eq. (16) completes the description of \mathcal{H}_s for one period. See Fig. 8 for a typical spatial variation of \mathcal{H}_s over several periods.

When the number of conducting stripes is finite, \mathcal{H}_s still has the wavelike behavior per period as described above. There are, however, essential differences. The function \mathcal{H}_s is not periodic. In addition, the wavelike behavior of \mathcal{H}_s is superimposed on a gradual change in the mean value of the bias field. This change in bias is due to the current $I = 2an_s J_o$ flowing in the n_s stripes. The magnitude of the effect can be estimated, using eq. (15), for a single stripe of width $(n_s - 1)s + 2a$. To illustrate the quality of the estimate, consider $n_s = 11$ and $s = 4a$. Choose the origin at the middle of the center stripe and calculate \mathcal{H}_s at the center of each stripe and at the corresponding abscissa for a single stripe of width $(n_s - 1)s + 2a = 42a$. Then, for $\alpha = 1$, the numerical results are shown in Table VI. Here, \mathcal{H}_1 is the normalized intensity for the single stripe. In the design of current-access devices, the shift in local bias must be taken into account. By returning the drive current through a nearby stripe of comparable width, this shift in local bias can be reduced.

To develop current-access propagation structures, large-scale models were made and the magnetic fields due to current flow in them probed with a coil. See Section III. One structure was a stripe pattern which served to calibrate the coil in the following way. The coil of radius r_o and height h was placed over a central stripe with the coil center

Table VI

Stripe Index n	Abscissa x/h	$ \bar{\mathcal{H}}_s $	$ \bar{\mathcal{H}}_1 $
0	0	0	0
± 1	± 4.0	0.367	0.385
± 2	± 8.0	0.760	0.802
± 3	± 12.0	1.218	1.293
± 4	± 16.0	1.829	1.995
± 5	± 20.0	2.928	3.582

directly over the stripe edge. This location permitted maximum flux linkage. One can now calculate the flux threading the coil from

$$\phi = \int B \, dA = 4J_o \int_0^{r_o} \mathcal{H}_s(u) \, du \int_0^{(r_o^2 - u^2)^{1/2}} \, dv, \quad (44)$$

where the u - v coordinate system has its origin at the coil center with u parallel to x and v parallel to y . The integration proceeds over the first quadrant of the coil cross section; hence, the multiplier of 4. Finally, eq. (44) will give the flux in maxwells when the current density is expressed in A/mm and the coil radius in cm. Numerical integration over u utilized the parameter values $r_o = 0.1$ in., $h = 0.2$ in., and $z_s = 0.025$ in. Using the stripe width $2a = 0.5$ in., one gains entry to Fig. 66 with $h/a = 0.8$ and $z_s/a = 0.1$ to determine the values of $\mathcal{H}_s(u)$. Numerical quadrature yields $\phi = (6.183J_o/\pi)\pi r_o^2$. The result is written in this form because division by πr_o^2 provides the flux density averaged throughout the coil volume. For a sinusoidal drive $J_o = J_m \sin \omega t$, the rms voltage induced in the coil is

$$E_{\text{rms}} = 6.183N_t\omega r_o^2 J_m \times 10^{-8}/\sqrt{2} = 2.1 \text{ mV} \quad (45)$$

for $N_t = 100$ turns, $\omega/2\pi = 30$ kHz, and $J_m = 1/25.4$ A/mm. This result compares favorably with the measured value of 2.0 mV. With the coil calibration in hand, one can assign absolute field values to voltage measurements from the large-scale models of proposed propagation structures. Apart from a slight increase in labor associated with a two-dimensional numerical integration over the coil cross section, one can also use eq. (25) for coil calibration. This has been done, and the results show a comparably small difference between the measured and calculated coil voltage.

XI. MISCELLANEOUS

This section contains items that do not seem to fit elsewhere. These include a three-level conductor circuit, a single level circuit with a lattice of etched islands, a deep implant layer for hard bubble suppression, and an analysis of bubble-bubble interactions.

11.1 Three-level conductor circuits

Advances in processing techniques will ultimately allow fabrication, at high yield, of structures with a complexity unattainable today. One structure that could become practical is the three-level circuit we discuss at this time.

Bias shift due to current flow in a conductor and its consequences was covered in Section III and, in particular, in Fig. 9. By using three-conductor levels and returning the current of the excited level through

the remaining two, one can ease the bias-shift problem and utilize for storage the full conductor width. Another advantage is simpler electronics; only unipolar drivers are needed.

Two possible propagation structures are shown in Fig. 78. Circuit (a) has slots side by side in each of the three levels. The number of + and - signs indicates the relative pole strengths at the ends of the slots when level 1 is energized with two units of current, half of which returns through each of the remaining levels. The wavelike curve sketched below is an approximation to H_z near the lower ends of the slots. If the levels are excited in turn with pulse currents I_j in the sequence [...123...], the wave will move to the right and bubble propagation will be to the right. Interchange any two drives in the sequence, such as [...213...], and bubbles move left. Circuit (b) with in-line oval holes in the three levels generates the indicated relative pole strengths when level 1 is excited. Here also, the sequence

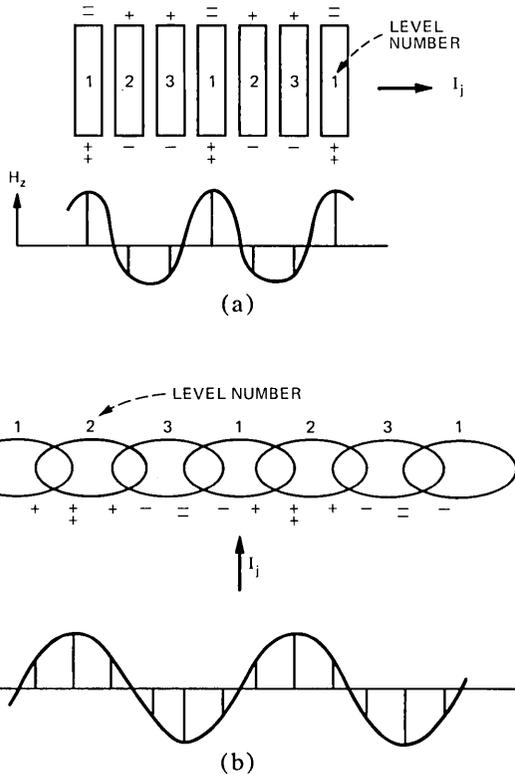


Fig. 78—Two types of three-level circuits are shown: in-line (a) and transverse (b) propagation. In either case, the drive current is returned through the remaining levels to provide z -field cancellation.

[...123...] will cause bubble motion to the right, and the reversal of any two drives in the sequence, to the left.

11.2 Single-level lattice-like circuits

In most single-level conductor circuits, a rather precise fit of the bubble with the feature which produces the static offset field is required. In other words, these circuits cannot tolerate a change in bubble size. A lone exception is that of Dekker in which permalloy points driven by an in-plane field generate the offset field. Other alternatives are possible. From the magnetostatic equivalence of bubbles and cylindrical holes etched in a garnet layer (craters), one can deduce that, at a distance, craters repel bubbles. Therefore, an array of craters can provide interstitial sites that precisely position bubbles. Bubble motion would then be somewhat analogous to electrons jumping from trap to trap in a semiconductor. Devices based on this principle are under study.

11.3 Deep implantation for hard bubble suppression

Data on the performance of implanted and nonimplanted films in dual-conductor devices were presented in Figs. 32 to 35. In particular, Figs. 32 and 34 point out that the minimum drive currents are higher when implanted films are used. This we attribute to a reduction in the effective drive fields because of the shielding effect of implanted layers. If the garnet is implanted at the garnet-substrate interface, this shielding will be eliminated. Experiments with a 2.25- μm thick YSmLuCaGe garnet implanted at 2×10^{16} with H^+ at 300 Kev confirmed this expectation. We anticipate that deep implantation will be especially useful if the garnet layers are less than a micron thick since the buried layer will precisely determine bubble height and, in addition, decouple bubbles from the sometimes troublesome substrate-garnet interface.

It is known that an in-plane field can be used to convert bubbles into the $S = 0$ state; however, the field must be very large if used with small bubble garnets. An analysis by Okabe⁴⁹ suggests that the magnitude of this field will be reduced with a very thick implant layer. A thick implant layer is practical only if it is buried.

11.4 Bubble-bubble interactions

We have already stressed the point that some conductor-access circuits can accommodate a wide range of bubble sizes. There is no "gap," in the sense of the permalloy circuits, to be overcome. In an 8- μm period, dual-conductor circuit, for example, bubbles ranging from 4 μm to less than 0.8 μm have been propagated. This tolerance to bubble size is especially useful when bubble-bubble interactions are a concern.

Bubble-bubble interaction imposes the following constraints on the circumstances under which a pair of adjacent bubbles in a track are stable in position when not driven and move in step when driven. In the absence of any drive excitation, the condition

$$H_c \geq 6\pi^2 M_s h r_o^3 / \lambda^4, \quad (46)$$

where r_o is the bubble radius assures positional stability. That is, the mutual-interaction field gradients are just balanced by the coercivity H_c . During propagation, the trailing bubble must be driven to overcome both its coercivity and an interaction gradient from the leading bubble. The addition of this interaction is equivalent to a doubling of H_c . The drive current must therefore generate a gradient $\partial H_z / \partial x$ given by

$$2r_o \partial H_z / \partial x \geq 16H_c / \pi. \quad (47)$$

It follows that, since $J \sim \partial H_z / \partial x$, $P/\text{bit} \sim (H_c r_o)^2$. From (46), it is seen that $r_o \sim H_c^{1/3}$ so

$$P/\text{bit} \sim H_c^{4/3}. \quad (48)$$

Equation (48) tells us that low-coercivity garnets are preferable to high-coercivity garnets since, with their use, the power dissipation will be reduced in spite of the interaction constraint (46).

XII. SUMMARY AND CONCLUSIONS

The purpose of this paper was to introduce new, very high speed, conductor-access devices where apertured conducting sheets replace the serpentine conductors of conventional conductor devices and the rotating-field coils of field-access devices. When apertured sheets are used, the bias field perturbations are highly localized, and therefore the drives can be ranged over wide limits. We have also shown that the impedance of the conducting layers can be made to match that of drivers operated from low-voltage ($\sim 5\text{V}$) power supplies. The design of single, dual, and triple level conductor circuits was discussed.

All the functions for a shift register or a major-minor chip were covered, although some of the designs have not been optimized. For example, we have not as yet detected at a megahertz stepping rate. However, RF detection does look promising, and the redesign of both the detection electronics and the detector is in progress, and the results will be reported later.

The material requirements for the current-access device technology have been discussed. The high-speed performance inherent in this technology places more emphasis on the material parameters of bubble mobility and dynamic coercivity. Using the $(\text{YLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$ system as a starting point, a technique has been presented to exercise

the bubble mobility while maintaining adequate uniaxial anisotropy. Compositions in the three systems, $(\text{YLuSmCa})_3(\text{FeGe})_5\text{O}_{12}$, $(\text{LaLu-SuCa})_3(\text{FeGe})_5\text{O}_{12}$, and $(\text{LaLuSm})_3(\text{FeGa})_5\text{O}_{12}$, have been compared and the limitations found in the systems discussed. A simplified film composition in the $(\text{LaLuSm})_3(\text{FeGa})_5\text{O}_{12}$ systems is observed that exhibits 1.7- μm diameter bubbles along with high mobility and low dynamic coercivity.

Plasma etching has been used to define circuit features in the AlCu films. The anisotropic etch characteristics of the plasma etch process have produced metal patterns dimensionally identical to the photolithographically defined etch masks. Coupled to the high material-etch rate selectivity, the plasma etch process appears to be ideal for producing the very small period devices that current-access technology promises. Circuits have also been successfully fabricated using Au conductors patterned by ion milling technology; for Al conductors, however, the plasma etching has been the more practical etch method. Thin, conformal, insulating layers of SiO_2 have been successfully deposited by plasma deposition at low substrate temperatures, and it appears that this process is particularly suited for fabricating conductor-access devices.

Current-access and field-access devices differ, of course, in many ways. In terms of analytical complexity, the nonlinear permalloy-bubble interactions of the field-access scheme have been removed in current-access devices because permalloy is not used. One may anticipate, therefore, simpler, more frequent, and more accurate analyses in this new technology.

It is too early to conclude that conductor-access devices will become the technology of the next generation bubble devices. In this context, we include all the conductor-access devices now under development here and elsewhere. It is possible, for example, that the apertured sheet drive approach applied to the bubble-lattice file⁷ will make that structure more attractive. There are some features that only conductor-access devices possess. If high speed is a necessity, as in video-based systems, conductor circuits are the only option available to the device designer at present. The anticipated small package size will be significant in applications for which space is a premium. The ability to run with logic-level power supplies is another plus for conductor devices. Finally, conductor circuits will require fewer precision analog control circuits since the critical control function to rotating field timing has been eliminated.

XIII. ACKNOWLEDGMENTS

Many people have contributed to this project. The authors thank W. Biolsi, W. L. Swartzwelder, and L. C. Luther for the growth of

garnet films, R. Wolfe for many informative discussions on ion implantation, G. P. Vella-Coleiro, R. D. Pierce, and W. Venard for material characterization, W. E. Hess for test electronics, J. A. Morrison for mathematical advice, and J. L. Blue for fruitful discussions about numerical analysis and programming.

We would also like to acknowledge the efforts of D. L. Fraser, Jr. and D. N. Wang who plasma-etched many of our early circuits, W. A. Johnson for fabricating Au conductor circuits by ion milling, R. O. Miller for constructing our plasma etch station, and R. S. Wagner for foresight and advice in developing the plasma etch process, and J. L. Smith and J. E. Geusic for words of encouragement during moments of frustration. Finally, special thanks should be given to R. T. Anderson for device-coding and to our patient typist, A. L. Becker.

XIV. MEMORIAL

This paper is dedicated to the memory of Edward M. Walters, a close friend and pioneer in magnetic bubbles.

REFERENCES

1. A. H. Bobeck, "Properties and Device Applications of Magnetic Domains in Orthoferrites," *B.S.T.J.*, 46, No. 8 (October 1967), pp. 1901-1925.
2. A. J. Perneski, "Propagation of Cylindrical Magnetic Domains in Orthoferrites," *IEEE Trans. on Magnetics*, *MAG-5*, No. 3 (September 1969), pp. 554-557.
3. R. Wolfe, J. C. North, W. A. Johnson, R. R. Spiwak, L. J. Varnerin, and R. F. Fisher, "Ion Implanted Patterns for Magnetic Bubble Propagation," *AIP Conf. Proc.*, No. 10 (1972), pp. 339-343.
4. A. H. Bobeck and I. Danylchuk, "Characterization and Test Results for a 272K Bubble Memory Package," *IEEE Trans. on Magnetics*, *MAG-13*, No. 5 (September 1977), pp. 1370-1372.
5. J. A. Copeland, J. G. Josenhans, and R. R. Spiwak, "Circuit and Module Design for Conductor-Groove Bubble Memories," *IEEE Trans. on Magnetics*, *MAG-9*, No. 3 (September 1973), pp. 489-492.
6. T. J. Walsh and S. H. Charap, "Novel Bubble Drive," *AIP Conf. Proc. No. 24* (1974), pp. 550-551.
7. O. Voegeli, B. A. Calhoun, L. L. Rosier, and J. C. Slonczewski, "The Use of Bubble Lattices for Information Storage," presented at the 1974 Conf. Magnetism and Magnetic Materials.
8. H. L. Hu, T. J. Beaulieu, D. W. Chapman, D. M. Franich, G. R. Henry, L. L. Rosier, and L. F. Shew, "1K Bit Bubble Lattice Storage Device: Initial Tests," *J. Appl. Phys.*, 49, No. 3 (March 1978), pp. 1913-1917.
9. E. H. L. J. Dekker, K. L. L. van Mierloo, and R. de Werdt, "Combination of Field and Current Access Magnetic Bubble Circuits," *IEEE Trans. on Magnetics*, *MAG-13*, No. 5 (September 1977), pp. 1261-1263.
10. J. A. Copeland, J. P. Elward, W. A. Johnson, and J. G. Ruch, "Single-Conductor Magnetic-Bubble Propagation Circuits," *J. Appl. Phys.*, 42, No. 4 (March 1971), pp. 1266-1267.
11. T. H. O'Dell, *Magnetic Bubbles*, London, England: MacMillan, 1974.
12. W. J. Tabor, A. H. Bobeck, G. P. Vella-Coleiro, and A. Rosencwaig, "A New Type of Cylindrical Magnetic Domain (Hard Bubble)," *AIP Conf. Proc.*, No. 10 (1972), pp. 442-457.
13. J. C. Slonczewski, J. C. Malozemoff, and O. Voegeli, "Statics and Dynamics of Bubbles Containing Bloch Lines," *AIP Conf. Proc.*, No. 10 (1972), pp. 458-477.
14. R. W. Patterson, "Annihilation of Bloch Lines In Hard Bubbles," *AIP Conf. Proc.*, No. 24 (1974), pp. 608-609.
15. H. L. Hu and E. A. Giess, "Hard Bubble Suppression and Controlled State Gener-

- ation of One Micron Bubbles in Ion-Implanted Garnet Films," IEEE Transactions on Magnetics, *MAG-11*, No. 5 (September 1975), pp. 1085-1087.
16. V. M. Benrud, G. L. Forslund, M. M. Hanson, R. L. Horst, A. D. Kaske, J. A. Kolling, D. S. Lo, M. J. Nordstrom, H. N. Oredson, W. J. Simon, C. H. Tolman, and E. J. Torok, "Oligatonic Film Memories," *J. Appl. Phys.*, *42*, No. 4 (March 1971), pp. 1364-1373.
 17. H. J. Levinstein, S. J. Licht, R. W. Landorf, and S. L. Blank, "Growth of High-Quality Garnet Thin Films from Supercooled Melts," *Appl. Phys. Lett.*, *19*, No. 11 (December 1971), pp. 486-488.
 18. S. L. Blank and J. W. Nielsen, "The Growth of Magnetic Garnets by Liquid Phase Epitaxy," *J. Crystal Growth*, *17* (1972), pp. 302-311.
 19. S. L. Blank, J. W. Nielsen, and W. A. Biolsi, "Preparation and Properties of Magnetic Garnet Films Containing Divalent and Tetravalent Ions," *J. Electrochem. Soc.*, *123*, No. 6 (June 1976), pp. 856-863.
 20. S. L. Blank, B. S. Hewitt, L. K. Shick, and J. W. Nielsen, "Kinetics of LPE Growth and its Influence on Magnetic Properties," *AIP Conf. Proc.*, No. 10, (1972), pp. 256-270.
 21. E. A. Giess, B. E. Argyle, D. C. Cronemeyer, E. Klokhholm, T. R. McGuire, D. F. O'Kane, T. S. Plaskett, and V. Sadagopan, "Europium-Yttrium Iron-Gallium Garnet Films Grown by Liquid Phase Epitaxy on Gadolinium Gallium Garnet," *AIP Conf. Proc.*, No. 5, (1971), pp. 110-114.
 22. R. L. Barnes, *Advances in X-Ray Analysis*, Vol. 15, K. J. Heinrich, C. F. Barrett, J. D. Newkirk and C. D. Rund, eds., New York: Plenum Press, 1972, pp. 330.
 23. R. D. Pierce, "Magnetic Characterization of Bubble Garnet Films in an LPE Growth Facility," *J. Crystal Growth*, *27* (1974), pp. 299-305.
 24. G. P. Vella-Coleiro and W. J. Tabor, "Measurement of Magnetic Bubble Mobility in Epitaxial Garnet Films," *Appl. Phys. Lett.*, *21*, No. 1 (July 1972), pp. 7-8.
 25. R. C. LeCraw and R. D. Pierce, "Temperature Dependence of Growth-Induced Magnetic Anisotropy in Epitaxial Garnet Films by Resonance Techniques," *AIP Conf. Proc.*, No. 5, (1971), pp. 200-204.
 26. A. H. Boeck, P. I. Bonyhard, and J. E. Geusic, "Magnetic Bubbles—An Emerging New Memory Technology," *Proc. IEEE*, *63*, No. 8 (August 1975), pp. 1176-1195.
 27. S. L. Blank, R. Wolfe, L. C. Luther, R. C. LeCraw, T. J. Nelson, and W. A. Biolsi, "Design and Development of Single-Layer, Ion-Implantable Small Bubble Materials for Magnetic Bubble Devices," *Proc. 24th Annual Conf. on Mag. and Mag. Matl.*, *J. Appl. Phys.*, March, 1979.
 28. M. Kestigian, A. B. Smith, and W. R. Bekebrede, "(YSmLu)₃(FeGa)₅O₁₂ for 1 to 3 μm-Diameter Bubble Devices," *J. Appl. Phys.*, *49*, No. 3 (March 1978), pp. 1873-1875.
 29. G. P. Vella-Coleiro, F. B. Hagedorn, S. L. Blank, and L. C. Luther, "Coercivity in 1.7-μm Bubble Garnet Films," *Proc. 24th Annual Conf. on Mag. and Mag. Matl.*, *J. Appl. Phys.*, March, 1979.
 30. E. R. Czerlinsky and W. G. Field, "Magnetic Properties of Ferromagnetic Garnet Single Crystals," *Solid State Phys. Electronics Telecommun.*, *Proc. Intern. Conf.*, Brussels, 3 (1958) (Pub. 1960), pp. 488-499.
 31. W. A. Johnson, private communication.
 32. A. K. Sinha, and T. T. Sheng, "The Temperature Dependence of Stresses in Aluminum Films on Oxidized Silicon Substrates," *Thin Solid Films*, *48*, No. 2 (February 1978), pp. 117-126.
 33. B. J. Roman, private communication.
 34. D. K. Rose, "Planar Processing for Magnetic Bubble Devices," *IEEE Trans. on Magnetics*, *MAG-12*, No. 6 (November 1976), pp. 618-621.
 35. C. J. Mogab, and W. R. Harshbarger, "Plasma Processes Set to Etch Finer Lines with Less Undercutting," *Electronics* (August 31, 1978), pp. 117-121.
 36. C. M. Melliar-Smith and C. J. Mogab, in *Thin Film Processes*, J. L. Vossen and W. Kern, eds., New York: Academic Press, 1978, pp. 497-555.
 37. H. J. Levinstein, D. N. Wang, Patent Pending.
 38. C. J. Mogab and F. B. Alexander, private communication.
 39. R. M. Goldstein and J. A. Copeland, "Permalloy Rail-Cylindrical Magnetic Domain Systems," *J. Appl. Phys.*, *42*, No. 6 (May 1971), pp. 2361-2367.
 40. R. M. Goldstein, M. Shoji, and J. A. Copeland, "Bubble Forces in Cylindrical Magnetic Domain Systems," *J. Appl. Phys.*, *44*, No. 11 (November 1973), pp. 5090-5095.
 41. D. E. Gray, Ed., *American Institute of Physics Handbook*, New York: MacMillan, 1963, p. 5-26.
 42. H. B. Dwight, *Tables of Integrals and Other Mathematical Data*, fourth ed., New York: MacMillan, 1961, pp. 12-13.

43. L. M. Milne-Thomson, *Theoretical Hydrodynamics*, fifth ed., New York: MacMillan, 1968, p. 158.
44. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, New York: Academic Press, 1965, pp. 81-85.
45. J. L. Blue, "Automatic Numerical Quadrature," *B.S.T.J.*, 56, No. 9 (November 1977), pp. 1651-1678.
46. H. W. Richmond, "On the Electrostatic Field of a Plane or Circular Grating Formed of Thick Rounded Bars," *Proc. London Math. Soc.*, 22 (April 1923), pp. 389-403.
47. A. A. Thiele, "Device Implications of the Theory of Cylindrical Magnetic Domains," *B.S.T.J.*, 50, No. 3 (March 1971), pp. 725-773.
48. T. Hsu, "Control of Domain Wall States for Bubble Lattice Devices," *AIP Conf. Proc. No. 24*, 1974, pp. 624-626.
49. Yoichi Okabe, "A Model to Predict the Upper Cap-Switch Field of Various Capping Layers," *IEEE Trans. on Magnetics, MAG-14*, No. 5 (September 1978), pp. 602-604.

Contributors to This Issue

Stuart L. Blank, B.S. (ceramic engineering), 1962, Alfred University, New York State College of Ceramics; M.S. (materials science), 1964, Ph.D. (materials science), 1967, University of California; Bell Laboratories, 1969—. Mr. Blank has been involved in investigating the growth of materials by liquid-phase epitaxy, in developing new materials for magnetic bubble device applications, and in transferring the epitaxial growth technology to Western Electric for production of magnetic bubble materials. He is also investigating crystal growth, phase transitions, and defects in oxide materials. At present, he is supervisor of the Epitaxial Materials and Processes group in the Electronic Materials and Processes Department. Member, American Ceramic Society, National Institute of Ceramic Engineers, American Association for Crystal Growth.

Andrew H. Bobeck, B.S. (electrical engineering), 1948, M.S. (electrical engineering), 1949, Purdue University; Bell Laboratories, 1949—. Mr. Bobeck has specialized in the development of magnetic components. His early work involved the first solid-state driven core memory and the twistor memory. His interest in magnetic logic and storage eventually led to the magnetic bubble concept. A significant contribution to magnetic bubble development was the discovery that garnet materials can be prepared with a growth-induced uniaxial anisotropy, an observation that led to the epitaxial garnet films in general use today. At present, he supervises an activity aimed at the development of very high density bubble devices. Distinguished Engineering Alumnus, 1968, Doctor of Engineering (honorary), 1972, Purdue University; Stuart Ballantine medal, 1973, Franklin Institute; co-recipient, 1975, Morris Liebmann award; National Academy of Engineering, 1975; Valdemar Poulsen Gold Medal, 1976, Danish Academy of Technical Sciences. Fellow, IEEE, Tau Beta Pi, Eta Kappa Nu. He has been awarded more than 115 patents.

A. Duane Butherus, B.S. (chemistry), 1961, Andrews University; Ph.D. (chemistry), 1967, Michigan State University; Bell Laboratories,

1967—. Mr. Butherus has done materials investigations in a range of fields including polymers, solid-state ion conductors, and other electrochemically related systems. His current interests are in plasma chemistry and bubble memory processing techniques.

Frank J. Ciak, Bell Laboratories, 1968—. Mr. Ciak has worked in the Device Design Group on the development of magnetic bubble memory devices and associated test equipment. He designed magnet-setting apparatus and in-process bubble device test stations currently in use at Western Electric, Reading, Pa.

Nicholas F. Maxemchuk, B.S.E.E., 1968, City College of New York; M.S.E.E., 1970, Ph.D. 1974, University of Pennsylvania; RCA Laboratories, 1968-1976; Bell Laboratories, 1976—. Since joining Bell Laboratories, Mr. Maxemchuk has been a member of the technical staff of the Electronic and Computer Systems Research Laboratory.

Carlo Scagliola, Dr. Ing. (Electronic Engineering) 1970, University of Pisa, Italy. Mr. Scagliola has been with CSELT (Centro Studi e Laboratori Telecomunicazioni), Turin, Italy since 1970, where he is presently Head, Speech Processing Department. He has been engaged in adaptive speech coding, in assessment of the quality of digitally coded speech, and in studies on automatic synthesis of the Italian language. He served as a consultant at Bell Laboratories from January 1977 through January 1978.

David H. Smithgall, B.S. (E.E.), 1967, M.S. (E.E.), 1968, Ph.D. (E.E.), 1970, Cornell University; Western Electric Engineering Research Center, 1970—. Mr. Smithgall has worked in various areas of process control and has published in the areas of optical waveguides, control, microprocessor applications, and optical fiber measurements. His current interests include the measurement, characterization, and control of optical fiber manufacturing processes. Member, IEEE.

Walter Strauss, B.E.E., 1948, College of the City of New York; Ph.D. (physics), Columbia University, 1961; Department of Electrical Engineering, College of the City of New York, 1948-1953; Bell Laboratories, 1960—. At Bell Laboratories, Mr. Strauss has been concerned with problems related to magnetoelasticity, piezoelectricity, and magnetic devices. Member, American Physical Society, Eta Kappa Nu, New York Academy of Sciences.

John Stuller, B.S.E.E., 1963, Massachusetts Institute of Technology; M.S.E.E.; 1967, University of Southern California; Ph.D., 1971, University of Connecticut; Northrop Nortronics Corp., 1963-1965; TRW Systems Group, 1965-1968; Perkin Elmer Corp. 1970-1971; Department of Electrical Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada, 1971-1974; Bell Laboratories, 1977—. At Bell Laboratories, Mr. Stuller has been engaged in analytical and simulation studies of both intraframe and interframe video coders. His primary area of interest is in the statistical theory of communications and its application.

Hans S. Witsenhausen, I.C.M.E., 1953 and Lic. Sc. Phys., 1956, Université Libre de Bruxelles; Ph.D., 1966, Massachusetts Institute of Technology; Electronic Associates, 1957-63; Electronic Systems Laboratory, 1963-65; Hertz Fellow, 1965-66; Bell Laboratories, 1966—. Mr. Witsenhausen is currently working in the Communications Analysis Research Department. He has published articles on hybrid computation, control theory, optimization, geometric inequalities, and other applied mathematical fields. He was Senior Fellow at Imperial College of Science and Technology, London, in 1972, visiting professor at M.I.T. in 1973, and Vinton Hayes senior fellow at Harvard University in 1975-76. Member, American Mathematical Society, IEEE, Sigma Xi.

Papers by Bell Laboratories Authors

BIOLOGY

Acoustical Properties of the Vocal Tract. M. M. Sondhi, *Physics* (November 1978), pp. 1-2.

Acoustic Characterization and Computer Simulation of the Air Volume Displaced by the Vibrating Vocal Cords: Lateral and Longitudinal Motion. J. L. Flanagan and K. Ishizaka, *Proceedings of the International Symposium on Articulatory Modeling* (July 1977), pp. 251-261.

A ^{13}C NMR Study of Gluconeogenesis in Isolated Rat Liver Cells. S. M. Cohen, S. Ogawa, and R. G. Shulman, *Frontiers of Biological Energetics*, Vol. 2, New York: Academic Press, 1978, pp. 1357-1364.

Early Experience with Sour and Bitter Solutions Increases Subsequent Ingestion. R. M. London, C. T. Snowdon, and J. M. Smithana, *Physiology and Behavior* (June 1979).

The Solution Conformation of Malformin A. A. E. Tonelli, *Biopolymers*, 17 (1978), pp. 1175-1179.

CHEMISTRY

Block Copolymer Theory 5. Spherical Domains. E. Helfand and Z. R. Wasserman, *Macromolecules*, 11 (September-October 1978), pp. 960-966.

^{13}C Chemical Shifts of the Polypropylene "Model" Compounds 3,5-Dimethylheptane and 3,5,7-Trimethylnone. A. E. Tonelli, *Macromolecules*, 12 (1979), pp. 83-85.

Calculated and Measured ^{13}C NMR Chemical Shifts of the 2,4,6-Trichloroheptanes and Their Implications for the ^{13}NMR Spectra of Poly(vinyl chloride). A. E. Tonelli, F. C. Schilling, W.H. Starnes, Jr., L. Shepherd, and I. M. Plitz, *Macromolecules*, 12 (1979), pp. 78-83.

CuCl: More Facts Generate More Thoughts on High Temperature Superconductivity. J. A. Wilson, *Phil. Mag. B.*, 38 (1978), pp. 427-444.

The High Temperature Deposition and Evaluation of Phosphorus-or Boron-Doped Silicon Dioxide Films. A. C. Adams, C. D. Capio, S. E. Haszko, G. I. Parisi, E. I. Povelonis, and M. D. Robinson, *J. Electrochem. Soc.*, 126 (February 1979), pp. 313-319.

Measuring the Phosphorus Concentration in Deposited Phosphosilicate Films. A. C. Adams and S. P. Murarka, *J. Electrochem. Soc.*, 126 (February 1979), pp. 334-338.

The Nuclear Modulation Effect in Electron Spin Echoes for Complexes of Cu^{2+} and Imidazole with ^{14}N and ^{15}N . W. B. Mims and J. Peisach, *J. Chem. Phys.*, 69, No. 11 (December 1, 1978), pp. 4921-4930.

Parametric Counting of Samples Generating Complex X-Ray Spectra. P. F. Schmidt, J. E. Riley, Jr., and D. J. McMillan, *Anal. Chem.*, 51 (February 1979), pp. 189-194.

Piezoelectric and Electrooptic Constants of Crystals. D. F. Nelson, *Landolt-Bornstein*, New Series, Berlin: Springer-Verlag, 1979, pp. 495-505.

Preparation and Characterization of Tetra (2,4-Pentanedionato)-Hexa(Benzotriazolato)-Penta-Cu (II). J. H. Marshall, *Inorg. Chem.*, 17 (December 1978), pp. 3711-3713.

X-Ray Absorption Studies of Metalloproteins. R. G. Shulman, *Trends in Biochemical Sciences*, 3, No. 12 (December 1978), pp. N282-N283.

COMPUTING

Application of Multilinear Algebra in Optimal Regulation of Nonlinear Polynomial Systems. M. R. Buric, Proceedings of 16th Annual Allerton Conference on Communications, Control and Computing, 1 (October 4, 1978), pp. 655-663.

Development of the Law of Computer Software Protection. R. O. Nimtz, Journal of Patent Officer Society, 61 (January 1979), pp. 3-43.

IMSMON-A Low Overhead Monitor for IMS Systems. B. Domanski, 1978 Compact Measure Group 1X Conference Proceedings on the Management and Evaluation of Computer Performance (December 5, 1978), pp. 115-123.

Model for the User Services Layer of a Network Architecture. K. Coates, 1969 ACM Computer Science Conference (February 20, 1979), p. 38.

On-Line PWB Design Verification and Associated Data Structuring Problems. J. C. Foster and S. Pardee, Proceedings, International Conference, Interactive Techniques in Computer Aided Design, Bologna, Italy, 78 CH1289-8C (September 21-23, 1978), pp. 265-271.

What Employers Look for When Hiring Computer Science Graduates. M. M. Irvine, Proceedings of 1978 Annual Conference Association for Computing Machinery, 1, pp. 404-405.

ELECTRICAL AND ELECTRONIC ENGINEERING

Alpha Particles Tracks in Silicon and Their Effect on Dynamic MOS RAM Reliability. D. S. Yaney, J. T. Nelson, and L. L. Vaniskike, IEEE Trans. Electron Dev., ED-26 (January 1979), pp. 10-16.

Defect Etch for (100) Silicon Evaluation. D. G. Schimmel, J. Electrochem. Soc., 126 (March 1979), pp. 479-483.

Design of an Integrated Circuit for the TIC Low-Power Line Repeater. P. C. Davis, J. F. Graczyk, and W. A. Griffin, IEEE J. of Solid State Circuits, SC-14 (February 1979), pp. 109-120.

Digital Communications—The Silent (R) Evolution? M. R. Aaron, IEEE Communications Magazine, 17, No. 1 (January 1979), pp. 16-26.

The Evolution of the Discrete Crystal Single-Sideband Selection Filter in the Bell System. T. H. Simmonds, Jr., Proc. IEEE, 67, No. 1 (January 1979), pp. 109-115.

Fatigue Behavior of Flex Cables and Circuits. W. Engelmaier, Electronic Packaging and Production, 19 (February 1979), pp. 110-118.

A 50 MHz Phase and Frequency Locked Loop. R. R. Cordell, C. N. Dunn, J. B. Forney, and W. G. Carrett, 1979 ISSCC Digest of Technical Papers 22 (February 14, 1979), pp. 234-235.

Final State Structure in XPS. G. K. Wertheim, Electron Spectroscopy, Theory, Techniques and Application, Vol. 2, New York: Academic Press, 1978, pp. 259-284.

Foreword-Special Issue on Quantum-Electronic Devices for Optical Fiber Communications. T. Li, IEEE J. of Quantum Electronics, QE-14 (November 1978), p. 790.

InGaAsP/InP Photodiodes: Microplasma-Limited Avalanche Multiplication at 1-1.3 m Wavelength. T. P. Lee, C. A. Burrus, A. G. Dentai, IEEE J. Quantum Electronics, QE-15 (January 1979), pp. 30-35.

An Integrated PCM Encoder Using Interpolation PCM Decoder. B. A. Wooley, D. C. Fowles, J. L. Henry, and C. E. Williams, IEEE J. Solid State Circuits, SC-14, No. 1 (February 1979), pp. 20-25.

P-N-P-N Optical Detectors and Light-Emitting Diodes. J. A. Copeland, A. G. Dentai, and T. P. Lee, IEEE J. Quantum Electronics, QE-14 (November 1978), pp. 810-813.

Self-Contained LED-Pumped Single-Crystal. J. Stone and C. A. Burrus, Fiber and Integrated Optics, 2 (1979), pp. 19-46.

Semiconductor-Laser Self-Pulsing Due to Deep-Level Traps. J. A. Copeland, Electron. Lett. (December 7, 1978), pp. 809-810.

Signal-To-Noise Ratio Statistics for Nondispersive Fading in Radio Channels with Cross Polarization Interference Cancellation. N. Amitay, IEEE Transactions on Communication, COM-27, No. 2 (February 1979), pp. 498-502.

Slow Positron Emission from Metal Surfaces. A. P. Mills, P. M. Platzman, and B. L. Brown, *Phys. Rev. Lett.*, *41* (1978), p. 1076.

A Stereo-Fiberscope with a Magnetic Interlens Bridge for Laryngeal Observation. O. Fujimura, T. Baer, S. Niimi, *J. Acoust. Soc. Amer.*, *65*, No. 2 (1979), pp. 478-480.

Supplement to Literature on Digital Signal Processing. J. F. Kaiser and H. D. Helms, Special Publication of IEEE ASSP Society, IEEE, New York, *JH4686-2 ASSP* (1979), p. 237.

MATERIALS SCIENCE

An AES Study of the Surface Composition of Ion-Etched Iron-Chromium Alloys: Effect of Absorbed CO. R. P. Frankenthal and D. E. Thompson, *J. Vacuum Sci. and Technol.*, *16* (January-February 1979), pp. 6-12.

Control of Substrate Tube Diameter During MCVD Preform Preparation. P. D. Lazay and W. G. French, Conference Digest (OSA—Topical Meeting on Optical Fiber Communication) (March 5, 1979), pp. 50-52.

Core Hole Screening in Lanthanide Metals. G. Creceline, G. K. Wertheim, and D. N. E. Buchanan, *Phys. Rev. B*, *18*, No. 12 (December 15, 1978), pp. 6519-6524.

Effect of pH on the Rate of Corrosion of Gold in Acid Sulfate Solutions. B. S. Duncan and R. P. Frankenthal, *J. Electrochem. Soc.*, *126* (January 1979), pp. 95-97.

The Intermediate Valence State in Rare Earth Compounds; Bulk and Surface Manifestations. G. K. Wertheim, *J. Electron. Spectr.*, *15* (January 1979), pp. 5-14.

The Kinetic Photochemistry of the Marine Atmosphere. T. E. Graedel, *J. Geophys. Res.*, *84* (1979), pp. 273-286.

Mechanical Testing Standards as Developed by the Voluntary Consensus System: Present and Future. A. Fox, Golden Gate Metals and Welding Conference Abstracts, Session 23 (January 29, 1979).

Processing and Magnetic Properties of Low-Loss and High Stability Mn-Zn Ferrites. B. B. Ghatge, *Mater. Sci. Res., Process. Cryst. Ceram.*, *11* (1978), pp. 369-379.

Sealed Extrusion-A Coating Method for Photoprinting UV Curable Liquid Resin Films. G. B. Fefferman and T. V. Lake, Fourth Radiat. Int. Conf. (September 1978), pp. 1-18.

Variable Resolution Capability for Multichannel Filter Spectrometers. P. S. Henry, *Rev. Sci. Instrum.*, *50* (February 1979), pp. 185-192.

GENERAL MATHEMATICS AND STATISTICS

Comment on Schruben and Margolin. C. L. Mallows, *J. Amer. Statist. Assoc.*, *73* (September 1978), p. 520.

Complexity Results for Bandwidth Minimization. M. R. Garey, R. L. Graham, D. S. Johnson, and D. E. Knuth, *SIAM J. Appl. Math.*, *34* (May 1978), pp. 477-495.

Factor Analysis and Principal Components: The Bilinear Methods. J. B. Kruskal, *The International Encyclopedia of Statistics*, The Free Press (1978), pp. 307-330.

Good and Optimal Ridge Estimators. R. L. Obenchain, *The Annals of Statistics*, *6* (September 1978), pp. 1111-1121.

A Note on Bisecting Minimum Spanning Trees. W. M. Boyce, M. R. Garey, and O. S. Johnson, *Networks*, *8* (1978), pp. 187-192.

Postscript to "Transformations of Data." J. B. Kruskal, *International Encyclopedia of Statistics*, New York: The Free Press, Vol. 2, 1978, pp. 1055-1056.

PHYSICS

Amorphous Metal Films by Getter-Sputtering at 25°K. J. J. Hauser, R. J. Schutz, and W. M. Augustyniak, *Phys. Rev.*, *18B* (October 15, 1978), pp. 3890-3896.

Angle-Resolved Photoemission from Surface and Adsorbates. N. V. Smith, *Journal de Physique*, *39C4* (July 1978), pp. 161-168.

Diamagnetic Structure of Rb in Intense Magnetic Fields. N. P. Economou, R. R. Freeman, and P. F. Liao, *Phys. Rev. A*, *18* (December 1978), pp. 2506-2509.

The Diffraction of He Atoms at a Si(111) 7 × 7 Surface. M. J. Cardillo and G. E. Becker, *Phys. Rev. Lett.*, *42* (February 1979), pp. 508-511.

- Diffraction of Photoelectrons Emitted from Core Levels of Te and Na Atoms Adsorbed on Ni (001).** D. P. Woodruff, D. Norman, B. W. Holland, N. V. Smith, H. H. Farrell, and M. M. Traum, *Phys. Rev. Lett.*, *41* (October 16, 1978), pp. 1130-1133.
- Dynamic Central Peaks and Phonon Interactions Near Structural Phase Transitions.** P. A. Fleury and K. B. Lyons, *Proc. of the Int'l. Conf. on Lattice Dynamics*, Paris, France (1978), p. 731.
- The Effect of Strain-Induced Band-Gap Narrowing on High Concentration Phosphorus Diffusion in Silicon.** R. B. Fair, *J. Appl. Phys.* *50* (February 1979), pp. 860-868.
- Effects of Atomic Order in α - and β -phase AgCd Alloys Studied by X-Ray Photoelectron Spectroscopy.** G. Crecelius and G. K. Wertheim, *Phys. Rev. B*, *18*, No. 12 (December 15, 1978), pp. 6525-6530.
- Effects of Magnetic Fields on Four Wave Mixing Processes in Atomic Vapors.** N. P. Economou, R. R. Freeman, and G. C. Bjorklund, *Optics Letters*, *3* (December 1978), pp. 209-211.
- Light Scattering Determinations of Dynamic Four Point Correlation Functions.** P. A. Fleury, *Proc. of the NATO Advanced Study Inst. on Correlation Functions and Quasi-particle Transactions in Condensed Matter*, New York: Plenum Press, 1978, pp. 325-365.
- The Magnetic Behavior of an $S = \frac{1}{2}$ Amorphous Antiferromagnet.** R. B. Kummer, R. E. Walstedt, S. Geschwind, V. Narayanamurti, G. E. Devlin, *Phys. Rev. Lett.*, *40*, No. 16 (April 17, 1978), pp. 1098-1100.
- Magnetic Field Quantum Beats in Two-Photon Free-Induction Decay.** N. P. Economou and P. F. Liao, *Optics Letters*, *3* (November 1978), pp. 172-174.
- Nitric Acid Vapor Line Parameters Measured By Co Laser Transmittance.** L. A. Farrow, R. E. Richton, and C. P. Karnas, *Appl. Opt.*, *18* (January 1, 1979) pp. 76-81.
- Piezooptic and Elastooptic Coefficients.** R. F. S. Hearmon and D. F. Nelson, *Landolt-Börnstein, New Series* (1979), pp. 505-551.
- Real-Time Autocorrelation Interferometer.** R. L. Fork and F. A. Beisser, *Appl. Opt.*, *17* (November 15, 1978), pp. 3534-3535.
- The Relation of Solitons to Polaritons in Coupled Systems.** D. F. Nelson, *Solitons and Condensed Matter Physics*, Berlin: Springer-Verlag, 1978, pp. 187-190.
- Thermal Convection and Crystal Growth.** P. G. Simpkins, and P. A. Blythe, *Physics News in 1977*, American Inst. of Physics Booklet, *R-28 1* (November 1977), pp. 52-53.
- Time-Resolved Photoluminescence Spectroscopy in Amorphous As_2S_3 .** M. A. Bosch and J. Shah, *Phys. Rev. Lett.*, *42*, pp. 118-121.
- X-ray Photoemission Study of Ce-Pnictides.** Y. Beer, R. Hanger, C. Zürdrer, M. Campagna, G. K. Wertheim, *Phys. Rev. B*, *18* (October 15, 1978), pp. 4433-4439.

PSYCHOLOGY

- Subjective Detection of Differences in Variance from Small Samples.** R. L. Fike and W. R. Ferrell, *Organ. Behav. Hum. Perf.*, *22* (December 1978), pp. 262-278.

SYSTEMS ENGINEERING AND OPERATIONS RESEARCH

- Communications Privacy.** A. Gersho, *IEEE Commun. Soc. Mag.*, *16* (November 1978), p. 2.
- The Complexity of the Network Design Problem.** D. S. Johnson, J. K. Lenstra, and A. H. G. Kan, *Networks*, *8* (1978), pp. 279-285.
- D. C. Welcomes MAT.** B. S. Brosius and G. H. Webster, *Telephony* (January 1, 1979), pp. 19-23.
- Digital Techniques for Speech Communications: Some New Dimensions.** J. L. Flanagan, *Proceedings of the IEEE International Conference on Cybernetics and Society* (September 1977), pp. 1-5.
- Double Phase Matching Function.** D. F. Nelson, *J. Opt. Soc. Amer.*, *68* (December 1978), pp. 1780-1781.
- A New Measurement Technique For Telephone Switching Systems.** R. J. Jaeger, Jr. and R. C. Stone, Jr., *Pacific Telecommunications Conference Proceedings* (January 8, 1979), pp. 3C24-3C33.
- The Use of MIL-STD-1050 to Control Average Outgoing Quality.** B. S. Liebesman, *J. Qual. Tech.*, *11*, No. 1 (January 1979), pp. 36-43.

Errata

F. R. K. Chung and F. K. Hwang, "On Blocking Probabilities for a Class of Linear Graphs," *B.S.T.J.*, 57, No. 8 (October 1978), pp. 2915-2925.

We acknowledge the help of M. Horgan of England who pointed out an error in our paper. On page 2918, we define

$$f(\beta_{\lfloor (t-1)/2 \rfloor}) = (1 - \alpha_{t/2})^{\beta_{\lfloor (t-1)/2 \rfloor}} \quad \text{if } t \text{ is even.}$$

The correct definition should be

$$f(\beta_{\lfloor (t-1)/2 \rfloor}) = (1 - \alpha_{t/2-1} \alpha_{t/2} \alpha_{t/2+1})^{\beta_{\lfloor (t-1)/2 \rfloor}} \quad \text{if } t \text{ is even.}$$

F.K.H.

F.R.K.C.

A. N. Netravali, F. W. Mounts, and K. A. Walsh, "Adaptation of Ordering Techniques for Facsimile Pictures with No Single Element Runs," *B.S.T.J.*, 58, No. 4 (April 1979), pp. 857-865.

Fig. 1 (continued), page 859, is actually Fig. 2 (continued), page 861, and Fig. 2 (continued) is Fig. 1 (continued).

B.S.T.J. BRIEF

A 30-GHz Scale-Model, Pyramidal, Horn-Reflector Antenna

By R. A. SEMPLAK

(Manuscript received January 24, 1979)

I. INTRODUCTION

In the early 1940s, the pyramidal horn-reflector antenna was invented¹ at Bell Laboratories, Holmdel, New Jersey. It is now in extensive use in the Bell System 4-, 6-, and 11-GHz, transcontinental, microwave, common-carrier, radio-relay network.² This antenna is a combination of a square electromagnetic horn and a reflector that is a section of the paraboloid of revolution. The apex of the square horn coincides with the focus of the paraboloid. The antenna is essentially a shielded, offset, parabolic antenna, so that very little of the energy incident on the reflector is reflected back into the feed to produce an impedance mismatch.

Radio interference from adjacent paths limits the number of converging routes of a common carrier microwave radio system, and in recent years, demands have been made to improve the sidelobe performance of the pyramidal horn-reflector antenna. In response, blinders (metallic extensions to the sidewalls of the horn) have been developed³ which provide a degree of far sidelobe reduction, i.e., sidelobes beyond 35 degrees from the axis of the main beam.

To continue the investigation for ways of reducing the sidelobe levels of this type antenna, a scale model was built. The model has a numerically-machined, precision reflector, and the scaling factor is 7.5, which means that measurements made at a frequency of 30 GHz will represent the performance of a full-size antenna measured at a frequency of 4 GHz. This paper presents and discusses the measurements made on the scaled model at 30 GHz. Comparisons are made with data obtained (at a frequency of 4 GHz) by others on full-sized antennas.⁴

In the discussion that follows, it should be remembered (from Ref. 4) that longitudinal polarization and longitudinal plane indicate that the electric field in the aperture and the plane of antenna rotation, for radiation measurements, are aligned with the pyramidal horn axis, whereas transverse polarization and transverse plane indicate that the electric field in the aperture and plane of antenna rotation are perpendicular to the horn axis.*

II. DISCUSSION

A study of the historical data on the radiation characteristics of the horn-reflector antenna indicates that, since its inception in the 1940s, disagreements between measured data and theoretical values have existed. For example, in the transverse plane with longitudinal polarization, agreement between measured data and theoretical value exists only out to the first sidelobe. With transverse polarization, good agreement extends out to the fourth sidelobe.

Since the aperture of the horn-reflector antenna is illuminated with a dominant waveguide mode, the theoretically obtainable off-axis radiation levels⁴ in both the transverse planes for transverse polarization are considerably higher than those obtained for longitudinal polarization in the transverse plane and transverse polarization in the longitudinal plane. For transverse polarization in the transverse plane, one has essentially the equivalent of an aperture with constant illumination across it. Hence, with the antenna mounted on its normal vertical position (i.e., the axis of the horn normal to the earth), the amplitude distributions across the aperture of the antenna are as follows: For longitudinal polarization, the electric field in the longitudinal direction is uniformly distributed across the aperture, and in the transverse direction the field is tapered to a low value at the edges; with transverse polarization, the electric field is uniform across the aperture in the transverse direction and is tapered to a low value at the edges in the longitudinal direction.

Of the many various possible contributors to high sidelobes, the two strongest contenders are: (i) higher-order modes being generated in the existing feedhorn, which was designed to accommodate the oversized circular waveguide† used in the system, and (ii) surface toler-

* As used in the microwave radio relay system, the horn-reflector antenna is mounted with the axis of the horn normal to the earth's surface. Hence, longitudinal and transverse polarizations could be called vertical and horizontal, respectively. However, the aperture field distribution for each polarization is different, and when the antenna is used as an earth station antenna for satellite communications, or as a radiometer, or simply to obtain radiation patterns in the longitudinal plane (the antenna is mounted on its side), aperture field distributions for so-called vertical and horizontal are now interchanged. To avoid this ambiguity, longitudinal and transverse polarizations which are referred to the axis of the horn are used.

† Higher-order modes can be excited in oversized waveguide; hence to avoid this condition when obtaining radiation patterns, one uses a transition from feedhorn to waveguide transducer.

ances of the parabolic reflector. To examine the first contender, two feedhorns were made, one a scaled version of the existing system feedhorn and the second a new design. This new feedhorn terminates the taper of the horn into the walls of rectangular K-band waveguide, thus ensuring the presence of only the dominant mode. In a sense, this is achieved by opening the parallel walls of the waveguide at a constant angle and extending them until they form the square aperture needed to mate with the rest of the antenna. This, of course, requires that the walls normal to the electric field open first. Carefully machined mandrels were made of both feeds over which copper was electroformed to produce the final models. An examination of the radiation characteristics of the scale-model antenna obtained from using both feeds indicated that the characteristics were essentially identical. Hence, the characteristics contained in the figures to be discussed were obtained using the scale version of the system feedhorn. In the figures, the theoretical values⁴ are shown by a dashed line, and the measured data of the full-sized horn-reflector antenna by a broken line. The measured data for the scale-model are shown by the solid line.

For the transverse plane, Figs. 1 and 2 show the measured characteristics for longitudinal and transverse polarizations, respectively. As indicated here, the agreement between scale-model measurements and theory is remarkably good. From these two figures, one can see the improvement in sidelobe level by comparing the results with the broken line curve which indicates the measured response for the full-sized antenna. For example, in Fig. 1 the improvement is about 8 dB and increasing in the region of the fourth sidelobe and beyond, i.e., beyond 8 degrees from the on-axis position. In Fig. 2, a similar comparison shows a 4-dB improvement beginning around 10 degrees.

In the longitudinal plane, the measurements for longitudinal and transverse polarizations were obtained for the scale-mode antenna but are not included here since these patterns, like those of the full-size antenna, are in very good agreement with theory. Similarly, the cross-polarization patterns obtained on the scale model were like those of the full-sized antenna and, since both agree well with theory, are also not included here.

It is evident from these figures that the scale-model measurements are in remarkable agreement with theory. The presumption that the high sidelobe levels of the full-size antenna were due in part to high-order modes being generated by the feedhorn is not valid. This is emphatically demonstrated by the agreement with theory which does not accommodate the presence of higher-order modes. The remaining strong contender for these high sidelobe levels is the accuracy of the full-size parabolic reflector surface. An examination of the full-size structure shows the surface to consist of two half skins fastened to parabolic back ribs by means of 262 (No. 14, Type B) sheet metal

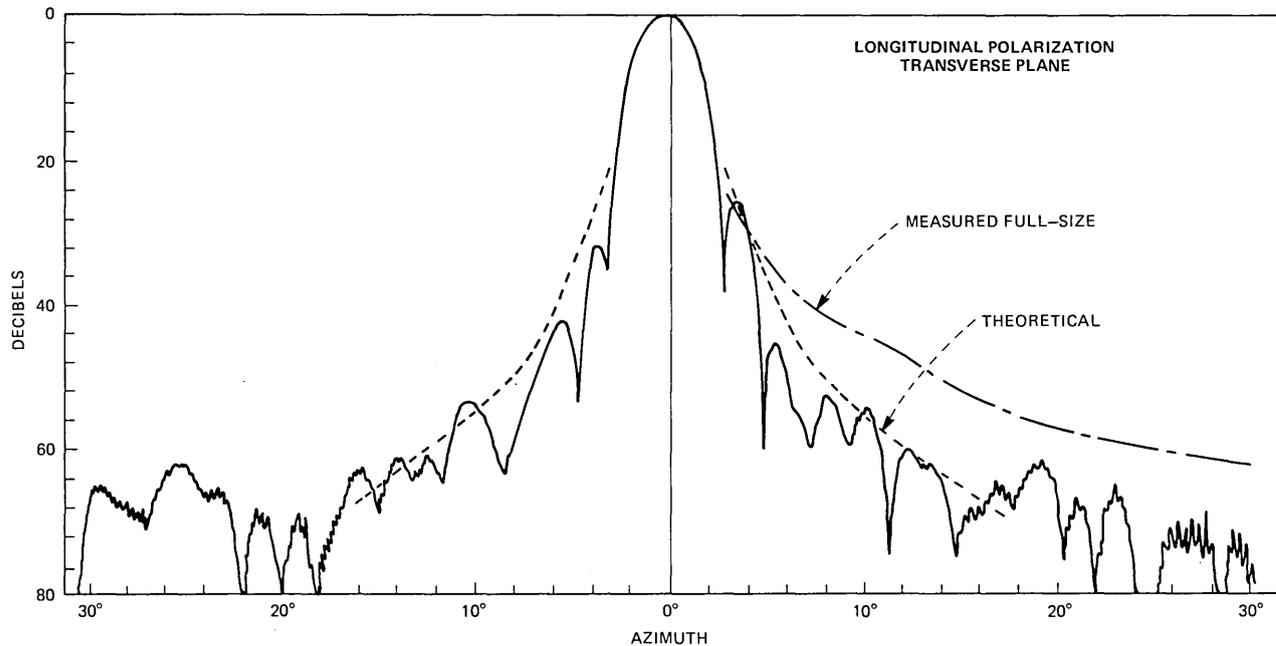


Fig. 1—Radiation pattern in the transverse plane for longitudinal polarization. The dashed line is theoretical, and the broken line represents measurements obtained on the full-size antenna.

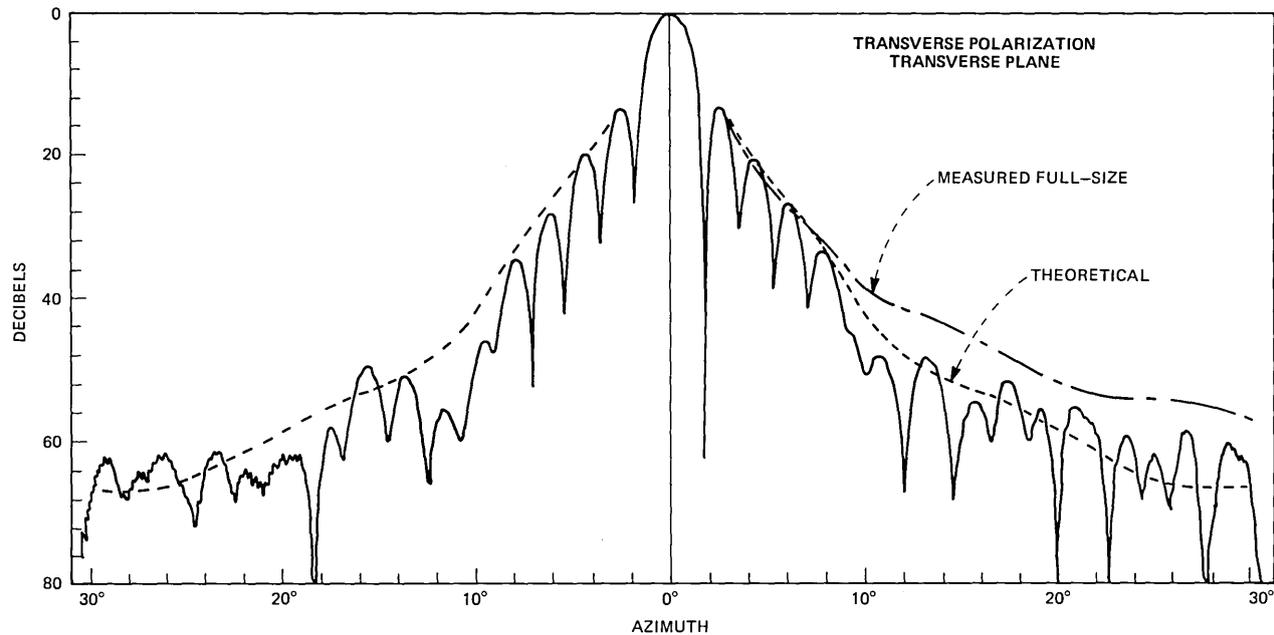


Fig. 2—Radiation pattern in the transverse plane for transverse polarization. The dashed line is theoretical, and the broken line represents measurements obtained on the full-size antenna.

screws and 120 (3/16-inch diameter Universal Head) rivets. Since these fasteners are distributed in a regular fashion, one could envision them as forming an array of scatterers which produce a low level, very broad beam and enhance the far sidelobe region. An investigation of this remaining aspect is not considered in the immediate future since it would entail a reproduction of these discontinuities on the precise surface of the scale-model. Also, there are many interesting experiments waiting to be done on this scale model which can now serve as a test bed.

III. CONCLUSIONS

Data obtained on a scale-model version of the full-sized horn-reflector (used in the 4-, 6-, and 11-GHz microwave common-carrier, radio relay system) have been presented. From the data, one once again observes how well scale models depict the full-size world. But of most importance was the achievement of agreement between measurement and theory which produced the key finding that past disagreement between theoretical results and experimental data for a full-size antenna in the sidelobe region could not be due primarily to higher order modes being excited in the feedhorn, but could be attributed to surface imperfections of the reflector.

IV. ACKNOWLEDGMENTS

I would like to thank N. F. Schlaack, now retired, of these Laboratories, for sharing his measurements on the full-sized horn-reflector antenna with me. It is with appreciation that I acknowledge A. Quigley for providing the two scaled feedhorns, and the team of P. J. Schoellner and J. A. Swig, Jr., for their shepherding of the fabrication of the scale-model parabolic reflector. The successful results obtained are due to the attention these three gave to their respective parts.

REFERENCES

1. H. T. Friis and A. C. Beck, U. S. Patent 2,236,393.
2. R. W. Friis and A. S. May, "A New Broad-Band Microwave Antenna System," AIEE Trans., Part I, 77, 1958, p. 97.
3. D. T. Thomas, "Analysis and Design of Elementary Blinders for Large Horn Reflector Antennas," B.S.T.J., 50, No. 8 (November 1971), p. 2979.
4. A. B. Crawford, D. C. Hogg, and L. E. Hunt, "A Horn-Reflector Antenna for Space Communications," B.S.T.J., 40, No. 4 (July 1961), p. 1095.

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, Computer & Control Abstracts, Current Contents/Engineering, Technology & Applied Sciences, Current Contents/Physical & Chemical Sciences, Current Index to Statistics, Current Papers in Electrical & Electronic Engineering, Current Papers on Computers & Control, Electrical & Electronic Abstracts, Electronics & Communications Abstracts Journal, The Engineering Index, International Aerospace Abstracts, Journal of Current Laser Abstracts, Language and Language Behavior Abstracts, Mathematical Reviews, Metals Abstracts, Science Abstracts, Science Citation Index, 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