

September 1984
Vol. 63 No. 7

AT&T
BELL LABORATORIES

TECHNICAL JOURNAL

A JOURNAL OF THE AT&T COMPANIES

Speech Recognition

Traffic

Sticky Storage Devices

Short-Channel MOSFETs

EDITORIAL COMMITTEE

A. A. PENZIAS,¹ *Committee Chairman*

M. M. BUCHNER, JR.¹

R. C. FLETCHER¹

J. S. NOWAK¹

R. P. CLAGETT²

D. HIRSCH⁴

B. B. OLIVER⁵

R. P. CREAM²

S. HORING¹

J. W. TIMKO³

B. R. DARNALL¹

R. A. KELLEY¹

V. A. VYSSOTSKY¹

B. P. DONOHUE, III³

J. F. MARTIN²

¹AT&T Bell Laboratories ²AT&T Technologies ³AT&T Information Systems

⁴AT&T Consumer Products ⁵AT&T Communications

EDITORIAL STAFF

B. G. KING, *Editor*

L. S. GOLLER, *Assistant Editor*

P. WHEELER, *Managing Editor*

A. M. SHARTS, *Assistant Editor*

B. G. GRUBER, *Circulation*

AT&T BELL LABORATORIES TECHNICAL JOURNAL (ISSN0005-8580) is published ten times each year by AT&T, 550 Madison Avenue, New York, NY 10022; C. L. Brown, Chairman of the Board; T. O. Davis, Secretary. The Computing Science and Systems section and the special issues are included as they become available. Subscriptions: United States—1 year \$35; 2 years \$63; 3 years \$84; foreign—1 year \$45; 2 years \$73; 3 years \$94. A subscription to the Computing Science and Systems section only is \$10 (\$12 foreign). Single copies of most issues of the Journal are available at \$10 (\$11 foreign). Payment for foreign subscriptions or single copies must be made in United States funds, or by check drawn on a United States bank and made payable to the Technical Journal and sent to AT&T Bell Laboratories, Circulation Dept., Room 1E335, 101 J. F. Kennedy Pky, Short Hills, NJ 07078.

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

Comments on the technical content of any article or brief are welcome. These and other editorial inquiries should be addressed to the Editor, AT&T Bell Laboratories Technical Journal, Room 1H321, 101 J. F. Kennedy Pky, Short Hills, NJ 07078. Comments and inquiries, whether or not published, shall not be regarded as confidential or otherwise restricted in use and will become the property of AT&T. Comments selected for publication may be edited for brevity, subject to author approval.

Printed in U.S.A. Second-class postage paid at Short Hills, NJ 07078 and additional mailing offices. Postmaster: Send address changes to the AT&T Bell Laboratories Technical Journal, Room 1E335, 101 J. F. Kennedy Pky, Short Hills, NJ 07078.

Copyright © 1984 AT&T.

AT&T Bell Laboratories

Technical Journal

VOL. 63

SEPTEMBER 1984

NO. 7

Copyright © 1984 AT&T. Printed in U.S.A.

On Hidden Markov Model and Dynamic Time Warping for Speech Recognition—A Unified View	1213
B.-H. Juang	
On the Performance of Isolated Word Speech Recognizers Using Vector Quantization and Temporal Energy Contours	1245
L. R. Rabiner, K. C. Pan, and F. K. Soong	
The Overload Performance of Engineered Networks With Nonhierarchical and Hierarchical Routing	1261
J. M. Akinpelu	
Methods in Traffic Calculations	1283
D. L. Jagerman	
On the Capacity of Sticky Storage Devices	1311
H. S. Witsenhausen	
Velocity-Saturated Characteristics of Short-Channel MOSFETs	1325
G. W. Taylor	
LETTERS	1405
PAPERS BY AT&T BELL LABORATORIES AUTHORS	1411
ERRATUM	1416
CONTENTS, OCTOBER ISSUE	1417

On the Hidden Markov Model and Dynamic Time Warping for Speech Recognition—A Unified View

By B.-H. JUANG*

(Manuscript received January 17, 1984)

This paper gives a unified theoretical view of the Dynamic Time Warping (DTW) and the Hidden Markov Model (HMM) techniques for speech recognition problems. The application of hidden Markov models in speech recognition is discussed. We show that the conventional dynamic time-warping algorithm with Linear Predictive (LP) signal modeling and distortion measurements can be formulated in a strictly statistical framework. It is further shown that the DTW/LP method is implicitly associated with a specific class of Markov models and is equivalent to the probability maximization procedures for Gaussian autoregressive multivariate probabilistic functions of the underlying Markov model. This unified view offers insights into the effectiveness of the probabilistic models in speech recognition applications.

I. INTRODUCTION

Research in speech recognition has produced numerous algorithms and commercially available speech recognizers that all work to some extent.¹ Among these, temporal alignment techniques such as Dynamic Time-Warping (DTW) algorithms²⁻⁴ and Markov modeling⁵⁻⁷ are two prevailing approaches that are practical and theoretically sound. Both techniques emphatically address nonstationarity in speech signals. The two techniques, however, operate in different manners, as we will discuss briefly.

* AT&T Bell Laboratories.

Copyright © 1984 AT&T. Photo reproduction for noncommercial use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

In a recognition system employing dynamic time warping, a warping procedure based upon a prechosen, well-defined distortion measure aligns the unknown test speech sequence in turn to each reference sequence. The distortion measure must be a meaningful metric of dissimilarity between sound representations, usually the short-time spectra. The objective is to find a reference sequence of a known category or word that has the least dissimilarity to the test sequence after being optimally time aligned. Time alignment involves (time-) warping functions that are dynamic but deterministic representations of the possible variation of sound durations that are evident between the reference and the test sequences. There is, hence, one warping function that best matches the reference and test sequences, resulting in the smallest dissimilarity. The smallest dissimilarity measurement among all categories determines the recognition (recognition by minimum distortion).

Markov modeling techniques, while they may still perform sound pattern comparisons, do not require explicit time alignment. Instead, a probabilistic transition and observation structure is defined for each reference category or word. Such a structure, called a Markov model, includes (1) a state transition probability matrix, (2) an initial probability vector, and (3) an observation probability matrix for discrete probability densities or a set of continuous densities defined by parameter sets, or a mixture of the two when different types of densities are used. During recognition, one computes for each given reference model the probability of observing the test sequence. The model that produces the maximum observation probability is the classification result (recognition by maximum probability).

These two techniques are similar in theory, despite their vastly different operations and results. Confusion from this similarity often renders comparative studies of the two techniques difficult and futile. The purpose of this paper is, then, to give a unified tutorial view of the two techniques and to establish a theoretical link between them such that more fruitful and meaningful comparison can be made and each technique will improve the other technique.

The paper is organized as follows. We first present statistical characteristics of Gaussian autoregressive sources in Section II, which serves as a foundation for the later developments. This topic is well studied in multivariate analysis, and an excellent treatment of it in the context of speech processing can be found in Ref. 8. In Section III we discuss maximum likelihood estimation of Gaussian autoregressive source parameters, and we explicitly demonstrate the relationship between some well-known probability density functions and distortion measures related to linear prediction (LP). In Section IV we discuss some fundamentals of probabilistic functions of Markov chains and

their applications in speech recognition. We again show that distortion measures can be cast in the framework of probabilistic functions of Markov chains. We finally discuss dynamic time warping in Section V and show that dynamic time warping employing LP-related measures is equivalent to the recognition by maximum probability procedure, with some specific constraints. Theoretical similarities and differences between the two techniques are then discussed in detail to complete the attempted unified view. We start with Gaussian autoregressive source because it is one of the best known sources and is useful in speech research. More general measure-theoretic steps could have been taken to establish a formal theoretical link between the two methods. It is, nevertheless, our opinion that the present framework of Gaussian autoregressive sources adequately gives a meaningful unified view of the two methods.

II. GAUSSIAN AUTOREGRESSIVE SOURCE

Consider a stationary, zero-mean, Gaussian signal source. The output of the source, subject to observation, is an N -sampled sequence $\{s_1, s_2, \dots, s_N\}$, where each s_i is a real random variable. The vector notation $\mathbf{s}^t = [s_1 \ s_2 \ \dots \ s_N] \in R^N$ denotes the observation. The probability density function of the random vector \mathbf{s} for *known* autocorrelation matrix \mathbf{C}_N is thus

$$f(\mathbf{s} | \mathbf{C}_N) = \lim_{\substack{\Delta s_i \rightarrow 0 \\ i=1,2,\dots,N}} \frac{\text{Pr}\{s_1 \leq \mathbf{s}_1 \leq s_1 + \Delta s_1, s_2 \leq \mathbf{s}_2 \leq s_2 + \Delta s_2, \dots, s_N \leq \mathbf{s}_N \leq s_N + \Delta s_N | \mathbf{C}_N\}}{\Delta s_1 \Delta s_2 \dots \Delta s_N} \\ = (2\pi)^{-N/2} |\mathbf{C}_N|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{s}^t \mathbf{C}_N^{-1} \mathbf{s} \right\}, \quad (1)$$

where $\mathbf{s}^t = [s_1 \ s_2 \ \dots \ s_N]$ is a realization of \mathbf{s}^t , $\mathbf{C}_N = [r_{ij}]_{i,j=1}^N$, and $r_{ij} = E\{s_i s_j\} = r_{|i-j|}$ due to stationarity.

The source is assumed to be M th-order autoregressive with coefficients $\mathbf{a}^t = [a_0 \ a_1 \ \dots \ a_M]$, where a_0 is always unity. Hence, as shown in Fig. 1, the source can be equivalently viewed as a white Gaussian noise source with unity variance, followed by an all-pole filter $1/A(z)$, where

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_M z^{-M},$$

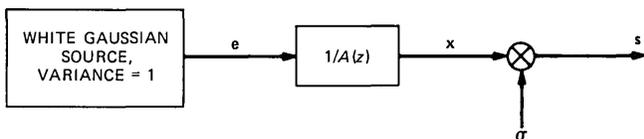


Fig. 1—Gaussian autoregressive source.

$$= \left[\begin{array}{c|c} \mathbf{H}_1 & 0 \\ \hline \mathbf{H}_2 & \mathbf{H}_3 \end{array} \right] \begin{array}{l} \} M \\ \} N - M \end{array}$$

$\underbrace{\hspace{1.5cm}}_M \quad \underbrace{\hspace{1.5cm}}_{N-M}$

The elements of \mathbf{e} are uncorrelated, and h_{ij} are properly scaled so that $E[\mathbf{e}_i \mathbf{e}_j] = E[\epsilon_i \epsilon_j] = E[\epsilon_i \epsilon_j] = \delta_{ij}$ for any appropriate i and j , thereby giving

$$\begin{aligned} \mathbf{I}_N &= E\{\mathbf{e}\mathbf{e}^t\} \\ &= (\sigma^2)^{-1} \mathbf{H} E\{\mathbf{s}\mathbf{s}^t\} \mathbf{H}^t \\ &= (\sigma^2)^{-1} \mathbf{H} \mathbf{C}_N \mathbf{H}^t, \end{aligned} \quad (4)$$

where \mathbf{I}_N is the $N \times N$ identity matrix. Equation (4) leads to

$$\mathbf{C}_N^{-1} = (\sigma^2)^{-1} \mathbf{H}^t \mathbf{H} \quad (5)$$

and

$$|\mathbf{C}_N| = |\mathbf{C}_N^{-1}|^{-1} = (\sigma^2)^N |\mathbf{H}|^{-2}. \quad (6)$$

But, since $|\mathbf{H}_3| = 1$ and $|\mathbf{H}| = |\mathbf{H}_1| \cdot |\mathbf{H}_3|$,

$$|\mathbf{C}_N| = (\sigma^2)^N |\mathbf{H}_1|^{-2}.$$

Note that matrix \mathbf{H}_1 also corresponds to the diagonalization of the $M \times M$ autocorrelation matrix $\mathbf{C}_M = [r_{ij}]$ for $i, j = 1, 2, \dots, M$; i.e.,

$$\mathbf{I}_M = (\sigma^2)^{-1} \mathbf{H}_1 \mathbf{C}_M \mathbf{H}_1^t,$$

where \mathbf{I}_M is the $M \times M$ identity matrix. Therefore,

$$|\mathbf{C}_M| = (\sigma^2)^M |\mathbf{H}_1|^{-2}$$

and

$$|\mathbf{C}_N| = (\sigma^2)^{N-M} |\mathbf{C}_M|. \quad (7)$$

Given an autocorrelation matrix \mathbf{C}_M , $|\mathbf{C}_M|$ can be easily obtained by first diagonalizing \mathbf{C}_M using Cholesky decomposition or, more efficiently, Levinson's recursion algorithm,⁹

$$\mathbf{B}^t \mathbf{C}_M \mathbf{B} = \beta = \begin{bmatrix} \beta_0 & & & & 0 \\ & \beta_1 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & & \beta_{M-1} \end{bmatrix},$$

where \mathbf{B} is an upper triangular matrix, the diagonal elements of which are all unity. Therefore,

$$|\mathbf{C}_M| = \prod_{i=0}^{M-1} \beta_i,$$

and as a result,

$$|\mathbf{C}_N| = (\sigma^2)^{N-M} \left(\prod_{i=0}^{M-1} \beta_i \right). \quad (8)$$

Note that β_i is equivalent to the minimum mean-square error resulting from an i th-order linear prediction of the signal. The probability density function of \mathbf{s} becomes

$$f(\mathbf{s} | \mathbf{C}_N) = (2\pi)^{-N/2} (\sigma^2)^{-(N-M)/2} \left(\prod_{i=0}^{M-1} \beta_i \right)^{-1/2} \exp\{-s^t \mathbf{H}^t \mathbf{H} \mathbf{s} / 2\sigma^2\}. \quad (9)$$

For the gain-independent expression,

$$f(x | \mathbf{C}_N) = (2\pi)^{-N/2} \left(\prod_{i=0}^{M-1} \frac{\beta_i}{\sigma^2} \right)^{-1/2} \exp\left\{-\frac{1}{2} x^t \mathbf{H}^t \mathbf{H} x\right\}. \quad (10)$$

We further write $x^t \mathbf{H}^t \mathbf{H} x$ explicitly as

$$\begin{aligned} x^t \mathbf{H}^t \mathbf{H} x = & \left(\sum_{j=0}^M a_j^2 \right) \left(\sum_{i=1}^N x_i^2 \right) + 2 \left(\sum_{j=0}^{M-1} a_j a_{j+1} \right) \left(\sum_{i=1}^{N-1} x_i x_{i+1} \right) + \dots \\ & + 2(a_0 a_M) \left(\sum_{i=1}^{N-M} x_i x_{i+M} \right) - Q, \end{aligned}$$

where Q represents negligible terms compared to others for $N \gg M$. Letting

$$r_a(i) \triangleq \sum_{j=1}^{M-i} a_j a_{j+i}, \quad (11)$$

$$r_x(i) \triangleq \sum_{j=1}^{N-i} x_j x_{j+i} = \frac{1}{\sigma^2} \sum_{j=1}^{N-i} s_j s_{j+i} = \frac{r_s(i)}{\sigma^2}, \quad (12)$$

and

$$\alpha(x; \mathbf{a}) = r_a(0)r_x(0) + 2 \sum_{i=1}^M r_a(i)r_x(i), \quad (13)$$

we then have an approximation for the density function,

$$f(\mathbf{s} | \mathbf{C}_N) \cong (2\pi)^{-N/2} (\sigma^2)^{-(N-M)/2} \left(\prod_{i=0}^{M-1} \beta_i \right)^{-1/2} \exp\left\{-\frac{1}{2} \alpha(\sigma^{-1} \mathbf{s}; \mathbf{a})\right\} \quad (14)$$

or

$$f(x | \mathbf{C}_N) \cong (2\pi)^{-N/2} \left(\prod_{i=0}^{M-1} \frac{\beta_i}{\sigma^2} \right)^{-1/2} \exp\left\{-\frac{1}{2} \alpha(x; \mathbf{a})\right\}. \quad (15)$$

This function can be evaluated easily if the source parameters are known.

III. MAXIMUM LIKELIHOOD

In many realistic situations, such as dealing with speech signals, the a priori information about the source is usually not available. What is involved in parameterization of speech signals for coding and recognition is mainly estimating and identifying the source parameters from *finite observations*. More specifically, the following two dominant problems arise almost ubiquitously in speech analysis research: (1) Estimation—Given an observation s , what is the best or the most probable set of source parameters that led to the observation? (2) Identification—Given two observations s_1 and s_2 , how close are the two observations? Are they close enough to be considered identical? Or, what is the probability that s_2 has been produced by the same source as s_1 ? The first problem is certainly very much studied in statistical estimation theory as well as in (deterministic) least-squares time-series analysis. Research in identification, particularly in the field of speech processing, resulted in some distance or distortion measures.¹⁰ Thus our main goal here is to integrate the formulation of the two problems in a probabilistic framework to better understand the probabilistic modeling techniques. In the following presentation, we shall focus on the maximum likelihood estimate of Gaussian autoregressive source parameters, as initiated in the previous section.

3.1 Estimation—Autocorrelation method

We discuss here only the autocorrelation method. For other varieties, we suggest that readers consult Refs. 9 and 11.

The observation sequence $s = \{s_1, s_2, \dots, s_N\}$ is assumed to be very long, i.e., $N \gg M$. It is further argued that observation of the source output, which is infinitely long, is made through some “smooth window,” so that the edge problem at the beginning of the observed sequence is avoided, maintaining that

$$\mathbf{e}_i = \sum_{j=0}^M a_j \mathbf{x}_{i-j} \quad \text{and} \quad \mathbf{s}_i = \sigma \mathbf{x}_i \quad \text{for all } i, \quad (16)$$

with $a_0 = 1$ and $\mathbf{x}_i = 0$ for $i \leq 0$ and for $i > N$. Hence, the diagonal elements in \mathbf{H} matrix are assumed to be all unity. Equation (6) then becomes

$$|\mathbf{C}_N| = (\sigma^2)^N,$$

and the probability density function, as in eq. (14), is now expressed as

$$f(s | \mathbf{C}_N) \simeq (2\pi)^{-N/2} (\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2} \alpha(\sigma^{-1} \mathbf{s}; \mathbf{a}) \right\}. \quad (17)$$

Since (17) is a function of \mathbf{a} and σ^2 and it closely approximates $f(s | \mathbf{C}_N)$, we shall, in the following, define $f(s | \mathbf{a}, \sigma^2)$ as (17). Furthermore, the gain-independent density function is thus

$$f(x | \mathbf{a}) = (2\pi)^{-N/2} \exp \left\{ -\frac{1}{2} \alpha(x; \mathbf{a}) \right\}. \quad (18)$$

It defines the probability density function of observing a vector x at the output of an all-pole filter $1/A(z)$ driven by a *unity variance* Gaussian i.i.d. sequence.

It is clear that, given an observation $s^{(0)}$, the maximum likelihood estimate of \mathbf{a} is the one that maximizes $f(s^{(0)} | \mathbf{a}, \sigma^2)$ or, equivalently, minimizes $\alpha(s^{(0)}; \mathbf{a})$ because σ here is only a scaling factor. In linear prediction terminology, the optimal $\mathbf{a}^{(0)}$ is obtained by minimizing the prediction-error energy $\alpha(s^{(0)}; \mathbf{a})$, defined by (13), and $\alpha(s^{(0)}; \mathbf{a}^{(0)}) = \min_{\mathbf{a}} \alpha(s^{(0)}; \mathbf{a})$ is called the minimum residual energy.⁹ Furthermore, to maximize $f(s^{(0)} | \mathbf{a}^{(0)}, \sigma^2)$ with respect to σ^2 , the optimal estimate $\sigma_{(0)}^2$, is easily found to be

$$\sigma_{(0)}^2 = \alpha(s^{(0)}; \mathbf{a}^{(0)})/N, \quad (19)$$

which leads to

$$\alpha(\sigma_{(0)}^{-1} s^{(0)}; \mathbf{a}^{(0)}) = N. \quad (20)$$

This can be verified intuitively by recognizing that there are N valid uncorrelated error samples, $\sigma_{(0)} \mathbf{e}_i = \sum_{j=0}^M a_j^{(0)} \mathbf{s}_{i-j}^{(0)}$, $i = 1, 2, \dots, N$, each having variance $\sigma_{(0)}^2$, resulting in an energy of $\alpha(s^{(0)}; \mathbf{a}^{(0)}) = N \sigma_{(0)}^2$.

3.2 Identification

In this section we establish the relationship between probability and distortion measures for identification purposes. In particular, for the present consideration of Gaussian autoregressive sources, we discuss the role of Itakura-Saito measure and the likelihood ratio measure¹² in probability density functions.

3.2.1 Itakura-Saito measure

In maximum likelihood estimation, we often maximize the log likelihood, $\log \{f(s^{(0)} | \mathbf{a}, \sigma^2)\}$, instead of $f(s^{(0)} | \mathbf{a}, \sigma^2)$ for convenience, particularly when the probability density function is jointly Gaussian, as in the present case. The log likelihood takes the form

$$\log \{f(s | \mathbf{a}, \sigma^2)\} = -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2} \alpha(\sigma^{-1} \mathbf{s}; \mathbf{a}). \quad (21)$$

Since $\mathbf{a}^{(0)}$ and $\sigma_{(0)}^2$ are the maximum likelihood estimate based upon $s^{(0)}$,

$$\begin{aligned}\log\{f(s^{(0)} | \mathbf{a}^{(0)}, \sigma_{(0)}^2)\} &= -\frac{N}{2} \log(2\pi\sigma_{(0)}^2) - \frac{1}{2} \alpha(\sigma_{(0)}^{-1}s^{(0)}; \mathbf{a}^{(0)}) \\ &= -\frac{N}{2} \log(2\pi\sigma_{(0)}^2) - \frac{N}{2} \\ &= \{\log f(s^{(0)} | \mathbf{a}, \sigma^2)\}_{\max}.\end{aligned}\quad (22)$$

The log likelihood difference between the maximum and other arbitrary values is thus

$$\begin{aligned}L_d &= \{\log f(s^{(0)} | \mathbf{a}, \sigma^2)\}_{\max} - \log f(s^{(0)} | \mathbf{a}, \sigma^2) \\ &= \log f(s^{(0)} | \mathbf{a}^{(0)}, \sigma_{(0)}^2) - \log f(s^{(0)} | \mathbf{a}, \sigma^2) \\ &= -\frac{N}{2} \log(2\pi\sigma_{(0)}^2) - \frac{N}{2} + \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2} \alpha(\sigma^{-1}s^{(0)}; \mathbf{a}) \\ &= \frac{N}{2} \left[\frac{1}{N} \alpha(\sigma^{-1}s^{(0)}; \mathbf{a}) + \log \sigma^2 - \log \sigma_{(0)}^2 - 1 \right]\end{aligned}\quad (23)$$

because

$$f(s^{(0)} | \mathbf{a}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2} \alpha(\sigma^{-1}s^{(0)}; \mathbf{a}) \right\}.\quad (24)$$

The bracketed term in (23) is, in fact, the well-known Itakura-Saito distortion measure¹⁰ between $\{\mathbf{a}^{(0)}, \sigma_{(0)}^2\}$, representing $s^{(0)}$, and $\{\mathbf{a}, \sigma^2\}$, representing another observation s ; i.e.,

$$\begin{aligned}d_{IS}(s^{(0)}; s) &= d_{IS}(s^{(0)}; \{\mathbf{a}, \sigma^2\}) = \frac{1}{N} \alpha(\sigma^{-1}s^{(0)}; \mathbf{a}) \\ &\quad + \log \sigma^2 - \log \sigma_{(0)}^2 - 1 \\ &= 2L_d/N.\end{aligned}\quad (25)$$

Therefore, the probability of observing $s^{(0)}$ at the output of a source with parameters $\{\mathbf{a}, \sigma^2\}$ is, in terms of the distortion measure,

$$\begin{aligned}f(s^{(0)} | \mathbf{a}, \sigma^2) &= (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{N}{2} [d_{IS}(s^{(0)}; \{\mathbf{a}, \sigma^2\}) \right. \\ &\quad \left. + \log \sigma_{(0)}^2 - \log \sigma^2 + 1] \right\} \\ &= G(\sigma^2, \sigma_{(0)}^2) \exp \left\{ -\frac{N}{2} d_{IS}(s^{(0)}; \{\mathbf{a}, \sigma^2\}) \right\},\end{aligned}\quad (26)$$

where

$$G(\sigma^2, \sigma_{(0)}^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{N}{2} [\log \sigma_{(0)}^2 - \log \sigma^2 + 1] \right\}. \quad (27)$$

3.2.2 Likelihood ratio measure

In many situations, the desired identification is only based upon the autoregressive parameters \mathbf{a} . This is equivalent to comparing two gain-normalized observation sequences. Let $\{\mathbf{a}^{(0)}, \sigma_{(0)}^2\}$ and $\{\mathbf{a}, \sigma^2\}$ be the maximum likelihood estimates corresponding to observations $s^{(0)}$ and s , respectively. The gain-normalized observations are then $x^{(0)} = s^{(0)}/\sigma_{(0)}$ and $x = s/\sigma$. From (18) and (20), the maximum log likelihood for $x^{(0)}$ is

$$\begin{aligned} \{\log f(x^{(0)} | \mathbf{a})\}_{\max} &= \log f(x^{(0)} | \mathbf{a}^{(0)}) \\ &= -\frac{N}{2} \log(2\pi) - \frac{1}{2} \alpha(x^{(0)}; \mathbf{a}^{(0)}) \\ &= -\frac{N}{2} \log(2\pi) - \frac{N}{2}. \end{aligned}$$

But, since

$$f(x^{(0)} | \mathbf{a}) = (2\pi)^{-N/2} \exp \left\{ -\frac{1}{2} \alpha(x^{(0)}; \mathbf{a}) \right\}, \quad (28)$$

the log likelihood difference between the maximum and another arbitrary value is then

$$\begin{aligned} L_d &= \{\log f(x^{(0)} | \mathbf{a})\}_{\max} - \log f(x^{(0)} | \mathbf{a}) \\ &= -\frac{N}{2} \log(2\pi) - \frac{N}{2} + \frac{N}{2} \log(2\pi) + \frac{1}{2} \alpha(x^{(0)}; \mathbf{a}) \\ &= \frac{N}{2} \left[\frac{1}{N} \alpha(x^{(0)}; \mathbf{a}) - 1 \right]. \end{aligned} \quad (29)$$

The above-bracketed term is the likelihood ratio measure widely employed in vector quantization vocoder designs; that is,

$$\begin{aligned} d_{LR}(s^{(0)}; s) &\triangleq d_{LR}(x^{(0)}; x) \\ &= d_{LR}(x^{(0)}; \mathbf{a}) \\ &= \frac{1}{N} \alpha(x^{(0)}; \mathbf{a}) - 1. \end{aligned} \quad (30)$$

The likelihood ratio measure and the Itakura-Saito measure are closely related,

$$\begin{aligned}
 d_{LR}(s^{(0)}; s) &= d_{LR}(x^{(0)}; x) \\
 &= d_{LR}(x^{(0)}; \mathbf{a}) \\
 &= d_{IS}(x^{(0)}; x) \\
 &= d_{IS}(x^{(0)}; \mathbf{a}).
 \end{aligned}$$

As a result, the probability of observing $x^{(0)}$ at the output of an all-pole filter $1/A(z)$ driven by a unity variance Gaussian i.i.d. sequence is

$$\begin{aligned}
 f(x^{(0)} | \mathbf{a}) &= (2\pi)^{-N/2} \exp \left\{ -\frac{N}{2} [d_{LR}(x^{(0)}; \mathbf{a}) + 1] \right\} \\
 &= P \exp \left\{ -\frac{N}{2} d_{LR}(x^{(0)}; \mathbf{a}) \right\}
 \end{aligned} \tag{31}$$

where

$$P = (2\pi)^{-N/2} \exp \left\{ -\frac{N}{2} \right\}. \tag{32}$$

Equations (26) and (31) are thus the fundamental link between probability and distortion measures.

3.3 Estimation and identification based upon multiple observations

We have presented parametric estimation of the probability density function based upon a single observation in the above. When several observations are available and known to be from the same source, the estimation turns out to be quite similar to the single-observation case.

Let $s^{(i)}$, $i = 1, 2, \dots, L$, denote the available observations. These observations are considered to be i.i.d. with probability density

$$f(s^{(i)} | \mathbf{a}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2} \alpha(\sigma^{-1}s^{(i)}; \mathbf{a}) \right\}.$$

The joint probability density of observations $s^{(1)}, s^{(2)}, \dots, s^{(L)}$ is thus

$$\begin{aligned}
 &f(s^{(1)}, s^{(2)}, \dots, s^{(L)} | \mathbf{a}, \sigma^2) \\
 &= \prod_{i=1}^L f(s^{(i)} | \mathbf{a}, \sigma^2) \\
 &= [(2\pi\sigma^2)^{-N/2}]^L \exp \left\{ -\frac{1}{2} \sum_{i=1}^L \alpha(\sigma^{-1}s^{(i)}; \mathbf{a}) \right\}.
 \end{aligned} \tag{33}$$

As with the single-observation case, the maximum likelihood estimate requires minimization of $\sum_{i=1}^L \alpha(\sigma^{-1}s^{(i)}; \mathbf{a})$. But maximizing

$f(s^{(1)}, s^{(2)}, \dots, s^{(L)} | \mathbf{a}, \sigma^2)$ is, in our interest here, equivalent to maximizing $[f(s^{(1)}, s^{(2)}, \dots, s^{(L)} | \mathbf{a}, \sigma^2)]^{1/L}$. Since

$$[f(s^{(1)}, s^{(2)}, \dots, s^{(L)} | \mathbf{a}, \sigma^2)]^{1/L} = (2\pi\sigma^2)^{-N/2} \exp \left\{ -\frac{1}{2} \left[\frac{1}{L} \sum_{i=1}^L \alpha(\sigma^{-1}s^{(i)}; \mathbf{a}) \right] \right\}, \quad (34)$$

the similarity between multiple- and single-observation estimation can easily be seen by comparing (17) and (34). From (13),

$$\begin{aligned} & \frac{1}{L} \sum_{i=1}^L \alpha(\sigma^{-1}s^{(i)}; \mathbf{a}) \\ &= \frac{1}{\sigma^2} \frac{1}{L} \sum_{i=1}^L \left\{ r_a(0)r_s^{(i)}(0) + 2 \sum_{j=1}^M r_a(j)r_s^{(i)}(j) \right\} \\ &= \frac{1}{\sigma^2} \left\{ r_a(0) \left[\frac{1}{L} \sum_{i=1}^L r_s^{(i)}(0) \right] + 2 \sum_{j=1}^M r_a(j) \left[\frac{1}{L} \sum_{i=1}^L r_s^{(i)}(j) \right] \right\}, \quad (35) \end{aligned}$$

where

$$r_s^{(i)}(j) = \sum_{n=1}^{N-j} s_n^{(i)} s_{n+j}^{(i)},$$

and $s_n^{(i)}$ is simply the n th sample in $s^{(i)}$. Equations (34) and (35) clearly demonstrate that the maximum likelihood estimate of the source parameters, given multiple observations, can be obtained by the same minimization procedure as in the single-observation situation, using the autocorrelation coefficients averaged over all available observations.

The same result holds for the gain-independent case where the joint density for observations $x^{(1)} = s^{(1)}/\sigma_{(1)}$, $x^{(2)} = s^{(2)}/\sigma_{(2)}$, \dots , $x^{(L)} = s^{(L)}/\sigma_{(L)}$ is, after taking the L th root,

$$[f(x^{(1)}, x^{(2)}, \dots, x^{(L)} | \mathbf{a})]^{1/L} = (2\pi)^{-N/2} \exp \left\{ -\frac{1}{2} \left[\frac{1}{L} \sum_{i=1}^L \alpha(x^{(i)}; \mathbf{a}) \right] \right\}. \quad (36)$$

Note that gain independence is maintained by normalizing each observation $s^{(i)}$ with its own estimate $\sigma_{(i)}$. Equation (35) thus becomes

$$\begin{aligned} & \frac{1}{L} \sum_{i=1}^L \alpha(x^{(i)}; \mathbf{a}) \\ &= r_a(0) \left[\frac{1}{L} \sum_{i=1}^L r_x^{(i)}(0) \right] + 2 \sum_{j=1}^M r_a(j) \left[\frac{1}{L} \sum_{i=1}^L r_x^{(i)}(j) \right] \end{aligned}$$

$$= r_a(0) \left[\frac{1}{L} \sum_{i=1}^L \frac{r_s^{(i)}(0)}{\sigma_{(i)}^2} \right] + 2 \sum_{j=1}^M r_a(j) \left[\frac{1}{L} \sum_{i=1}^L \frac{r_s^{(i)}(j)}{\sigma_{(i)}^2} \right]. \quad (37)$$

Equations (35) and (37) lead to the same optimization procedure as the centroid computation in code-book design for vector quantization.¹³ In centroid computation, however, the objective is to minimize the average distortion, while in the current probabilistic framework, the probability is to be maximized. The equivalence between probability and distortion measures is again witnessed.

Once the source parameters are estimated, the probability density is defined just as in the single-observation case, and expressed in (26) or (31) in terms of distortion measures.

IV. PROBABILISTIC FUNCTIONS OF MARKOV CHAINS

Consider a first-order K -state Markov chain governed by a transition probability matrix $\mathbf{V} = [v_{ij}]$, $i, j = 1, 2, \dots, K$, and an initial probability vector $\mathbf{u}^t = [u_1, u_2, \dots, u_K]$. Obviously,

$$\sum_{j=1}^K u_j = 1, \quad u_j \geq 0 \quad \text{for all } j \quad (38)$$

and

$$\sum_{j=1}^K v_{ij} = 1 \quad \text{for any } i, \quad (39)$$

because v_{ij} is the probability of making a transition from state i to state j given that the current state is i . For any integer state sequence $\Theta = \theta_0\theta_1 \dots \theta_T$, where $\theta_i \in \{1, 2, \dots, K\}$, the probability of Θ being generated by the Markov chain can be easily calculated by

$$\Pr(\Theta | \mathbf{V}, \mathbf{u}) = u_{\theta_0} v_{\theta_0\theta_1} v_{\theta_1\theta_2} \dots v_{\theta_{T-1}\theta_T}. \quad (40)$$

Now suppose $\Theta = \theta_0\theta_1 \dots \theta_T$ cannot be observed directly. Instead, we observe a stochastic process $\mathbf{S} = \mathbf{s}_1\mathbf{s}_2 \dots \mathbf{s}_T$, produced by an underlying state sequence $\theta_1\theta_2 \dots \theta_T$. Each state, say i , manifests itself through a probability density function $f_i(s)$. We use $\mathbf{F} = \{f_i(\cdot)\}$ to denote such a set of density functions. The probability density of observing $\mathbf{S} = S \triangleq s_1s_2 \dots s_T$ given a specific state sequence Θ generated by the Markov chain with transition probability matrix \mathbf{V} and initial probability \mathbf{u} is thus

$$f(S | \Theta, \mathbf{V}, \mathbf{u}, \mathbf{F}) = f_{\theta_1}(s_1) f_{\theta_2}(s_2) \dots f_{\theta_T}(s_T). \quad (41)$$

Each s_i here is a vector without ambiguity. It follows that the probability density of observing \mathbf{S} given \mathbf{V} and \mathbf{u} is

$$\begin{aligned}
f(S | \mathbf{V}, \mathbf{u}, \mathbf{F}) &= \sum_{\text{all } \Theta} f(S, \Theta | \mathbf{V}, \mathbf{u}, \mathbf{F}) \\
&= \sum_{\text{all } \Theta} f(S | \Theta, \mathbf{V}, \mathbf{u}, \mathbf{F}) \Pr(\Theta | \mathbf{V}, \mathbf{u}, \mathbf{F}) \\
&= \sum_{\text{all } \Theta} f(S | \Theta, \mathbf{V}, \mathbf{u}, \mathbf{F}) \Pr(\Theta | \mathbf{V}, \mathbf{u}) \\
&= \sum_{\theta_0, \theta_1, \dots, \theta_T=1}^K u_{\theta_0} v_{\theta_0 \theta_1} f_{\theta_1}(s_1) v_{\theta_1 \theta_2} f_{\theta_2}(s_2) \dots v_{\theta_{T-1} \theta_T} f_{\theta_T}(s_T). \quad (42)
\end{aligned}$$

The stochastic process \mathbf{S} is characterized by the density $f(S | \mathbf{V}, \mathbf{u}, \mathbf{F})$ and the set of probability density functions \mathbf{F} , which is assumed to be known and independent of the Markov chain in the above. The triple $(\mathbf{V}, \mathbf{u}, \mathbf{F}) \triangleq \mathbf{M}$ is then called a (hidden) Markov model,¹⁴ and the conditional density for the stochastic process \mathbf{S} may be written as $f(S | \mathbf{M})$.

The application of hidden Markov models in speech recognition can now be formulated. It is treated as a classification problem. We wish to recognize utterances known to have been selected from some vocabulary \mathbf{W} of B words $W^{(1)}, W^{(2)}, \dots, W^{(B)}$. (We use "words" here for convenience. They may not be words in the traditional sense, but merely some lengths of speech utterances.) Every word $W^{(i)}$ is represented by a model \mathbf{M}_i . An observation sequence $S = s_1 s_2 \dots s_T$ of an unknown word is given. We then apply the maximum likelihood rule and classify S as word $W^{(i)}$ iff

$$f(S | \mathbf{M}_i) \geq f(S | \mathbf{M}_j) \quad \text{for any } j = 1, 2, \dots, B.$$

Such an application presents two problems: evaluating $f(S | \mathbf{M}_i)$ and estimating model \mathbf{M} that maximizes the likelihood of a given observation S . Sections II and III showed similar problems.

The computational load in evaluating $f(S | \mathbf{M})$ appears to be exponential in T , as we see from (42), which is a sum over all possible state sequences of length T . With the so-called forward-backward algorithm by Baum,¹⁵ however, it is only linear in T . The estimation of model parameters \mathbf{V} , \mathbf{u} , and \mathbf{F} , on the other hand, is less straightforward, and no closed form solution has been found so far. An iterative reestimation algorithm by Baum¹⁵ and Baum et al.¹⁶ is usually employed to attack this estimation problem for a certain class of hidden Markov models, including those with Gaussian autoregressive densities of (1), (17), or (18). We shall briefly discuss the forward-backward algorithm and the reestimation formula for Markov models with Gaussian autoregressive density. Similar developments with applications in speaker identification can be found in the work of Poritz.¹⁷ More rigorous developments of these techniques, as well as their theoretical verifications, can be found in Refs. 15 and 16.

4.1 Forward-backward recursion and trellis structure

Define the forward probabilities $\xi_0(i) = u_i$, $i = 1, 2, \dots, K$, and

$$\xi_t(i) = \sum_{j=1}^K \xi_{t-1}(j) v_{ji} f_i(s_t) \quad (43)$$

for $i = 1, 2, \dots, K$ and $t = 1, 2, \dots, T$. Clearly,

$$\xi_t(i) = f(s_1, s_2, \dots, s_t, \theta_t = i | \mathbf{M}).$$

Similarly, define backward probabilities

$$\begin{aligned} \eta_t(i) &= f(s_{t+1}, s_{t+2}, \dots, s_T | \theta_t = i, \mathbf{M}) \\ &= \sum_{j=1}^K \eta_{t+1}(j) v_{ij} f_j(s_{t+1}) \end{aligned} \quad (44)$$

and $\eta_T(i) = 1$, for $i = 1, 2, \dots, K$, and $t = T - 1, T - 2, \dots, 0$. $\xi_t(i)$ and $\eta_t(i)$ satisfy

$$\xi_t(i) \eta_t(i) = f(S, \theta_t = i | \mathbf{M}). \quad (45)$$

Therefore, we have

$$\begin{aligned} f(S | \mathbf{M}) &= \sum_{i=1}^K f(S, \theta_t = i | \mathbf{M}) \\ &= \sum_{i=1}^K \xi_t(i) \eta_t(i) \end{aligned} \quad (46)$$

for any t . In particular, by letting $t = T$,

$$f(S | \mathbf{M}) = \sum_{i=1}^K \xi_T(i) \quad (47)$$

so that $f(S | \mathbf{M})$ can be evaluated from forward probabilities alone and the computation load is thus linear in T .

This forward-backward evaluation technique takes advantage of a trellis structure when reducing the computational burden. The complexity of a tree structure, as Fig. 2 illustrates, grows exponentially in T . It treats distinctive paths differently, as if at instance t , the number of available state indices were K^t , as we noted by the parenthesized index in Fig. 2. The original evaluation formula of (42) displays a tree structure since the summation is directly over all θ_i , $i = 0, 1, 2, \dots, T$, in the range of 1 through K . This tree structure can be transformed easily into a trellis structure, as Fig. 3 depicts for $K = 4$, by recognizing the fact that at any instance t and any state θ_t there are always only K possible next states regardless of the past transition history. The branches in the tree structure merge into K nodes (states) at every instance. The definition of the forward probabilities can be stated as

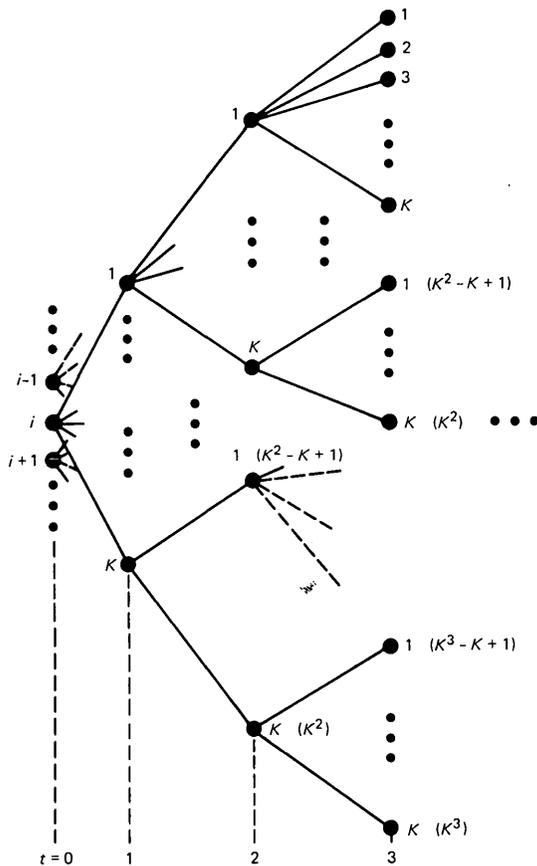


Fig. 2—Tree structure showing exponential growth of complexity.

follows: the density of the event that $\theta_t = i$ and s_1, s_2, \dots, s_t are observed is obtained by summing, over all j , the densities of the event that $\theta_{t-1} = j$ and s_1, s_2, \dots, s_{t-1} are observed, multiplied by the transition probability from state j to i and by the density of $s_t = s_t$ at state i . Thus, as Fig. 3 shows for $K = 4$, $\xi_t(1) = \xi_{t-1}(1)v_{11}f_1(s_t) + \xi_{t-1}(2)v_{21}f_1(s_t) + \xi_{t-1}(3)v_{31}f_1(s_t) + \xi_{t-1}(4)v_{41}f_1(s_t)$. Every $\xi_t(i)$ can be computed recursively from $\xi_{t-1}(i)$, $i = 1, 2, \dots, K$. Each progression from $t - 1$ to t requires the same amount of computation and, therefore, total computation is linear in T .

The tree/trellis structure also shows that dynamic programming techniques such as the Viterbi algorithm or the (M, L) algorithm, etc., can be employed efficiently to find a particular path Θ_{op} such that $f(S, \Theta_{op} | \mathbf{M}) = \max_{\text{all } \theta} f(S, \Theta | \mathbf{M})$. Later we will show that dynamic time

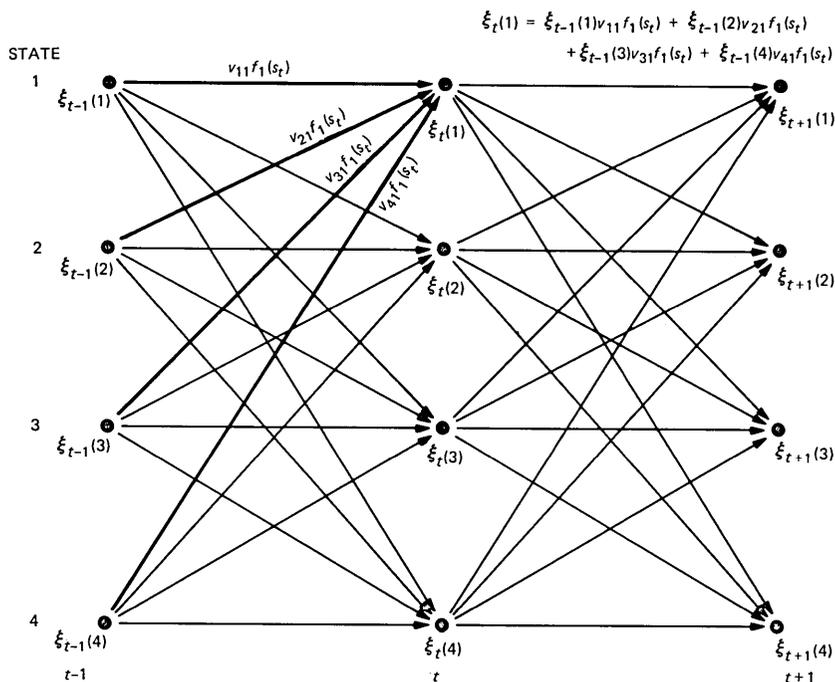


Fig. 3—Trellis structure for evaluating probability in hidden Markov models.

warping displays similar requirements of finding an optimal path (the warping function) to minimize the accumulative distortion.

4.2 Reestimation—Baum-Welch algorithm

We discuss here only the estimation of Markov models with Gaussian autoregressive densities; that is, in $\mathbf{F} = \{f_i\}_{i=1}^K$, every f_i takes the form of (17)* and is characterized by parameters $(\mathbf{a}_i, \sigma_i^2)$. Equivalently, we write $\mathbf{F} = \{(\mathbf{a}_i, \sigma_i^2)\}_{i=1}^K$. The objective of model estimation here is to find a model $\mathbf{M} = (\mathbf{V}, \mathbf{u}, \mathbf{F})$ that maximizes the likelihood of a given sequence S , for fixed number of states K and order of autoregression M .

Given a sequence S and an arbitrary model \mathbf{M} , the Baum-Welch reestimation algorithm iteratively finds another model $\mathbf{M}' = (\mathbf{V}', \mathbf{u}', (\mathbf{a}'_i, \sigma_i'^2))$ that leads to $f(S|\mathbf{M}') \geq f(S|\mathbf{M})$. The algorithm continually improves the estimate and converges to a local optimum. Let

$$\begin{aligned} \gamma_t(i) &\triangleq f(S, \theta_t = i | \mathbf{M}) \\ &= \xi_t(i) \cdot \eta_t(i) \end{aligned} \quad (48)$$

* The gain-independent case can be easily developed in the same way that the previous sections showed.

for $t = 0, 1, 2, \dots, T$, and $i = 1, 2, \dots, K$. Further, define

$$\begin{aligned} \gamma_t(i, j) &\triangleq f(S, \theta_{t-1} = i, \theta_t = j | \mathbf{M}) \\ &= \xi_{t-1}(i) v_{ij} f_j(s_t) \eta_t(j) \end{aligned} \quad (49)$$

for $i, j = 1, 2, \dots, K$, and $t = 1, 2, \dots, T$. A new estimate of transition probability v'_{ij} for \mathbf{V}' is obtained by

$$v'_{ij} = \frac{\sum_{t=1}^T \gamma_t(i, j)}{\sum_{t=1}^T \gamma_t(i)}. \quad (50)$$

And by applying Theorem 3.1 of Ref. 16 one chooses $(\mathbf{a}'_i, \sigma_i'^2)$, for each $i = 1, 2, \dots, K$, such that (see Appendix A)

$$\alpha(\bar{s}_i; \mathbf{a}'_i) = \min_{\mathbf{a}} \alpha(\bar{s}_i; \mathbf{a}) \quad (51)$$

and

$$\sigma_i'^2 = \frac{\alpha(\bar{s}_i; \mathbf{a}'_i)}{N \sum_{t=1}^T \gamma_t(i)}, \quad (52)$$

where \bar{s}_i represents a composite observation whose autocorrelation coefficients are $\bar{r}_i(j)$,

$$\begin{aligned} \bar{r}_i(j) &= \sum_{t=1}^T \gamma_t(i) r_t(j) \\ &= \sum_{t=1}^T \gamma_t(i) \sum_{n=1}^{N-j} s_{t,n} s_{t,n+j}. \end{aligned} \quad (53)$$

Note that $s_{t,n}$ is the n th sample in observation s_t . The composite autocorrelation $\bar{r}_i(j)$ is, as seen from (53), a weighted average autocorrelation. The weight is the density or relative frequency of the observation S being at state i at instance t . The concept of relative frequency may be helpful in relating (53) to (35) where a uniform average is involved. After $\bar{r}_i(j)$ for $i = 1, 2, \dots, K$, and $j = 0, 1, 2, \dots, M$ have been calculated, each $(\mathbf{a}'_i, \sigma_i'^2)$ is found by using the same maximum likelihood estimation procedure as in Section II.

In estimating the Markov model parameters based upon multiple observations $S^{(1)}, S^{(2)}, \dots, S^{(L)}$, we try to maximize the joint density given the same model $f(S^{(1)}, S^{(2)}, \dots, S^{(L)} | \mathbf{M})$. Since $S^{(i)}$ are independently observed,

$$\begin{aligned}
 f(S^{(1)}, S^{(2)}, \dots, S^{(L)} | \mathbf{M}) \\
 = \prod_{i=1}^L f(S^{(i)} | \mathbf{M}).
 \end{aligned} \tag{54}$$

The Baum-Welch algorithm can be equally applied and the key equation of (50) for new estimate of transition probability becomes (see Appendix B)

$$v_{ij} = \frac{\sum_{t=1}^T \left[\sum_{l=1}^L \gamma_t^{(l)}(i, j) \right]}{\sum_{t=1}^T \left[\sum_{l=1}^L \gamma_t^{(l)}(i) \right]}, \tag{55}$$

where

$$\gamma_t^{(l)}(i) = \frac{f(S^{(l)}, \theta_t = i | \mathbf{M})}{f(S^{(l)} | \mathbf{M})} \tag{56}$$

and

$$\gamma_t^{(l)}(i, j) = \frac{f(S^{(l)}, \theta_{t-1} = i, \theta_t = j | \mathbf{M})}{f(S^{(l)} | \mathbf{M})}. \tag{57}$$

The new estimate of $(\mathbf{a}'_i, \sigma_i'^2)$ of (51) and (52) then follows the same procedure as previously discussed for the maximum likelihood estimate based upon multiple observations. In particular, one can show (see Appendix B) that the new improved estimate $(\mathbf{a}'_i, \sigma_i'^2)$, according to the reestimation algorithm, satisfies

$$\alpha(\bar{s}_i; \mathbf{a}'_i) = \min_{\mathbf{a}} \alpha(\bar{s}_i, \mathbf{a}) \tag{58}$$

and

$$\sigma_i'^2 = \frac{\alpha(\bar{s}_i; \mathbf{a}'_i)}{N \sum_{t=1}^T \sum_{l=1}^L \gamma_t^{(l)}(i)}, \tag{59}$$

where \bar{s}_i now represents a composite observation whose autocorrelation coefficients are $\bar{r}_i(j)$,

$$\begin{aligned}
 \bar{r}_i(j) &= \sum_{t=1}^T \sum_{l=1}^L \gamma_t^{(l)}(i) r_t^{(l)}(j) \\
 &= \sum_{t=1}^T \sum_{l=1}^L \gamma_t^{(l)}(i) \sum_{n=1}^{N-j} s_{t,n}^{(l)} s_{t,n+j}^{(l)},
 \end{aligned} \tag{60}$$

with $s_{t,n}^{(l)}$ being the n th sample in the observation vector s_t of sequence $S^{(l)}$.

V. DYNAMIC TIME WARPING AND HIDDEN MARKOV MODELS

In this section, we establish the relationship between dynamic time warping using linear predictive coding (LPC) distance measures and hidden Markov models with Gaussian autoregressive densities. We give a unified view on these two techniques such that comparative discussions on the two techniques can be made easily.

Consider two speech sequences, $W = w_1 w_2 \dots w_{T_w}$ and $Y = y_1 y_2 \dots y_{T_y}$, called the reference and the test sequence, respectively. Through some warping function $\phi(\cdot)$, or $\zeta(\cdot)$,

$$t_y = \phi(t_w), \quad t_w = 1, 2, \dots, T_w \quad (61)$$

$$t_w = \zeta(t_y), \quad t_y = 1, 2, \dots, T_y, \quad (62)$$

a correspondence between W and Y can be established. Let $d[\cdot; \cdot]$ be a distortion measure. The conventional dynamic time warping uses dynamic programming techniques to determine ϕ or ζ such that

$$D_\phi = \sum_{t_w=1}^{T_w} d[w_{t_w}; y_{\phi(t_w)}] \quad (63)$$

or

$$D_\zeta = \sum_{t_y=1}^{T_y} d[y_{t_y}; w_{\zeta(t_y)}] \quad (64)$$

is minimized. In recognition, we classify an utterance Y as word $W^{(i)}$ in a vocabulary $\{W^{(1)}, W^{(2)}, \dots, W^{(B)}\}$ if

$$\begin{aligned} [D_\phi^{(i)}]_{\min} &= \min_{\phi^{(i)}} \left\{ \sum_{t_w=1}^{T_w^{(i)}} d[w_{t_w}^{(i)}; y_{\phi^{(i)}(t_w)}] \right\} \\ &\leq \min_{\phi^{(j)}} \left\{ \sum_{t_w=1}^{T_w^{(j)}} d[w_{t_w}^{(j)}; y_{\phi^{(j)}(t_w)}] \right\} \quad \text{for } j = 1, 2, \dots, B \end{aligned} \quad (65)$$

or

$$\begin{aligned} [D_\zeta^{(i)}]_{\min} &= \min_{\zeta^{(i)}} \left\{ \sum_{t_y=1}^{T_y^{(i)}} d[y_{t_y}; w_{\zeta^{(i)}(t_y)}] \right\} \\ &\leq \min_{\zeta^{(j)}} \left\{ \sum_{t_y=1}^{T_y^{(j)}} d[y_{t_y}; w_{\zeta^{(j)}(t_y)}] \right\} \quad \text{for } j = 1, 2, \dots, B. \end{aligned} \quad (66)$$

The choice of warping directions, i.e., using ϕ or ζ , is rather arbitrary and often is taken into consideration together with some continuity conditions.^{2,3}

Now, consider a T_w -state hidden Markov model $\mathbf{M}_w = (\mathbf{V}_w, \mathbf{u}_w, \mathbf{F}_w)$,

where \mathbf{V}_w is a $T_w \times T_w$ matrix $\mathbf{V}_w = [v_{ij}]$, $v_{ij} = 1/T_w$, for any $i, j = 1, 2, \dots, T_w$, \mathbf{u}_w is a T_w -dimensional vector with $u_i = 1/T_w$ for any $i = 1, 2, \dots, T_w$, and \mathbf{F}_w is the set of Gaussian autoregressive densities $\{f_i\}_{i=1}^{T_w}$. Each f_i is defined by a parameter pair $(\mathbf{a}_{w,i}, \sigma_{w,i}^2)$, which is the maximum likelihood estimate based upon observation w_i . We further consider a particular state sequence, called progressive sequence $\Theta_w = \theta_0 \theta_1 \dots \theta_{T_w}$, where $\theta_i = i$ and where $\theta_0 \in \{1, 2, \dots, T_w\}$ is arbitrary. Then,

$$\begin{aligned} f(W, \Theta_w | \mathbf{M}_w) \\ = (1/T_w)^{T_w+1} \prod_{i=1}^{T_w} f_i(w_i | \mathbf{a}_{w,i}, \sigma_{w,i}^2) \end{aligned}$$

and from (22),

$$\begin{aligned} \log f(W, \Theta_w | \mathbf{M}_w) \\ = -\frac{N}{2} \left[\sum_{i=1}^{T_w} \log(2\pi\sigma_{w,i}^2) + T_w \right] - (T_w + 1)\log T_w \\ = \max_{\Theta \in \Theta_{(T_w)}} \{\log f(W, \Theta | \mathbf{M}_w)\}, \end{aligned} \quad (67)$$

where $\Theta_{(T_w)}$ denotes the set of all state sequences with length T_w . We define a similar model $\mathbf{M}_y = (\mathbf{V}_y, \mathbf{u}_y, \mathbf{F}_y)$ for Y , in which $v_{ij} = 1/T_y$, $u_i = 1/T_y$ for $i, j = 1, 2, \dots, T_y$, and $\mathbf{F}_y = \{(\mathbf{a}_{y,i}, \sigma_{y,i}^2)\}_{i=1}^{T_y}$, where each $(\mathbf{a}_{y,i}, \sigma_{y,i}^2)$ is estimated based on y_i . We thus also have

$$\begin{aligned} f(Y, \Theta_y | \mathbf{M}_y) \\ = (1/T_y)^{T_y+1} \prod_{i=1}^{T_y} f_i(y_i | \mathbf{a}_{y,i}, \sigma_{y,i}^2) \end{aligned}$$

and

$$\begin{aligned} \log f(Y, \Theta_y | \mathbf{M}_y) \\ = -\frac{N}{2} \left[\sum_{i=1}^{T_y} \log(2\pi\sigma_{y,i}^2) + T_y \right] - (T_y + 1)\log T_y \\ = \max_{\Theta \in \Theta_{(T_y)}} \{\log f(Y, \Theta | \mathbf{M}_y)\}. \end{aligned} \quad (68)$$

A correspondence between progressive state sequences Θ_w and Θ_y is made through warping functions ϕ or ζ , as defined in (61) or (62). Within such a framework,

$$\begin{aligned}
 f(Y, \zeta(\Theta_y) | \mathbf{M}_w) &= \left(\frac{1}{T_w}\right)^{T_y+1} \prod_{i=1}^{T_y} f_{\zeta(i)}(y_i | \mathbf{a}_{w,\zeta(i)}, \sigma_{w,\zeta(i)}^2) \quad (69)
 \end{aligned}$$

and

$$\begin{aligned}
 \log f(Y, \zeta(\Theta_y) | \mathbf{M}_w) &= -\frac{N}{2} \sum_{i=1}^{T_y} \log(2\pi\sigma_{w,\zeta(i)}^2) - \frac{1}{2} \sum_{i=1}^{T_y} \alpha(\sigma_{w,\zeta(i)}^{-1} y_i; \mathbf{a}_{w,\zeta(i)}) \\
 &\quad - (T_y + 1)\log T_w. \quad (70)
 \end{aligned}$$

In the above we have used $\zeta(\Theta_y)$ to denote the ζ -warped progressive sequence Θ_y . Note that (68) is also a maximum over all models with fixed \mathbf{V}_y and \mathbf{u}_y . The difference between (68) and (70),

$$\begin{aligned}
 \log f(Y, \Theta_y | \mathbf{M}_y) - \log f(Y, \zeta(\Theta_y) | \mathbf{M}_w) &= \frac{N}{2} \sum_{i=1}^{T_y} \left\{ \frac{1}{N} \alpha(\sigma_{w,\zeta(i)}^{-1} y_i; \mathbf{a}_{w,\zeta(i)}) + \log \sigma_{w,\zeta(i)}^2 \right. \\
 &\quad \left. - \log \sigma_{y,i}^2 - 1 \right\} + (T_y + 1)(\log T_w - \log T_y), \quad (71)
 \end{aligned}$$

is thus nonnegative if $T_w \geq T_y$ (sufficient). In realistic situations, $\log T_w \simeq \log T_y$, so that

$$\begin{aligned}
 \log f(Y, \Theta_y | \mathbf{M}_y) - \log f(Y, \zeta(\Theta_y) | \mathbf{M}_w) &\simeq \frac{N}{2} \sum_{i=1}^{T_y} \left\{ \frac{1}{N} \alpha(\sigma_{w,\zeta(i)}^{-1} y_i; \mathbf{a}_{w,\zeta(i)}) + \log \sigma_{w,\zeta(i)}^2 - \log \sigma_{y,i}^2 - 1 \right\} \\
 &= \frac{N}{2} \sum_{i=1}^{T_y} d_{IS}[y_i; w_{\zeta(i)}] \\
 &= \frac{N}{2} D_{\zeta}. \quad (72)
 \end{aligned}$$

Therefore, the accumulative distortion D_{ζ} in dynamic time warping is directly related to the likelihood difference between the two models in generating Y sequence. To express the density $f(Y | \mathbf{M}_w)$ in terms of D_{ζ} we further have

$$\begin{aligned}
 f(Y | \mathbf{M}_w) &= \sum_{\text{all } \zeta} f(Y, \zeta(\Theta_y) | \mathbf{M}_w) \\
 &\simeq f(Y, \Theta_y | \mathbf{M}_y) \left[\sum_{\text{all } \zeta} \exp\left(-\frac{N}{2} D_{\zeta}\right) \right]. \quad (73)
 \end{aligned}$$

The same results can be extended to $f(W | \mathbf{M}_y)$,

$$\begin{aligned}
 f(W|\mathbf{M}_y) &= \sum_{\text{all } \phi} f(W, \phi(\Theta_y) | \mathbf{M}_y) \\
 &\simeq f(W, \Theta_w | \mathbf{M}_w) \left[\sum_{\text{all } \phi} \exp \left(-\frac{N}{2} D_\phi \right) \right]. \quad (74)
 \end{aligned}$$

Equation (72) demonstrates that determining a warping function ζ in dynamic time warping is equivalent to finding only the best state sequence that maximizes the density $f(Y, \zeta(\Theta_y) | \mathbf{M}_w)$. On the other hand, the density $f(Y | \mathbf{M}_w)$ can be calculated by summing $\exp \{- (N/2) D_\zeta\}$ over all possible warping paths ζ and then multiplying this sum of exponential terms by a constant $f(Y, \Theta_y | \mathbf{M}_y)$. Since $f(Y, \Theta_y | \mathbf{M}_y)$ is defined by the unknown sequence Y only, it does not affect the classification problem for recognition formulated above. The accumulative distortions D_ζ are then the key determining factor.

Although we have defined the transition probabilities v_{ij} to be all equal for any i and j , it is not absolutely necessary for the results of (72). It has, however, a simple but important interpretation in the calculation of $f(Y | \mathbf{M}_w)$. Equal v_{ij} for $i, j = 1, 2, \dots, T_w$ allows the input sequence Θ_y to be warped in every possible way and in every possible permutation. One thus may not expect a good recognition performance based upon the density $f(Y | \mathbf{M}_w) = \sum_{\zeta} f(Y, \zeta(\Theta_y) | \mathbf{M}_w)$, as the time order of the observed speech sequence is crucially important. (A reversed "we" may sound very close to "you"!) For word recognition, some constraints on the transitions are therefore desirable. Markov models for two types of serial constraints, for example, are shown in Fig. 4. In Fig. 4a, single and double transitions are allowed. Similar results linking dynamic time warping and Markov modeling can be obtained in this case if

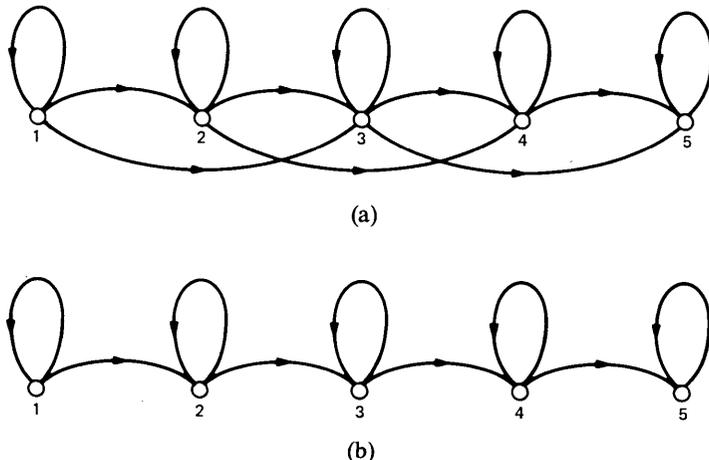


Fig. 4—Markov chains for two types of serial constraints: (a) single and double transitions permitted, and (b) only single transitions permitted.

$$V = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \dots & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \dots \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ \vdots & & & \ddots & & \\ & & & & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ & & & & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & & & & 0 & 0 & 1 \end{bmatrix}.$$

For the single transition case of Fig. 4b,

$$V = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \vdots & & \ddots & \\ & & & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & & & 0 & 0 & 1 \end{bmatrix}.$$

These kinds of constraints all have their counterparts in dynamic time warping, appearing as the continuity conditions for determining the warping function. It is expected that these constant transition probabilities lead to a similar approximation, as appeared in (72). Unlike (71), however, the lack of exactness in (72) is now caused by the increased transition probabilities at the end of the sequence. The idea of unconstrained endpoint algorithm in dynamic time warping to correct for the abrupt change in transition "possibilities" at word boundaries is, hence, noted.

VI. SUMMARY

We have given a unified theoretical view on the two dominant speech recognition techniques, namely dynamic time warping and Markov modeling. We described the role of some well-known distortion measures in the context of probability densities. After the relationship between probability density and distortion measures is made explicit, the similarities between the two techniques can be seen. We have shown that if the underlying transition structure is equiprobable, dynamic time warping is equivalent to the probabilistic modeling

technique except that it searches for the best transition path to minimize the accumulative distortion, while the probabilistic technique sums the density along every possible path. The results show that the two techniques may not be mutually exclusive, particularly when the density functions are exponential and LP-related distortion measures are used. The discussion may be helpful in bringing about a better understanding of each technique and possible future improvements.

REFERENCES

1. W. A. Lea, ed., "Speech Recognition: Past, Present, and Future," in *Trends in Speech Recognition*, Englewood Cliffs, NJ: Prentice-Hall, 1980, pp. 39-98.
2. F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-23*, No. 1 (February 1975), pp. 67-72.
3. L. R. Rabiner and S. E. Levinson, "Isolated Connected Word Recognition—Theory and Selected Applications," *IEEE Trans. Commun., COM-29*, No. 5 (May 1981), pp. 621-59.
4. H. Sakoe and S. Chiba, "A Dynamic Programming Approach to Continuous Speech Recognition," *Proc. Int. Congress on Acoustics, Budapest, Hungary, Paper 20 C-13*, 1971.
5. F. Jelinek, "Continuous Speech Recognition by Statistical Methods," *Proc. IEEE*, 64 (April 1976), pp. 532-56.
6. J. K. Baker, "Stochastic Modeling for Automatic Speech Understanding," in *Speech Recognition*, R. Reddy, ed., New York: Academic Press, 1975.
7. L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent Isolated Word Recognition," *B.S.T.J.*, 62, No. 4 (February 1983), pp. 429-56.
8. F. Itakura, *Speech Analysis and Synthesis Systems Based on Statistical Method*. Doctor of Engineering Dissertation (Dept. of Engineering, Nagoya University, Japan, 1972) (in Japanese).
9. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, New York: Springer-Verlag, 1976.
10. R. M. Gray et al., "Distortion Measures for Speech Processing," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-28* (August 1980), pp. 367-76.
11. J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, 63 (April 1975), pp. 561-80.
12. B. H. Juang, D. Y. Wong, and A. H. Gray, Jr., "Distortion Performance of Vector Quantization for LPC Voice Coding," *IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-30*, No. 2 (April 1982), pp. 294-304.
13. A. Buzo et al., "Speech Coding Based Upon Vector Quantization," *IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-28* (October 1980), pp. 562-74.
14. S. E. Levinson, L. R. Rabiner, and M. M. Sondhi, "An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition," *B.S.T.J.*, 62, No. 4 (April 1983), pp. 1035-74.
15. L. E. Baum, "An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of a Markov Process," *Inequalities*, 3 (1972), pp. 1-8.
16. L. E. Baum et al., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *Ann. Math. Statist.*, 41 (1970), pp. 164-71.
17. A. B. Poritz, "Linear Predictive Hidden Markov Models and the Speech Signal," *ICASSP-82 Proc.*, May 1982, pp. 1291-4.

APPENDIX A

Reestimation—Single Observation

We first show that the new, improved estimate ($\mathbf{a}'_i, \sigma_i'^2$) satisfies (51) and (52). We follow Ref. 16 and define the Q -function as

$$Q(\mathbf{M}, \mathbf{M}') = \sum_{\text{all } \Theta} f(S, \Theta | \mathbf{M}) \log f(S, \Theta | \mathbf{M}'), \quad (75)$$

where

$$\begin{aligned} & \log f(S, \Theta | \mathbf{M}') \\ &= \sum_{j=1}^K (\log u'_j) \delta(\theta_0 - j) \\ & \quad + \sum_{k=1}^K \sum_{i=1}^K \sum_{t=1}^T (\log v'_{ki}) \delta(\theta_{t-1} - k) \delta(\theta_t - l) \\ & \quad + \sum_{i=1}^K \sum_{t=1}^T [\log f'_i(s_t)] \delta(\theta_t - i), \end{aligned} \quad (76)$$

$\delta(\cdot)$ is the Kronecker delta, and

$$\begin{aligned} f'_i(s) &= f(s | \mathbf{a}'_i, \sigma_i'^2) \\ &= (2\pi)^{-N/2} (\sigma_i')^{-N} \exp \left\{ -\frac{1}{2} \alpha(\sigma_i'^{-1} s; \mathbf{a}_i) \right\} \end{aligned} \quad (77)$$

according to (17). It has been shown that¹⁶

$$\begin{aligned} f(S | \mathbf{M}') &= \sum_{\text{all } \Theta} u'_{\theta_0} \prod_{t=1}^T v'_{\theta_{t-1}\theta_t} f'_{\theta_t}(s_t) \\ &\geq \sum_{\text{all } \Theta} u_{\theta_0} \prod_{t=1}^T v_{\theta_{t-1}\theta_t} f_{\theta_t}(s_t) \\ &= f(S | \mathbf{M}) \end{aligned}$$

if $Q(\mathbf{M}, \mathbf{M}') \geq Q(\mathbf{M}, \mathbf{M})$. Therefore, we may obtain a new, improved estimate by maximizing $Q(\mathbf{M}, \mathbf{M}')$ with respect to \mathbf{M}' . Equation (76) shows that the contributions due to \mathbf{u}' , \mathbf{V}' , and \mathbf{F}' are separated and maximization of $Q(\mathbf{M}, \mathbf{M}')$ can thus be carried out independently with respect to each parameter set. In particular,

$$\begin{aligned} Q(\mathbf{M}, \mathbf{M}') &= Q_u(\mathbf{M}, \mathbf{u}') + Q_v(\mathbf{M}, \mathbf{V}') + Q_f(\mathbf{M}, \mathbf{F}') \\ &= Q_u(\mathbf{M}, \mathbf{u}') + Q_v(\mathbf{M}, \mathbf{V}') + \sum_{i=1}^K Q_{f_i}(\mathbf{M}, f'_i), \end{aligned}$$

where

$$Q_u(\mathbf{M}, \mathbf{u}') = \sum_{\text{all } \Theta} f(S, \Theta | \mathbf{M}) \sum_{j=1}^K (\log u'_j) \delta(\theta_0 - j),$$

$$Q_v(\mathbf{M}, \mathbf{V}') = \sum_{\text{all } \Theta} f(S, \Theta | \mathbf{M}) \left[\sum_{k=1}^K \sum_{l=1}^K \sum_{t=1}^T (\log v'_{kl}) \delta(\theta_{t-1} - k) \delta(\theta_t - l) \right],$$

and

$$Q_{f_i}(\mathbf{M}, f'_i) = \sum_{\text{all } \Theta} f(S, \Theta | \mathbf{M}) \sum_{t=1}^T [\log f'_i(s_t)] \delta(\theta_t - i).$$

We shall not repeat the treatment of \mathbf{u}' and \mathbf{V}' here, but only discuss the maximization of $Q_{f_i}(\mathbf{M}, f'_i)$. We rewrite $Q_{f_i}(\mathbf{M}, f'_i)$ as

$$\begin{aligned} Q_{f_i}(\mathbf{M}, f'_i) &= \sum_{t=1}^T \left[\sum_{\text{all } \Theta} f(S, \Theta | \mathbf{M}) \delta(\theta_t - i) \right] \log f'_i(s_t) \\ &= \sum_{t=1}^T f(S, \theta_t = i | \mathbf{M}) \log f'_i(s_t) \\ &= \sum_{t=1}^T f(S, \theta_t = i | \mathbf{M}) \\ &\quad \cdot \left\{ -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma_i'^2 - \frac{1}{2} \alpha(\sigma_i'^{-1} s_t; \mathbf{a}'_i) \right\}. \end{aligned} \quad (78)$$

Maximizing $Q_{f_i}(\mathbf{M}, f'_i)$ with respect to \mathbf{a}'_i is equivalent to minimizing

$$\begin{aligned} &\sum_{t=1}^T f(S, \theta_t = i | \mathbf{M}) \alpha(s_t; \mathbf{a}'_i) \\ &= \sum_{t=1}^T f(S, \theta_t = i | \mathbf{M}) \\ &\quad \cdot \left\{ r_{\alpha'}(0) r_t(0) + 2 \sum_{j=1}^M r_{\alpha'}(j) r_t(j) \right\} \\ &= r_{\alpha'}(0) \left[\sum_{t=1}^T f(S, \theta_t = i | \mathbf{M}) r_t(0) \right] \\ &\quad + 2 \sum_{j=1}^M r_{\alpha'}(j) \left[\sum_{t=1}^T f(S, \theta_t = i | \mathbf{M}) r_t(j) \right], \end{aligned} \quad (79)$$

where $r_{\alpha'}(j)$ and $r_t(j)$ are the j -lag autocorrelation coefficient of \mathbf{a}'_i and s_t , respectively. Since (79) is simply $\alpha(\bar{s}_i; \mathbf{a}'_i)$, (51) is thus proved. Equation (52) then follows from maximizing (78) with respect to $\sigma_i'^2$ given \mathbf{a}'_i .

APPENDIX B

Reestimation—Multiple Observations

We apply the Baum-Welch algorithm to maximize

$$f(S^{(1)}, S^{(2)}, \dots, S^{(L)} | \mathbf{M}) = \prod_{i=1}^L f(S^{(i)} | \mathbf{M}), \tag{80}$$

given multiple observations $S^{(1)}, S^{(2)}, \dots, S^{(L)}$. For brevity, we use $\{S\}$ to denote the set of observations $S^{(i)}, i = 1, 2, \dots, L$. In addition, for each observation sequence $S^{(i)}$, there is a corresponding probable state sequence $\Theta^{(i)}$, and

$$f(S^{(i)} | \mathbf{M}) = \sum_{\text{all } \Theta^{(i)}} f(S^{(i)}, \Theta^{(i)} | \mathbf{M}).$$

We further use $\{\Theta\}$ to denote the set of probable state sequences behind the set of observations $\{S\}$. Then (80) becomes

$$f(\{S\} | \mathbf{M}) = \sum_{\text{all } \{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}). \tag{81}$$

Accordingly, we define the Q -function as follows:

$$Q(\mathbf{M}, \mathbf{M}') = \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \log f(\{S\}, \{\Theta\} | \mathbf{M}'),$$

where

$$\begin{aligned} \log f(\{S\}, \{\Theta\} | \mathbf{M}') &= \sum_{j=1}^K \sum_{i=1}^L (\log u_j) \delta(\theta_i^{(j)} - j) \\ &+ \sum_{k=1}^K \sum_{l=1}^K \sum_{i=1}^L \sum_{t=1}^T (\log v_{kl}') \delta(\theta_{i-1}^{(k)} - k) \delta(\theta_i^{(l)} - l) \\ &+ \sum_{m=1}^K \sum_{i=1}^L \sum_{t=1}^T [\log f'_m(s_t^{(i)})] \delta(\theta_t^{(i)} - m). \end{aligned} \tag{82}$$

and $f'_m(s)$ is defined as in (77). We shall address the maximization of $Q(\mathbf{M}, \mathbf{M}')$ with respect to $\mathbf{V}' = [v'_{kl}]$ and $\mathbf{F}' = \{f'_i\}$. Extension of previous results on the initial probabilities to the current case of multiple observations is straightforward. Again, (82) shows separate contributions from different parameter sets and, hence, we write

$$Q(\mathbf{M}, \mathbf{M}') = Q_u(\mathbf{M}, \mathbf{u}') + Q_v(\mathbf{M}, \mathbf{V}') + Q_f(\mathbf{M}, \mathbf{F}'),$$

where

$$\begin{aligned}
 Q_u(\mathbf{M}, \mathbf{u}') &= \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_{j=1}^K \sum_{i=1}^L (\log u'_j) \delta(\theta_0^{(i)} - j) \\
 Q_v(\mathbf{M}, \mathbf{V}') &= \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_k^K \sum_l^K \sum_i^L \sum_t^T (\log v'_{ki}) \\
 &\quad \cdot \delta(\theta_{t-1}^{(i)} - k) \delta(\theta_t^{(i)} - l) \\
 &= \sum_{k=1}^K \left\{ \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_l^K \sum_i^L \sum_t^T (\log v'_{ki}) \right. \\
 &\quad \left. \cdot \delta(\theta_{t-1}^{(i)} - k) \delta(\theta_t^{(i)} - l) \right\} \\
 &= \sum_{k=1}^K Q_{vk}(\mathbf{M}, \mathbf{v}'_k),
 \end{aligned}$$

and

$$\begin{aligned}
 Q_F(\mathbf{M}, \mathbf{F}') &= \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_{m=1}^K \sum_{i=1}^L \sum_{t=1}^T [\log f'_m(s_t^{(i)})] \delta(\theta_t^{(i)} - m) \\
 &= \sum_{m=1}^K \left\{ \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_i^L \sum_t^T [\log f'_m(s_t^{(i)}) \delta(\theta_t^{(i)} - m)] \right\} \\
 &= \sum_{m=1}^K Q_{f_m}(\mathbf{M}, f'_m).
 \end{aligned}$$

In the above, $\mathbf{v}'_k = [v'_{k1} \ v'_{k2} \ \dots \ v'_{kK}]^t$ and

$$\sum_{i=1}^K v_{ki} = 1 \quad \text{for } k = 1, 2, \dots, K. \quad (83)$$

Let Ω be the Lagrangian of $Q_{vk}(\mathbf{M}, \mathbf{v}'_k)$ with respect to the constraint (83),

$$\Omega = Q_{vk}(\mathbf{M}, \mathbf{v}'_k) + \lambda \left[\sum_{i=1}^K v_{ki} - 1 \right].$$

We need to solve the equation

$$\frac{\partial \Omega}{\partial v'_{kj}} = \frac{\partial Q_{vk}}{\partial v'_{kj}} + \lambda = 0. \quad (84)$$

Then, we have

$$\frac{\partial Q_{vk}}{\partial v'_{kj}} = -\lambda = - \sum_{j=1}^K v'_{kj} \lambda = \sum_{j=1}^K v'_{kj} \frac{\partial Q_{vk}}{\partial v'_{kj}}. \quad (85)$$

Since,

$$\begin{aligned}
 \frac{\partial Q_{vk}}{\partial v'_{kj}} &= \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_{i=1}^L \sum_{t=1}^T \frac{1}{v'_{kj}} \delta(\theta_{t-1}^{(i)} - k) \delta(\theta_t^{(i)} - j) \\
 &= \frac{1}{v'_{kj}} \sum_{i=1}^L \sum_{t=1}^T \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \delta(\theta_{t-1}^{(i)} - k) \delta(\theta_t^{(i)} - j) \\
 &= \frac{1}{v'_{kj}} \sum_{i=1}^L \sum_{t=1}^T f(S^{(i)}, \theta_{t-1}^{(i)} = k, \theta_t^{(i)} = j | \mathbf{M}) \\
 &\quad \cdot f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M}), \tag{86}
 \end{aligned}$$

and

$$\begin{aligned}
 \sum_{j=1}^K v'_{kj} \frac{\partial Q_{vk}}{\partial v'_{kj}} &= \sum_{j=1}^K \sum_{i=1}^L \sum_{t=1}^T f(S^{(i)}, \theta_{t-1}^{(i)} = k, \theta_t^{(i)} = j | \mathbf{M}) \\
 &\quad \cdot f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M}) \tag{87} \\
 &= \sum_{i=1}^L \sum_{t=1}^T f(S^{(i)}, \theta_{t-1}^{(i)} = k | \mathbf{M}) f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M}),
 \end{aligned}$$

we arrive at the solution of (55). The term $f(\{S\} | \mathbf{M})$ in the above has no effect as it appears in both (86) and (87).

Next, we deal with the maximization of $Q_{ff}(\mathbf{M}, f'_j)$. Note from above that

$$\begin{aligned}
 Q_{ff}(\mathbf{M}, f'_j) &= \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \sum_{i=1}^L \sum_{t=1}^T [\log f'_j(s_t^{(i)}) \delta(\theta_t^{(i)} - j)] \\
 &= \sum_{t=1}^T \sum_{i=1}^L \left\{ \sum_{\{\Theta\}} f(\{S\}, \{\Theta\} | \mathbf{M}) \delta(\theta_t^{(i)} - j) \right\} \log f'_j(s_t^{(i)}) \\
 &= \sum_{t=1}^T \sum_{i=1}^L f(S^{(i)}, \theta_t^{(i)} = j | \mathbf{M}) \\
 &\quad \cdot \left\{ -\frac{N}{2} \log 2\pi - \frac{N}{2} \log \sigma_j'^2 - \frac{1}{2} \alpha (\sigma_j'^{-1} s_t^{(i)}; \mathbf{a}_j') \right\} \\
 &\quad \cdot f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M}), \tag{88}
 \end{aligned}$$

an expression similar to (78). Maximizing $Q_{ff}(\mathbf{M}, f'_j)$ with respect to \mathbf{a}_j' is equivalent to minimizing

$$\begin{aligned}
& \sum_{t=1}^T \sum_{i=1}^L f(S^{(i)}, \theta_t^{(i)} = j | \mathbf{M}) [f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M})] \alpha(s_t^{(i)}; \mathbf{a}_j') \\
&= r_{\alpha'}(0) \cdot \left[\sum_{t=1}^T \sum_{i=1}^L f(S^{(i)}, \theta_t^{(i)} = j | \mathbf{M}) [f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M})] r_t^{(i)}(0) \right] \\
&\quad + 2 \sum_{m=1}^M r_{\alpha'}(m) \cdot \left[\sum_{t=1}^T \sum_{i=1}^L f(S^{(i)}, \theta_t^{(i)} = j | \mathbf{M}) \right. \\
&\quad \left. \cdot [f(\{S\} | \mathbf{M}) / f(S^{(i)} | \mathbf{M})] r_t^{(i)}(m) \right], \tag{89}
\end{aligned}$$

where $r_{\alpha'}(m)$ and $r_t^{(i)}(m)$ are the m -lag autocorrelation coefficient of \mathbf{a}_j' and $s_t^{(i)}$ sequences, respectively. Similar to the development in Appendix A, (89) is simply $\alpha(\bar{s}_j; \mathbf{a}_j')$ and (58) must be satisfied in order to maximize $Q_{ji}(\mathbf{M}, f_j')$ with respect to \mathbf{a}_j' . Equation (59) then follows from maximizing (88) with respect to $\sigma_j'^2$ given \mathbf{a}_j' .

AUTHOR

Biing-Hwang Juang, B.Sc. (Electrical Engineering), 1973, National Taiwan University, Republic of China; M.Sc. and Ph.D. (Electrical and Computer Engineering), University of California, Santa Barbara, 1979 and 1981, respectively; Speech Communications Research Laboratory (SCRL), 1978; Signal Technology, Inc., 1979–1982; AT&T Bell Laboratories, 1982; AT&T Information Systems Laboratories, 1983; AT&T Bell Laboratories, 1983—. Before joining AT&T Bell Laboratories, Mr. Juang worked on vocal tract modeling at Speech Communications Research Laboratory, and on speech coding and interference suppression at Signal Technology, Inc. Presently, he is a member of the Acoustics Research department, where he is researching speech communications techniques and stochastic modeling of speech signals.

On the Performance of Isolated Word Speech Recognizers Using Vector Quantization and Temporal Energy Contours

By L. R. RABINER,* K. C. PAN,* and F. K. SOONG*

(Manuscript received February 3, 1984)

In this paper we present results of a series of experiments in which combinations of vector quantization and temporal energy contours are incorporated into the standard framework for the word recognizer. We consider two distinct word vocabularies, namely, a set of 10 digits, and a 129-word airlines vocabulary. We show that the incorporation of energy leads to small but consistent improvements in performance for the digits vocabulary; the incorporation of vector quantization (in a judicious manner) leads to small degradation in performance for both vocabularies, but at the same time reduces overall computation of the recognizer by a significant amount. We conclude that a high-performance, moderate-computation, isolated word recognizer can be achieved using vector quantization and the temporal energy contour.

I. INTRODUCTION

The most popular form for an isolated word recognition system is the classic statistical pattern recognition implementation shown in Fig. 1. In this model the speech signal is first analyzed by the feature measurement block, which produces a test pattern consisting of a temporal sequence of (spectral) feature vectors characteristic of the speech sounds in the word. Most typically, the feature measurement system is either a bank of highly overlapping (in frequency) bandpass filters, or a Linear Predictive Coding (LPC) analysis. In either case

* AT&T Bell Laboratories.

Copyright © 1984 AT&T. Photo reproduction for noncommercial use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

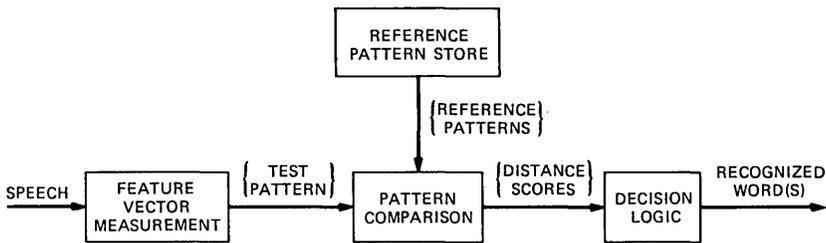


Fig. 1—Pattern recognition model of isolated word recognition system.

the feature vector, for a given time interval, is an estimate of the short-time spectrum of the speech signal. The job of the pattern comparison block is to register, in time, the test pattern with each of a set of stored reference patterns, and to determine a similarity (distance) score for each such pair of patterns. It has been shown that one must use some type of dynamic programming algorithm to achieve the required degree of time alignment for arbitrary word vocabularies.¹ Associated with the time alignment is a spectral distance measure for comparing frames of the test and reference patterns. Such distance measures are often as simple as summing spectral magnitude differences, or as complex as the likelihood ratio measure.² The final stage of the pattern recognition model of Fig. 1 is the decision rule that makes a recognition decision, or possibly a set of decisions, based on the distance or similarity scores provided by the pattern comparison block. The most widely used decision rule is the nearest-neighbor rule, which chooses the recognized word as the one whose pattern has the smallest distance score. Alternative decision rules are variants of the K-Nearest-Neighbor (KNN) rule.³

A wide variety of isolated word recognizers have been designed, based on the structure given in Fig. 1, and have been shown to yield good performance for several types of word vocabularies and talker sets.^{4,5} The major obstacle to the widespread use of such recognizers for simple applications (home computers, terminals, etc.) is the inherent cost of the implementation. This cost, either in terms of computation or actual dollars, is primarily due to the cost of implementing the pattern comparison block with a dynamic programming algorithm. Several alternative recognition structures have been proposed for reducing the cost of the recognizer. These include replacing the non-parametric model of Fig. 1 with a parametric model [e.g., a Hidden Markov Model (HMM)],⁶ using recognition structures without time alignment procedures,⁷ and using some coding technique on the feature vectors to significantly reduce computation in the dynamic programming algorithm.⁸ The first two alternative recognition strategies are still under investigation, but at the present time they yield degraded

performance for several standard vocabularies in the speaker-independent mode. The third alternative is the subject of this paper. Shikano has presented some results on trade-offs between computation and performance achievable using coding techniques. We extend his results and apply them to two useful and interesting word vocabularies, and consider their applicability in a speaker-independent mode.

The actual recognition structure used in this paper is an LPC-based system, which uses the likelihood distance metric in a Dynamic Time Warping (DTW) implementation of the pattern similarity block of Fig. 1. The coding technique used to reduce computation in the DTW algorithm is LPC Vector Quantization (VQ).^{9,10} The way to reduce computation is to replace each feature vector in the reference pattern by one of a set of fixed LPC vectors from a code book (designed from an appropriate training set). If we similarly replace each feature vector of the test pattern by the closest vector in the code book, then, by precomputing the matrix of distances of each code-book vector to every other code-book vector, the distance computation of the DTW algorithm becomes a simple table-lookup operation. Since the distance computation dominates the overall computation of the recognizer, significant reductions in computation are achieved with this technique. It remains to be shown that performance degradation (due to the distortion introduced by vector quantization) can be kept small.

This paper also discusses the application of temporal energy contours to the recognizer structure of Fig. 1. Previous work by Brown and Rabiner¹¹ shows that by treating the energy contour (normalized over the entire word duration) as a new feature, and by incorporating this energy feature into the distance metric as an independent, additive feature, performance of the conventional DTW recognizer was improved. Work by Rabiner et al.¹² shows how vector quantization design algorithms can incorporate energy directly into the standard code-book design procedure, yielding a joint quantization of the LPC vector and its energy value. In this paper we integrate both these results into a common framework. We also implement an isolated word recognizer, incorporating vector quantization to reduce computation and using temporal energy contours to achieve performance comparable to that of the DTW system without using either VQ or energy.

The organization of this paper is as follows. In Section II we describe the implementation of the isolated word recognizer using vector quantization and temporal energy contours. In Section III we describe and give results from a series of evaluation tests on two distinct sets of word vocabularies to show the performance of the overall word recognizer in a speaker-independent mode. In Section IV we discuss the results and compare them to those obtained in other studies of com-

putational reduction techniques. Finally, in Section V we summarize our findings.

II. STRUCTURE OF THE OVERALL ISOLATED WORD RECOGNIZER

Figure 2 is a block diagram of the word recognizer incorporating vector quantization of the LPC-feature vectors and using temporal energy contours. The speech signal, recorded off a dialed-up telephone line and digitized at 6.67-kHz rate, is first blocked into 45-ms frames and analyzed to give LPC vectors every 15 ms (100 samples) using the autocorrelation method. A Hamming window is applied to the 45-ms (300 samples) section of speech, which has been preemphasized using a first-order digital network, and a set of $(p + 1)$ autocorrelations are computed. In our implementation we use $p = 8$ poles for the telephone bandwidth signal. The signal energy (unnormalized) for the l th frame of speech is the zeroth-order autocorrelation, $R_l(0)$. We denote the p th-order LPC vector for the l th frame as \mathbf{a}_l , and the log energy for the l th frame (after normalization, which will be described later) as \hat{E}_l .

The next stage in the processing of Fig. 2 is vector quantization of the LPC vector (with or without the energy parameter). To perform vector quantization, we need a predesigned code book of vectors and an appropriate distance metric for comparing the LPC vectors of the speech signal with the prestored code-book vectors. If we denote an arbitrary test vector as the pair $\hat{\mathbf{a}} = (\mathbf{a}, \hat{E}^T)$, and an arbitrary code-book vector as the pair $\hat{\mathbf{b}} = (\mathbf{b}, \hat{E}^R)$, then an appropriate distance for comparing $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ is¹¹

$$d(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = \left(\frac{\mathbf{b}^t V^T \mathbf{b}}{\mathbf{a}^t V^T \mathbf{a}} - 1 \right) + \alpha f(|\hat{E}^T - \hat{E}^R|), \quad (1)$$

where V^T is the Toeplitz matrix of autocorrelations of the test frame, α is a suitable weighting factor on the energy distance, and $f(x)$ is a nonlinearity of the type

$$f(x) = \begin{cases} 0, & |x| < E_{LO} \\ |x| - E_{LO} + E_{OF}, & E_{LO} \leq x \leq E_{HI} + E_{LO} - E_{OF} \\ E_{HI}, & |x| > E_{HI} + E_{LO} - E_{OF}, \end{cases} \quad (2)$$

where E_{LO} , E_{HI} , and E_{OF} are appropriately chosen thresholds and energy offsets.

The first term in brackets in eq. (1) is the conventional Itakura LPC likelihood ratio (in its linear form),² and the second term is an energy distance that is added to the LPC distance. The weighting factor, α , accounts for the fact that energy distances bear significantly less information than LPC distances. The nonlinear function, $f(x)$,

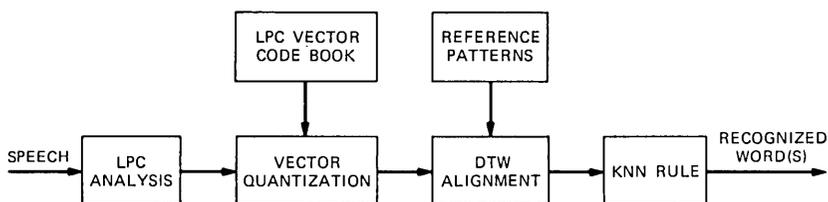


Fig. 2—LPC/DTW recognizer using vector quantization.

accounts for the fact that small energy differences (i.e., less than E_{LO} dB) are insignificant and hence should add no distance, and that large energy distances should be clipped at some appropriate value. Energy distances between these extremes are linearly weighted. In practice, because the Itakura likelihood ratio is essentially unbounded in value, a clipping function, g , is also applied in eq. (1) so that

$$g(x) = \begin{cases} x & x < D_{CLIP} \\ D_{CLIP} & x > D_{CLIP}, \end{cases} \quad (3)$$

where x is the expression in brackets in eq. (1).

A complete specification of the distance metric of eq. (1) requires specification of values for the LPC clipping threshold, D_{CLIP} , and the energy thresholds, E_{LO} , E_{HI} , and E_{OF} . These values are obtained by experimentation in a small pilot test, and specific values will be given in Section III.

Once the distance metric of eq. (1) is given, the implementation of the vector quantization stage of Fig. 2 using a code book with M^* vectors, for the l th feature vector, is a computation of the form

$$d^* = \min_{1 \leq m \leq M^*} d(\hat{\mathbf{a}}_l, \hat{\mathbf{b}}_m) \quad (4a)$$

$$m^* = \arg \left\{ \min_{1 \leq m \leq M^*} d(\hat{\mathbf{a}}_l, \hat{\mathbf{b}}_m) \right\} \quad (4b)$$

$$\hat{\mathbf{a}}_l = \hat{\mathbf{b}}_{m^*}, \quad (4c)$$

i.e., we find the code-book vector $\hat{\mathbf{b}}_{m^*}$ such that its distance to the analysis vector $\hat{\mathbf{a}}_l$ is minimum over all code-book entries, and we replace $\hat{\mathbf{a}}_l$ by $\hat{\mathbf{b}}_{m^*}$. (Equivalently, all we need to save is the index m^* , which gives the code-book vector with the minimum distance, since the code-book vectors are fixed.)

Once a test feature vector has been vector quantized, and similarly, each reference feature vector has been vector quantized, then the calculation of distance between test and reference feature vectors (as required in the DTW alignment procedure) becomes simply a table-lookup procedure from a table of all possible distances between code-book vectors. Thus, if we define the M^* by M^* matrix of code-book

vector distances as

$$D_{CB}(i, j) = d(\hat{\mathbf{b}}_i, \hat{\mathbf{b}}_j), \quad 1 \leq i \leq M^*, \quad 1 \leq j \leq M^*, \quad (5)$$

and precompute $D_{CB}(i, j)$ and store it, then the distance between a test vector coded by code-book vector i^* , and a reference vector coded by code-book vector j^* , is simply $D_{CB}(i^*, j^*)$, and is computed in the time for a single table-lookup. Hence, the number of distance computations [nominally needing about $(p + 1)$ multiplications and additions] is essentially reduced to zero. In this manner the computation of distance from the DTW alignment is substantially reduced. The technique of vector quantizing both the test and reference patterns, and then using the matrix of distances for a table-lookup computation is called a double-SPLIT VQ by Shikano.⁸ For the double-SPLIT method, a full LPC analysis (i.e., the Levinson recursion) does not need to be carried out since the prediction residual is common to all distances in the minimization of eq. (4) and hence, need not be computed.

It should be noted that the process of vector quantization of $\hat{\mathbf{a}}_l$ leads to a distortion error, ϵ_l , given by

$$\epsilon_l = \hat{\mathbf{a}}_l - \hat{\mathbf{b}}_m \quad (6)$$

since the actual feature vector does differ from the code-book vector. One way of avoiding this distortion in the test feature vector, due to Sakoe,¹³ is to save the vector of distances, $\hat{d}(l, m)$, of the form

$$\hat{d}(l, m) = d(\hat{\mathbf{a}}_l, \hat{\mathbf{b}}_m) \quad (7)$$

for all frames, l , of the test pattern, and all code-book indices, m . In this manner whenever a distance is required between the true test feature vector, $\hat{\mathbf{a}}_l$, and a reference vector quantized to code-book vector, $\hat{\mathbf{b}}_q$, then the distance can be looked up in the distance vector for frame l as the q th entry. Thus, we eliminate storage for the M^* by M^* distances of the code-book vectors, but we instead need storage for the M^* by L distances, for a word of L frames, of the vector quantizer. Since $L < M^*$, in most cases, this simplification of the vector quantization generally both increases performance of the recognizer (since no distortion of the test vectors is incurred) and decreases storage of the system. This technique is called a single-SPLIT VQ by Shikano.⁸

The remaining steps in the recognizer of Fig. 2 are essentially those of a conventional DTW-based word recognizer. The DTW alignment compares the test pattern (in some type of VQ format) to each reference pattern (coded as a series of code-book vectors) and generates a distance score. The KNN rule examines the best K scores for each vocabulary word and gives an ordered list of word distances based on the average of the K scores. The "recognized" word is selected as the word whose best- K patterns have the smallest average score.

2.1 Generation of LPC-vector code book

The generation of vectors in the code book of the vector quantizer is straightforward and follows the procedures outlined in Refs. 9, 10, and 12. We used a training set of 39,000 LPC vectors with energy values. The vectors were extracted from isolated digit sequences spoken by 100 talkers (50 male, 50 female). Code books of size $M^* = 2, 4, 8, 16, 32, 64,$ and 128 were generated. All results presented in this paper are for code-book size 128. Results, on digit evaluation tests, for smaller-size code books are given in Ref. 12.

2.2 Normalization of the energy contours for words

The raw energy value for the l th frame of a word, E_l , is computed as

$$E_l = 10 \cdot \log_{10}(R_l(0)), \quad l = 1, 2, \dots, L, \quad (8)$$

where L is the number of frames in the word. The normalization of energy is performed by finding the maximum energy value, E_{MAX} , over the word as

$$E_{\text{MAX}} = \max_{1 \leq l \leq L} (E_l) \quad (9)$$

and by subtracting E_{MAX} from E_l to give

$$\hat{E}_l = E_l - E_{\text{MAX}}. \quad (10)$$

In this manner the peak energy value of each word is 0 dB, and the recognition system is relatively insensitive to differences in gain between recordings. Of course the computation of eq. (9) means that word energy contour normalization cannot take place until the end of the word is located. This constraint poses no real difficulty since there are ways of implementing an approximate gain normalization in "real time" based on some realistic assumptions about the rate of change of system gain.

III. EVALUATION TESTS OF THE RECOGNIZER

To evaluate the effects of the vector quantizer and the use of energy contours on the performance of the isolated word recognizer, we performed a series of three sets of recognition tests. For the first two sets of tests, the word vocabulary was the ten digits (zero through nine), and for the third set of tests, a 129-word airlines vocabulary was used.^{14,15} All tests were conducted in a speaker-independent mode in which all test recordings were made off a standard, dialed-up, local telephone line. A set of 12 speaker-independent reference patterns, obtained from a conventional clustering analysis¹⁶ of 100 tokens per word, were used for each vocabulary word.

We obtained three test sets, denoted TS1, TS2, and TS3, in the following way. TS1 consisted of 100 talkers (50 male, 50 female) who each spoke each digit once. These talkers were the same ones who generated the training tokens used to create both the isolated-digits reference patterns, and the code-book vectors. However, different recordings were used for the training sets than for the test set. The second test set, TS2, consisted of ten talkers (five male, five female) who each spoke each digit 20 times. These talkers were not members of the 100-talker training set and were chosen from a set of 100 talkers used in a large-scale evaluation of a combined digit recognition, talker identification system.¹⁷ The set of ten talkers was chosen on the basis of preliminary experimentation, since they had an error rate somewhat above the average of the 100 talkers used in the experiment. In this manner it was hoped that the TS2 data would amplify differences in test performance results.

The third test set, TS3, consisted of 20 talkers (10 male, 10 female) who each spoke the entire airlines vocabulary a single time. These 20 test talkers were different from those who provided the training tokens used to give the word reference templates.

3.1 Results on digits (TS1)

A series of recognition tests were performed in which temporal energy and vector quantization were tried in all combinations with the basic LPC-based DTW recognizer. A total of six test results are given in Table Ia for the following cases:

Run 1—Standard LPC-based DTW recognizer without energy and without VQ.

Table I—Average digit error rates for the top β word candidates for six runs

Run Number	Energy Used	VQ Ref.	VQ Test	Error Rate (%) for Top β Candidates					
				1	2	3	4	5	6
(a) TS1 data									
1	No	No	No	2.7	0.8	0.2	0.2	0.2	0
2	Yes	No	No	2.1	0.4	0.2	0.2	0.1	0
3	No	Yes	No	3.2	0.9	0.2	0.1	0.1	0
4	Yes	Yes	No	2.4	0.5	0.1	0.1	0	0
5	No	Yes	Yes	4.0	1.1	0.3	0.2	0	0
6	Yes	Yes	Yes	3.5	0.6	0.1	0.1	0.1	0
(b) TS2 data									
1	No	No	No	3.6	1.1	0.3	0.1	0	0
2	Yes	No	No	2.8	1.1	0.5	0.2	0.1	0
3	No	Yes	No	3.8	1.2	0.3	0.2	0.1	0
4	Yes	Yes	No	4.2	1.5	0.7	0.3	0.1	0
5	No	Yes	Yes	4.1	1.1	0.4	0.2	0.1	0
6	Yes	Yes	Yes	4.0	2.0	0.9	0.3	0.1	0.1

Run 2—Energy contour used, but no VQ.

Run 3—Energy contour not used and VQ used only on the reference pattern—i.e., this is a single-SPLIT VQ.

Run 4—Energy contour used combined with VQ only on the reference pattern.

Run 5—Energy contour not used and VQ used on both test and reference patterns—i.e., this is a double-SPLIT VQ.

Run 6—Energy contour used combined with VQ on both test and reference patterns.

For each of the energy-based runs, the parameters of the energy distance were set to

$$E_{LO} = 6(\text{dB}), \quad E_{HI} = 20(\text{dB}), \quad E_{OF} = 0(\text{dB}),$$

and the LPC distance-clipping threshold, D_{CLIP} , was set to 2.5. (Some experimentation was done with values of $E_{OF} = 6$ dB, as used in Ref. 12, but results were almost always worse using this parameter setting because of the sensitivity of the DTW path to matching energy contours with the 6-dB energy offset).

An examination of the results given in Table Ia, which gives digit error rates in percent for the top β word candidates ($\beta = 1$ to 6) shows the following:

1. For each of the three consecutive pairs of runs (where each pair differs only in regard to the inclusion of the temporal energy contour), the inclusion of energy reduces the error rate in the top candidate by about 0.6 percent (± 0.1 percent).

2. Applying VQ to the reference alone (Runs 3 and 4) increases the average digit error rate in the top candidate by about 0.4 percent. However, using energy, the error rate in the top candidate (2.4 percent) is still below the error rate for Run 1, the standard DTW recognizer without energy or VQ.

3. Applying VQ to both test and reference patterns (Runs 5 and 6) increases the average digit error rate in the top candidate by about 1.3 percent over that for Runs 1 and 2. In these cases the performance is degraded from that of the recognizer without either temporal energy or VQ.

3.2 Results on digits (TS2)

The results for the same six runs using the 2000 digits of TS2 are given in Table Ib. For this set of data the average digit error rate for Run 1 is about 1 percent higher than for TS1 data. This is due to the inclusion of talkers in the database with higher than average error rates. When energy is included in the recognizer (without VQ), the average top candidate error rate falls by 0.8 percent.

The results of Runs 3 and 4 show that using VQ on the reference

patterns alone leads to a small increase in error rate (0.2 percent) for the case without energy and a larger increase in error rate (1.4 percent) for the case with energy. In these two cases, the runs using the energy contour provided essentially the same performance as the run without the energy contour.

Runs 5 and 6 show a slight increase in error rate for the case without energy (Run 5 compared to Run 3) and a slight decrease in error rate for the case with energy (Run 6 compared to Run 4). For these two cases, the performance is essentially identical.

A summary set of curves showing the error rate versus candidate position, β , for the best sets of results of Table I is given in Fig. 3, where, for each consecutive pair of runs, we have plotted the best results. These sets of curves show the slight degradations introduced by applying a VQ to the reference alone and to both the reference and test patterns.

3.3 Results on airlines words (TS3)

For the airlines vocabulary a set of four runs were made. These runs correspond to the first four runs on the digits vocabulary. No tests were made with VQ of both test and reference patterns for this vocabulary. The results of the four runs are given in Table II and plotted in Fig. 4.

The results show that using energy contours for this medium-size, complex vocabulary led to essentially no significant improvement in performance for either of the pairs of runs. For the case of no VQ of the references, the performance with energy was 0.2 percent worse than without energy; for the case of using VQ on the references, the performance with energy was 0.5 percent better than without energy.

It can also be seen that using VQ of the references led to a 4- to 4.5-percent increase in error rate for the top candidate and somewhat smaller increases for higher-position candidates. These results indicate that a VQ with 128 code-book entries is just too small for a vocabulary of this size and complexity.

IV. DISCUSSION OF RESULTS

The results presented in Section III showed the following:

1. The addition of the temporal energy contour as an additive feature to the LPC vector generally improved the performance of the recognition system by a small amount. This result was more the case for the digits vocabulary than for the airlines vocabulary.

2. For the digits vocabulary, using VQ on just the reference pattern (the single-SPLIT case) slightly increased the error rate; for the airlines vocabulary a significant increase in error rate occurred, indi-

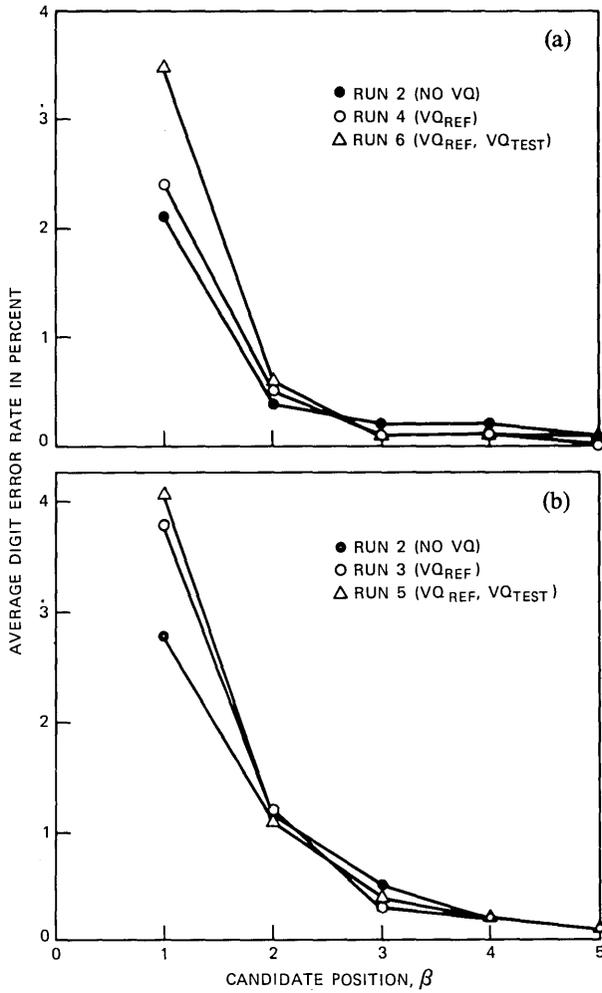


Fig. 3—Average word error rate versus candidate position for the digits vocabulary for (a) TS1 and (b) TS2.

Table II—Average word error rates (percent) for tests on airline vocabulary

Run Number	Energy Used	VQ Ref.	Error Rate (%) for Top β Candidates				
			1	2	3	4	5
1	No	No	10.0	4.5	2.9	2.2	1.6
2	Yes	No	10.2	4.2	2.4	1.9	1.6
3	No	Yes	14.5	6.0	3.5	2.6	2.1
4	Yes	Yes	14.0	6.3	4.0	3.0	2.2

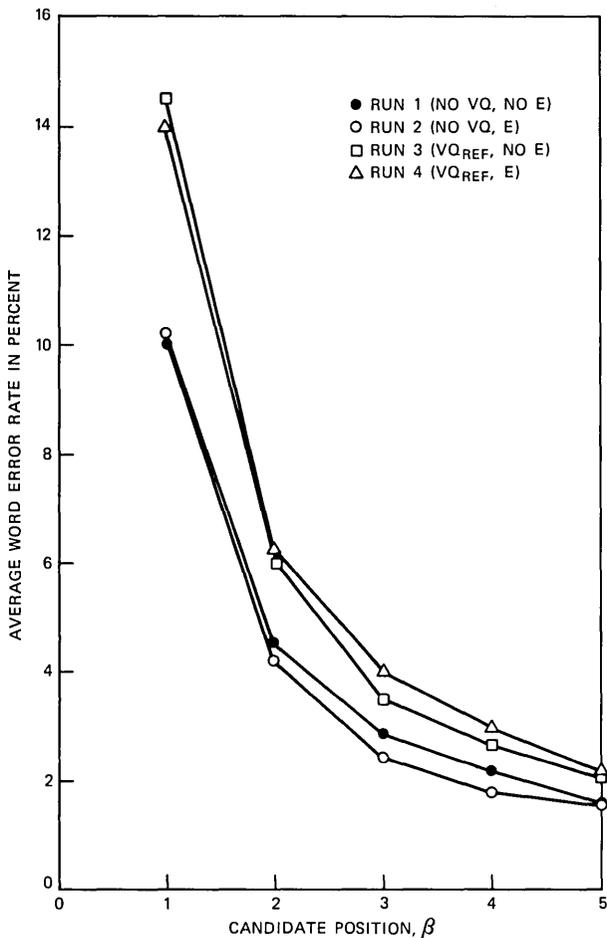


Fig. 4—Average word error rate versus candidate position for the airlines vocabulary.

cating that the size of the VQ was too small for the size and complexity of the vocabulary.

3. For the digits vocabulary, using VQ on both the test and reference patterns (the double-SPLIT case) increased the error rate to a larger degree.

These results indicate that using VQ on just the reference pattern, combined with using the energy contour as an additional feature, can lead to a recognition system with only marginally poorer performance than the system without VQ (at least for the digits vocabulary), and we assert that if a large enough VQ were used for the airlines vocabulary (e.g., M^* on the order of 512), similar results would have been attained.

The basic results are consistent with the findings of Shikano,⁸ who studied a speaker-trained system with a large vocabulary (641 words) of Japanese city names, and with earlier work on the digits vocabulary.¹²

If we compare the results reported here to alternative computationally efficient approaches—such as the method proposed by Shore and Burton,⁷ or the Hidden Markov Model (HMM) approach¹²—we see that the performance of the DTW approach using a VQ and temporal energy is significantly better than the alternatives. Shore and Burton presented results on a 20-word vocabulary (consisting of the digits and ten control words) for an eight-male talker population and had a 12-percent word error rate; for the digits vocabulary the error rate was still 4.1 percent. Rabiner et al.¹² reported error rates of about 3.5 percent for TS1 data using the HMM approach on a digits vocabulary, and about 15 percent on the airlines vocabulary. Thus, at the current time, since the computation of the LPC recognizer with DTW processing using energy and VQ is comparable to that of alternative approaches, and since its performance is better, it is the most attractive proposal for significant reductions in computation in an isolated word recognizer.

A key issue when using a VQ in the recognizer is the savings in computation over the conventional DTW approach without VQ. To quantify this concept, we define the following terms:

M^* = Number of vectors in code book

V = Number of vocabulary words

Q = Number of reference templates per vocabulary word

\bar{L} = Average number of frames in a word

p = Order of LPC analysis.

For the conventional DTW approach, the computation in each DTW (for each reference pattern) is approximately:

$$\bar{C}_{\text{DTW}} = \bar{C}_{\text{DIST}} + \bar{C}_{\text{COMB}} \quad (11a)$$

$$\bar{C}_{\text{DIST}} = \frac{\bar{L}^2}{3} \cdot (p + 1)(*, +), \quad (11b)$$

where \bar{C}_{DIST} is the computation for distances in the DTW, and \bar{C}_{COMB} is the computation for combinatorics. In a serial processor the computation for combinatorics is on the order of one-fifth the computation for distances. Hence, a good approximation is

$$\bar{C}_{\text{DTW}} = \frac{6}{5} \cdot \frac{\bar{L}^2}{3} (p + 1)(*, +) \quad (12)$$

per DTW, or a total of

$$\begin{aligned}
 C_{DTW} &= \bar{C}_{DTW} \cdot V \cdot Q \\
 &= \frac{2}{5} \bar{L}^2 (p + 1) V \cdot Q(*, +)
 \end{aligned}
 \tag{13}$$

to recognize a test word.

Using the VQ in the recognizer leads to a front-end computation load of

$$C_{VQ} = M^* \bar{L} (p + 1) (*, +) \tag{14}$$

to code the \bar{L} frames of the test, and a reduction of computation in the DTW to

$$C_{DTW/VQ} \cong \frac{\bar{L}^2}{15} (p + 1) V \cdot Q(*, +) \tag{15}$$

since all distance computation is eliminated.

The ratio of computation, R , of the DTW with VQ to the DTW without VQ is given as

$$R = \frac{C_{DTW/VQ} + C_{VQ}}{C_{DTW}} \tag{16a}$$

$$\begin{aligned}
 &= \frac{\frac{\bar{L}}{15} \cdot V \cdot Q + M^*}{\frac{2}{5} \bar{L} \cdot V \cdot Q}
 \end{aligned}
 \tag{16b}$$

For the digits vocabulary, with $\bar{L} = 40$, $M^* = 128$, $p = 8$, $V = 10$, and $Q = 12$, we get

$$R_{DIGITS} = \frac{(128 + 320)}{1920} \approx 0.233,$$

i.e., a 4.3 to 1 reduction in computation. For the airlines vocabulary, we get about a 5.5 to 1 reduction in computation (even if we make $M^* = 512$). Hence, we conclude that the inclusion of VQ in the DTW-based word recognizer can indeed significantly reduce the computation without significantly lowering recognizer performance.

V. SUMMARY

In this paper we show that by adding a vector quantization stage to the standard DTW-based isolated word recognizer, and by incorporating temporal energy as an additional feature to the LPC vector, a high-performance, yet significantly reduced computation word recognizer can be implemented. By using the so-called single-SPLIT methods, in which the VQ is only directly applied to the reference patterns,

we show that the resulting VQ distortion can be made sufficiently small such that only an insignificant increase in word error rate results.

REFERENCES

1. G. M. White and R. B. Neely, "Speech Recognition Experiments With Linear Prediction, Bandpass Filtering, and Dynamic Programming," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-24*, No. 2 (April 1976), pp. 183-8.
2. F. I. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-23*, No. 1 (February 1975), pp. 67-72.
3. L. R. Rabiner, S. E. Levinson, A. E. Rosenberg, and J. G. Wilpon, "Speaker-Independent Recognition of Isolated Words Using Clustering Techniques," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-27*, No. 4 (August 1979), pp. 336-49.
4. G. R. Doddington and T. B. Schalk, "Speech Recognition: Turning Theory to Practice," *IEEE Spectrum*, 18 (September 1981), pp. 26-32.
5. L. R. Rabiner and S. E. Levinson, "Isolated Word Connected Word Recognition—Theory and Selected Applications," *IEEE Trans. Commun., COM-29*, No. 5 (May 1981), pp. 621-59.
6. L. R. Rabiner, S. E. Levinson, and M. M. Sondhi, "On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition," *B.S.T.J.*, 62, No. 4 (April 1983), pp. 1075-105.
7. J. E. Shore and D. K. Burton, "Discrete Utterance Speech Recognition Without Time Alignment," *IEEE Trans. Inform. Theory, IT-29*, No. 4 (July 1983), pp. 473-91.
8. K. Shikano, "Spoken Word Recognition Based Upon Vector Quantization of Input Speech," *Trans. Comm. Speech Res.*, (December 1982), pp. 473-80.
9. Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantization," *IEEE Trans. Commun., COM-28*, No. 1 (January 1980), pp. 84-95.
10. B. Juang, D. Wong, and A. H. Gray, Jr., "Distortion Performance of Vector Quantization for LPC Voice Coding," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-30*, No. 2 (April 1982), pp. 294-303.
11. M. K. Brown and L. R. Rabiner, "On the Use of Energy in LPC-Based Recognition of Isolated Words," *B.S.T.J.*, 61, No. 10 (December 1982), pp. 2971-87.
12. L. R. Rabiner, M. M. Sondhi, and S. E. Levinson, "A Vector Quantizer Combining Energy and LPC Parameters and Its Application to Isolated Word Recognition," *AT&T Bell Lab. Tech. J.*, 63, No. 5 (May-June 1984), pp. 721-35.
13. H. Sakoe, "Device for Recognizing an Input Pattern With Approximate Patterns Used for Reference Patterns on Mapping," U.S. Patent 4, 256, 924, issued March 17, 1981.
14. S. E. Levinson, A. E. Rosenberg, and J. L. Flanagan, "Evaluation of a Word Recognition System Using Syntax Analysis," *B.S.T.J.*, 57, No. 5 (May-June 1978), pp. 1619-26.
15. J. G. Wilpon, L. R. Rabiner, and A. Bergh, "Speaker-Independent Isolated Word Recognition Using a 129-Word Airline Vocabulary," *J. Acoust. Soc. Amer.*, 72, No. 2 (August 1982), pp. 390-6.
16. S. E. Levinson, L. R. Rabiner, A. E. Rosenberg, and J. G. Wilpon, "Interactive Clustering Techniques for Selecting Speaker-Independent Reference Templates for Isolated Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing, ASSP-27*, No. 2 (April 1979), pp. 134-41.
17. A. E. Rosenberg, K. L. Shipley, and D. E. Bock, "A Speech Data Base Facility Using a Computer Controlled Cassette Tape Deck," *J. Acoust. Soc. Amer., Suppl. 1*, 72, (Fall 1982), p. 580.

AUTHORS

Kok-Chin Pan, S.B. and S.M., 1984 (Electrical Engineering and Computer Science), The Massachusetts Institute of Technology. From 1981 to 1984, Mr. Pan participated in a cooperative program in Electrical Engineering and Computer Science at AT&T Bell Laboratories. He worked on digital circuit design and vector quantization applied to speech recognition. He is currently a Teaching Assistant in the department of Electrical Engineering and Com-

puter Science at MIT. His interests include speech processing and recognition and VLSI system design. Member, Tau Beta Pi.

Lawrence R. Rabiner, S.B. and S.M., 1964, Ph.D., 1967 (Electrical Engineering), The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1962—. Presently, Mr. Rabiner is engaged in research on speech communications and digital signal processing techniques. He is coauthor of *Theory and Application of Digital Signal Processing* (Prentice-Hall, 1975), *Digital Processing of Speech Signals* (Prentice-Hall, 1978), and *Multirate Digital Signal Processing* (Prentice-Hall, 1983). Member, National Academy of Engineering, Eta Kappa Nu, Sigma Xi, Tau Beta Pi. Fellow, Acoustical Society of America, IEEE.

Frank K. Soong, B.S., 1973, National Taiwan University, M.S., 1977, University of Rhode Island, Ph.D., 1983, Stanford University, all in Electrical Engineering; AT&T Bell Laboratories, 1982—. From 1973 to 1975 Mr. Soong served as a teacher at the Chinese Naval Engineering School at Tsoying, Taiwan. In 1982 he joined the technical staff at AT&T Bell Laboratories, where he engaged in research in speech coding and recognition. Member, IEEE.

The Overload Performance of Engineered Networks With Nonhierarchical and Hierarchical Routing

By J. M. AKINPELU*

(Manuscript received December 10, 1982)

We report the results of a study of the performance of engineered nonhierarchical and hierarchical routing networks under overloads. This study was motivated by results obtained from mathematical models for small, symmetric, uniformly loaded, nonhierarchical networks with transparent switching systems, showing the existence of network instabilities. We extend the mathematical models to more general nonhierarchical networks, and show with analysis and an extant simulation model that such instabilities are also found in nonsymmetric, nonhierarchical networks. We then use our models to consider whether engineered nonhierarchical networks exhibit such unstable behavior. No instabilities are found in the engineered nonhierarchical networks considered here. However, the nonhierarchical networks consistently demonstrate a drop in carried load between 10- and 15-percent overloads. Our analysis of comparably engineered hierarchical networks shows that these networks do not exhibit a drop in carried load under overloads (in the absence of switching system dynamics). Finally, we show that using trunk reservation for first-routed traffic allows the formulation of a control strategy that provides a high level of network carried load during overloads.

I. INTRODUCTION

Hierarchical routing has been used since the early days of the toll network. Before the early 1960s, hierarchical routing was very advantageous for a number of reasons. First, it allowed switching systems

* AT&T Bell Laboratories.

Copyright © 1984 AT&T. Photo reproduction for noncommercial use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

to determine the path for a call very simply and quickly, using only the call's destination. Furthermore, the call did not loop back to a switching system previously traversed. Second, hierarchical routing combined small traffic parcels on efficient final trunk groups. Third, hierarchical routing networks were relatively easy to engineer.¹

Recent advances in switching and signaling technology, as well as the tremendous growth in toll traffic, have provided new incentives to increase network design efficiency. Since the mid-1970s, efforts have been under way at AT&T Bell Laboratories to develop a new method of engineering large-scale nonhierarchical networks. These efforts culminated in the unified algorithm,² which takes advantage of traffic noncoincidence and routes calls over least-cost paths. This results in nonhierarchical networks, which are less expensive than hierarchical networks. In the nonhierarchical networks, a call can theoretically use any path connecting its origination and destination (although in the unified algorithm only 1- and 2-link paths are allowed). A call blocked at an intermediate switching system is cranked back to its origin so it can take the next path in its route. In addition, the unified algorithm allows routing to take advantage of time-sensitive load variations.³

The nonhierarchical networks give service comparable to that of hierarchical networks under engineered traffic conditions. However, it was necessary to promote a better understanding of the performance of nonhierarchical networks under other traffic conditions. In particular, there was concern that network instabilities that existed in small symmetric nonhierarchical networks might also exist in nonhierarchical networks engineered using the unified algorithm.

To help provide this understanding, we studied the performance of engineered nonhierarchical and hierarchical networks under general network overloads, using mathematical and simulation models. We also investigated the effect on these networks of one control, namely, trunk reservation for first-routed traffic.

II. BACKGROUND

The concern about the stability of nonhierarchical networks was stimulated by the work of Krupp,⁴ and Nakagome and Mori.⁵ They carried out approximate analyses of nonhierarchical routing applied to small, uniformly loaded networks, which are easily analyzed because of their simple symmetric designs. Krupp also considered a nonsymmetric 3-node model. Their models did not include switching system dynamics. Their analyses revealed, in some cases, the existence of network instabilities when trunk-group (or network) blocking probability is considered as a function of offered load. That is, for certain loads the network has two realizable states: (1) a low network blocking state, in which almost all calls use their shorter first-choice path; and

(2) a congested state, in which a large proportion of calls use longer alternate paths, and many calls are blocked.

For example, Fig. 1 shows the relation between the offered load (per point-to-point pair) and trunk-group blocking obtained for a 10-node symmetric network with 100 trunks per trunk group, using their analyses. Each point-to-point pair has a direct first-choice path and eight 2-link alternate paths. The curve shows a range of offered loads that correspond to multiple blocking probabilities, indicating the potential for unstable behavior in this load range. To understand the instability more clearly, we used a simulation model developed by Krupp⁴ to simulate the 10-node network.

We first observed the number of calls carried by the network as offered load increased. Figure 2 shows the results of a simulation of the network with an initial load of 84 erlangs (starting with an empty network). The load was increased to 85 erlangs after 65 holding times. These loads are near the top of the multiple-valued region in Fig. 1. As Fig. 2 illustrates, a drastic change in carried load occurs after the offered load is increased. With the initial load, there is a fairly constant throughput of approximately 3750 calls. The trunk-group blocking is low, and few calls are alternate-routed. After the load change, the number of alternate-routed calls increases—first only slightly, and

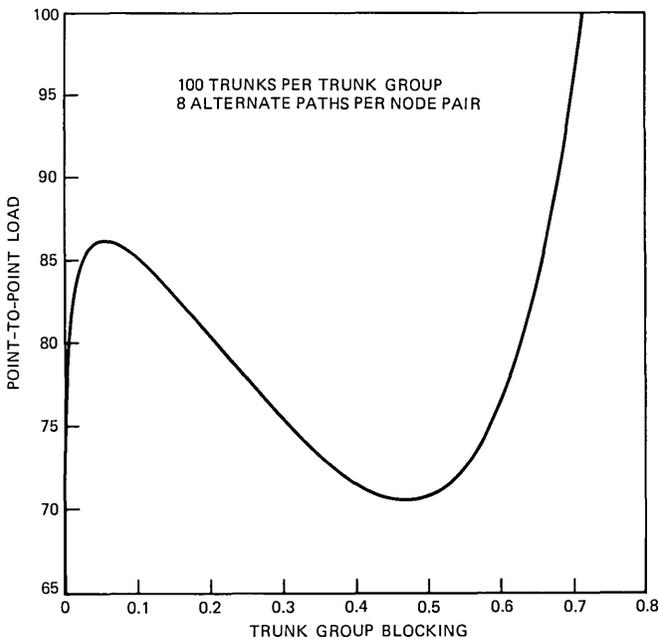


Fig. 1—Performance of a 10-node symmetric network.

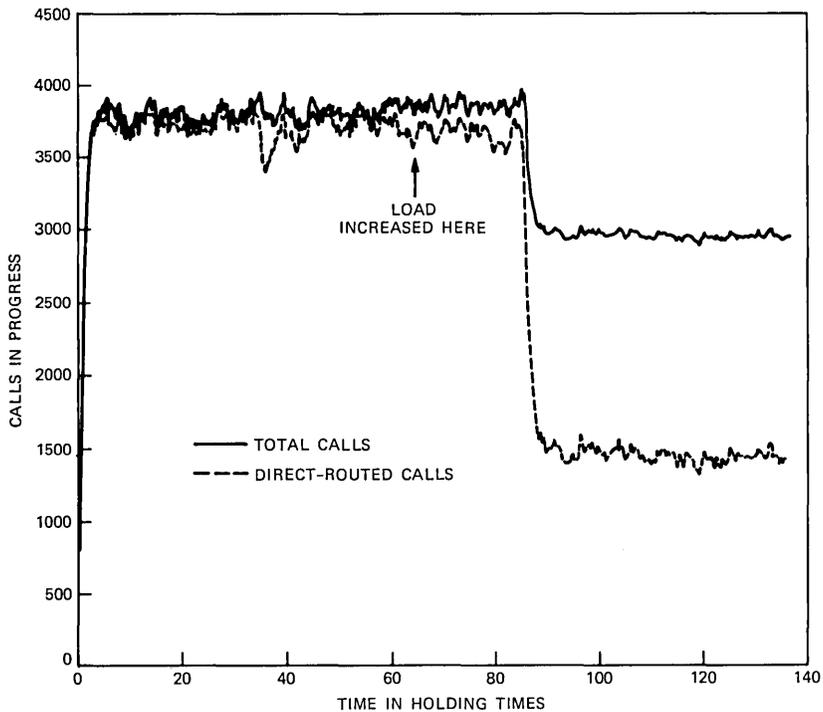


Fig. 2—Simulation of a 10-node symmetric network with load = 84 erlangs/pair; load increased to 85 erlangs/pair.

then dramatically. The average number of calls carried falls to 3000 and half the carried calls are alternate-routed, indicating a high trunk-group blocking. This agrees very closely with the predicted load at which the transition from low to high blocking should occur as offered load increases.

Another consequence of the mathematical solution is illustrated in Fig. 3, which shows the results of a simulation run for the same network with an initial point-to-point offered load of 90 erlangs, starting with an empty network. After approximately 14 holding times, the load is dropped to 80 erlangs. This simulation was conducted to determine the behavior of the network if a load in the multiple-valued region is offered while the network is congested. The initial load of 90 erlangs was used to congest the network, with a resulting carried load of only 3000 calls. When the load is dropped to 80 erlangs, congestion persists for another 16 holding times before the carried load increases to 3600 calls, corresponding to operation in an uncongested state.

These results raised interest in the performance of nonhierarchical routing in more general networks. To address this question, we developed a mathematical model that extends the models in Refs. 4 and 5

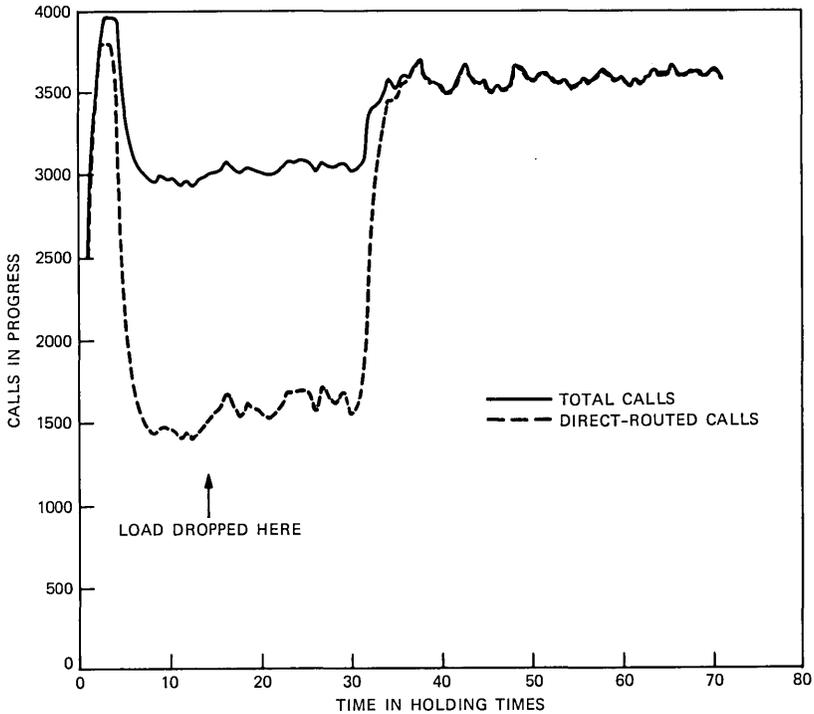


Fig. 3—Simulation of a 10-node symmetric network with load = 90 erlangs/pair; load dropped to 80 erlangs/pair.

to more general nonhierarchical networks. This model, described in the Appendix, analyzes network performance both without controls and with trunk reservation. We also used a simulation model for nonhierarchical networks developed by Krupp. The mathematical model is solved iteratively, starting with an initial estimate of the trunk-group blocking probabilities, to determine the trunk-group offered loads and blocking probabilities in equilibrium. The presence of network instabilities is demonstrated by the existence of multiple solutions for the same parameters (loads, trunk-group sizes, routing) obtained by using different initial estimates of the trunk-group blocking probabilities.

To establish the existence of network instabilities in more general nonhierarchical networks, we applied our models to an 8-node, non-symmetric, nonuniformly loaded, nonhierarchical network. The number of trunks in each trunk group ranged fairly uniformly from 50 to 995 trunks. It should be noted that this was not an engineered network. Point-to-point loads were chosen so that each point-to-point pair would experience a blocking no greater than 0.005 on the direct trunk group in the absence of alternate-routing, and an *overload* means

additional load above these loads. Each point-to-point pair was given a direct first-choice path and four 2-link alternate paths.

We considered the performance of this network under general overloads of up to 10 percent, as shown in Fig. 4. The function relating percent overload to network blocking is double-valued for overloads of up to 4 percent, indicating the existence of instabilities for loads in this range. (Solutions corresponding to low network blocking were obtained by starting with low trunk-group blocking estimates, while solutions corresponding to high network blocking were obtained by starting with high trunk-group blocking estimates.) These results were substantiated by simulation. For example, Fig. 5 shows a simulation at 1-percent overload, starting with an empty network. Initially, the network experiences low blocking, with carried load at about 13,500. Then, after about 17 holding times, the network enters the congested state; carried load drops to 11,000, and a large proportion of calls are alternate-routed.

Now that we had developed a methodology, and demonstrated that instabilities occur in more general nonhierarchical networks, we were ready to address the following question: Does the type of instability seen in these small nonhierarchical networks occur in engineered, nonhierarchical networks? To answer this question, we applied our models to three representative network models. These models are described in Section III.

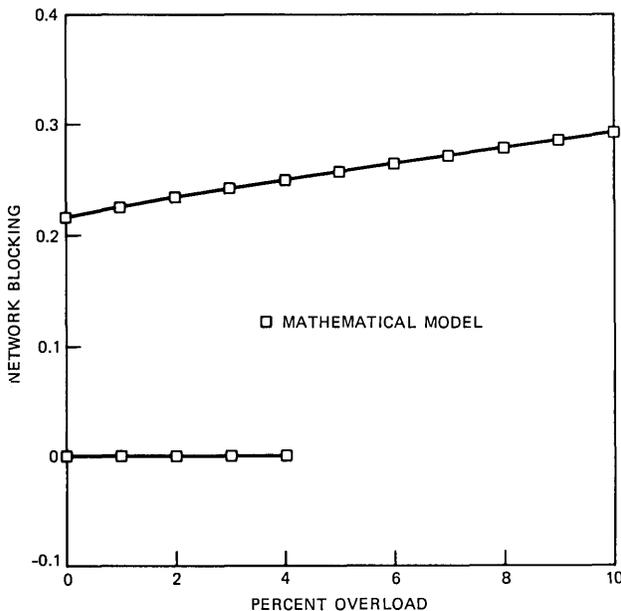


Fig. 4—Network blocking for an 8-node nonhierarchical network.

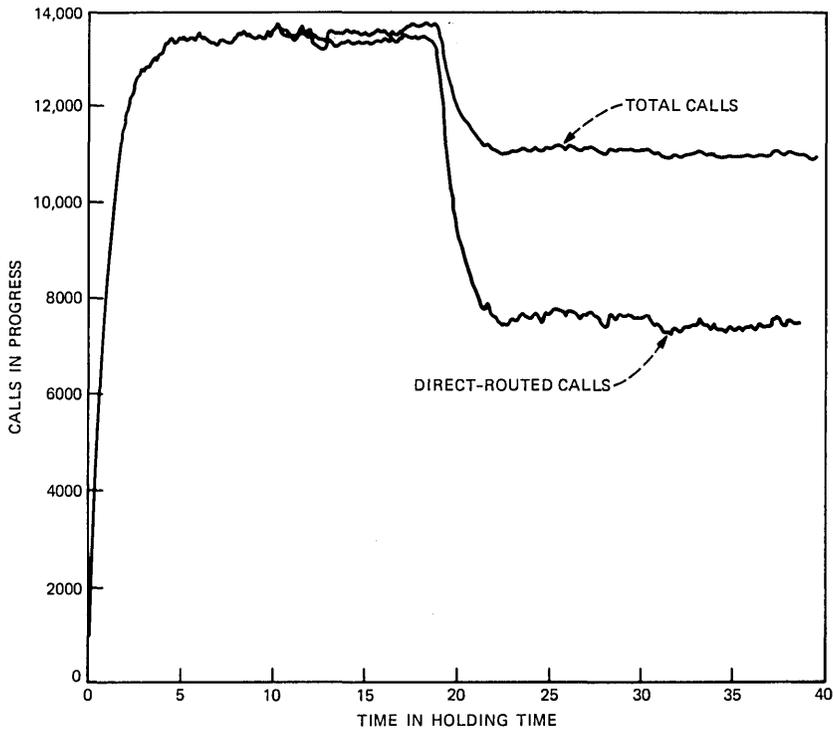


Fig. 5—Simulation of an 8-node network at 1-percent overload.

III. NETWORK MODELS

Three network models were used in our study. The 30-node network model was developed by Gechter and Modarressi at AT&T Bell Laboratories to represent a 4ESS* network. It consists of the 10 regional centers and 20 of the sectional centers in the then existing hierarchical network, and is based on 1977-1978 Trunk Servicing System data. Other characteristics of the network include a wide geographic dispersion of the switching systems, large point-to-point traffic parcels, and fairly uniformly distributed point-to-point loads.

The 25-node network model is made up of a subset of the switching systems that comprise the 215-node hybrid network model recently developed at AT&T Bell Laboratories. The 215-node network was engineered using loads for October 1989, projected from 1978 Centralized Message Data System data. This network includes the 140 4ESSs planned for deployment, as well as 75 stored-program-controlled toll tandem switching systems.

* Trademark of AT&T Technologies, Inc.

The 140-node network model consists of the 140 *4ESS*s in the 215-node network.

Versions of the two smaller network models engineered for both nonhierarchical and hierarchical routing were studied, while the 140-node network model was only considered with nonhierarchical routing. The 25- and 140-node networks were designed using the unified algorithm.² The 30-node nonhierarchical network was engineered using another nonhierarchical network design developed at AT&T Bell Laboratories. An average of 7-percent reserve capacity was added to the trunk groups of the nonhierarchical networks. For the 30- and 25-node network models, the nonhierarchical network has about 9 percent fewer trunks than its corresponding hierarchical network. The nonhierarchical 25-node network was the initial choice for the 1986 Phase One deployment of dynamic nonhierarchical routing (DNHR). The 140-node nonhierarchical network represented the 1989 DNHR network envisioned in the predivestiture environment.

Comparisons of the trunk-group sizes (number of trunks per trunk group) and route sizes (number of paths per route) for the 25-, 30-, and 140-node nonhierarchical networks are shown in Tables I and II. The 25- and 30-node nonhierarchical networks bound the 140-node nonhierarchical network both in terms of trunk-group sizes and route sizes.

IV. RESULTS WITHOUT CONTROLS

4.1 Introduction

We applied the mathematical and simulation models to the 30-, 25-, and 140-node networks. Network performance was investigated under both uniform and nonuniform general overloads. All three network models were studied under uniform general overloads, which were obtained by multiplying the engineered loads by a constant factor.

Table I—Trunk-group size distributions for the nonhierarchical networks

<i>x</i>	Proportion of Trunk-Group Sizes $\leq x$		
	25-Node Network	30-Node Network	140-Node Network
100	0.346	0.928	0.746
200	0.580	0.986	0.868
300	0.708	1.000	0.913
400	0.786		0.938
500	0.847		0.951
600	0.871		0.960
700	0.902		0.967
800	0.932		0.972
900	0.939		0.977
1000	0.949		0.980

Table II—Route size distributions for the nonhierarchical networks

x	Proportion of Route Sizes $\leq x$		
	25-Node Network	30-Node Network	140-Node Network
1	0.213	0.012	0.028
2	0.385	0.021	0.100
3	0.584	0.045	0.242
4	0.791	0.101	0.407
5	0.929	0.196	0.565
6	0.976	0.297	0.718
7	0.993	0.377	0.816
8	1.000	0.441	0.883
9		0.524	0.926
10		1.000	0.971

In addition, the 30-node network was studied under nonuniform general overloads, which were thought to emulate a more realistic overload situation. To represent the nonuniform overloads, we used a set of high-day loads for the 30-node network. The high-day loads were generated using the distribution derived for the ratio of the high-day loads to the average-business-day (engineered) loads and a Gamma distribution to achieve the proper ratio of the peak-day loads to the average of the ten-high-day loads. The high-day loads, which are on average 5 percent higher than the engineered loads, were used to represent a 5-percent overload. Other overloads were obtained by multiplying these loads by the appropriate factor. Uniform and non-uniform general overloads of up to 200 percent were considered. As we discuss below, the mathematical and simulation models give excellent agreement, indicating that the assumptions in the mathematical model do not distort the fundamental network behavior.

4.2 The 30-node network

Figure 6 presents the results for the 30-node networks under uniform overloads. The symbols in the figures indicate results obtained from the mathematical and simulation models. For a given set of loads, the mathematical model always converged to the same solution, regardless of the initial trunk-group blocking estimates, indicating that no network instabilities exist. This is demonstrated by the single-valued function in Fig. 6. However, both the mathematical and simulation results show a striking difference in the performance of the nonhierarchical and hierarchical networks. The two networks show similar performance up to about a 10-percent overload, with carried load increasing with increasing offered load. At that point the number of calls carried in the nonhierarchical network falls sharply, because of an increase in the number of multilink calls. The drop continues until around 100-percent overload, where the carried load begins to increase

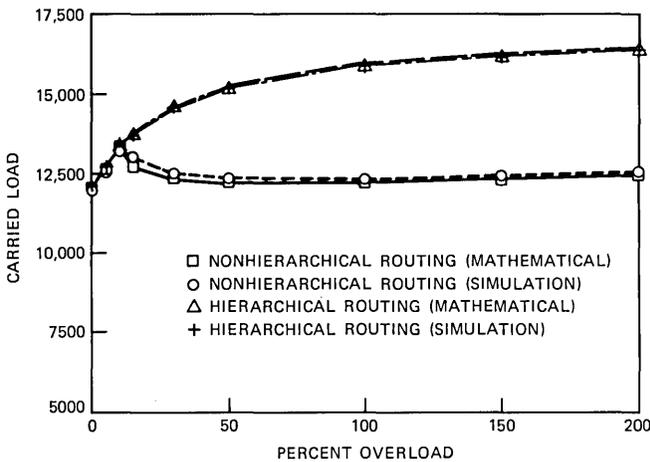


Fig. 6—Performance of the 30-node network under uniform overloads.

gradually. This increase results because the network has become congested to the point that the probability of finding available trunks for a multilink call is very small, so that now 1-link calls begin to be favored over multilink calls. Such behavior is not seen in the hierarchical network, because, intrinsically, it can better limit the number of multilink calls under overloads. This is due to the presence of final trunk groups and of primary high-usage trunk groups that do not carry alternate-routed traffic, the absence of cranking back, and more trunks than are in the nonhierarchical network. The number of calls carried in the hierarchical network increases steadily as offered load increases over the entire range of overloads considered.

Figure 7 demonstrates this point. Here we have plotted the ratio of the number of multilink calls to the number of 1-link calls as determined from the mathematical model for various overloads. This ratio grows sharply for the nonhierarchical network for overloads of up to 50 percent, then levels off and begins to decline. On the other hand, for the hierarchical network, this ratio rises slowly before leveling off at around 30-percent overload. These results are consistent with a simulation study by Weber using 3-, 4-, 5-, and 6-node networks, which showed that hierarchical networks perform more efficiently under overloads than nonhierarchical networks.⁶

Figure 8 displays the performance of the 30-node network under the nonuniform overloads. Qualitatively, the results are the same as those obtained under uniform overloads. In fact, the difference in the number of calls carried under uniform and nonuniform overloads is very small at the higher overloads. The nonhierarchical network under nonuniform overloads also exhibits a drop in carried load at about 10-percent

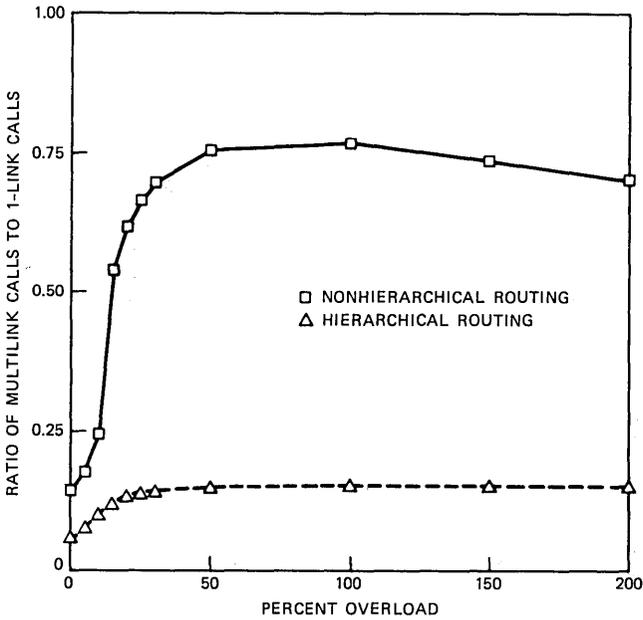


Fig. 7—Ratio of multilink calls to 1-link calls for the 30-node network.

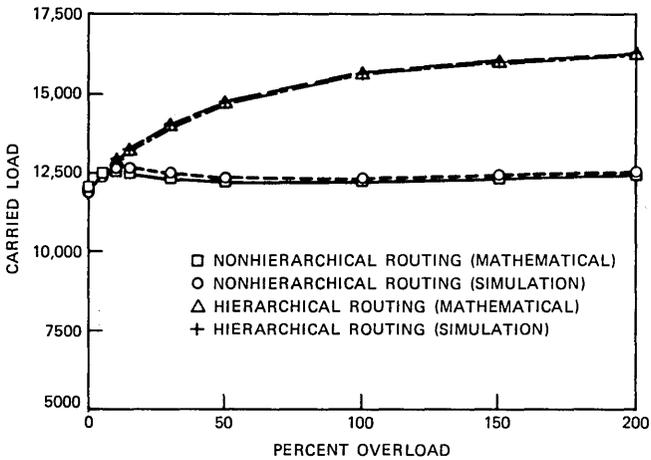


Fig. 8—Performance of the 30-node network under nonuniform overloads.

overload, though not as severe. The drop is attenuated because, with the nonuniform overloads, carried load cannot increase to the maximum level seen with the uniform overloads.

4.3 The 25-node network

Figure 9 presents the results for the 25-node network under uniform

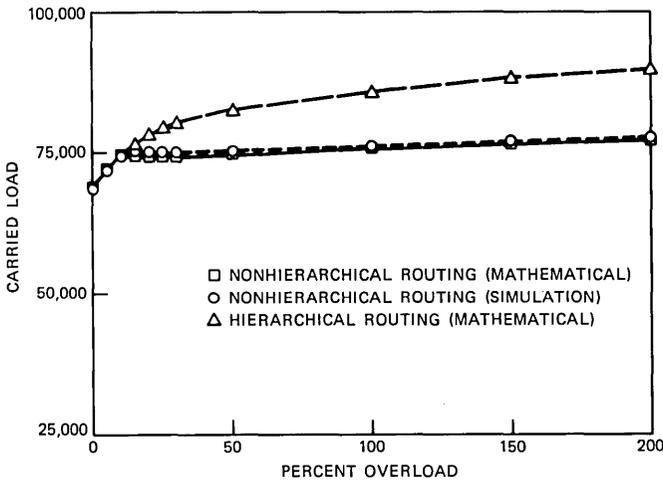


Fig. 9—Performance of the 25-node network under uniform loads.

overloads. As with the 30-node network, the 25-node nonhierarchical network shows a drop in carried load: at about 10-percent overload in the mathematical model and at about 15-percent overload in the simulations. However, the drop is not as sharp as that seen in the 30-node nonhierarchical network. We attribute this to the smaller number of paths per route in the 25-node nonhierarchical network (see Table II), which limits the potential for a large number of multilink calls. The hierarchical network exhibits a continuous increase in carried load with increasing offered load.

4.4 The 140-node network

Figure 10 displays the performance of the 140-node nonhierarchical network under uniform overloads. These results were obtained from the mathematical model. No simulations of this network were made because of its large size. The results agree qualitatively with those derived for the 30- and 25-node nonhierarchical networks. Again, carried load declines at around 10-percent overload and increases at the larger overloads. Figure 10 also shows the number of 1-link calls in the network as derived from the mathematical model. The direction of change in the number of 1-link calls is almost always the same as the direction of change in the total number of calls. Again we conclude that the degree to which network capacity is efficiently used is related to the network's ability to favor 1-link calls over multilink calls.

4.5 Comparison of the nonhierarchical networks

Our analysis of the engineered nonhierarchical networks without controls shows that the type of instability discussed in Section II does

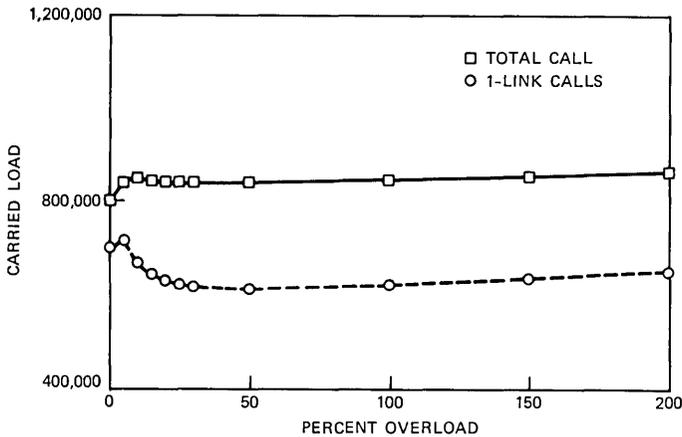


Fig. 10—Performance of the 140-node network under uniform overloads.

not occur—the offered load versus carried load functions are all single-valued. We have seen, however, that the nonhierarchical networks without controls demonstrate a drop in carried load under overload but with different degrees of severity. The most severe drop occurs in the 30-node network, where carried load drops 8.37 percent below the level at 10-percent overload before leveling off (based on the mathematical model). The 25- and 140-node networks demonstrate respective drops of 0.58 and 1.13 percent. Krupp has shown that, for symmetric nonhierarchical networks, all other things being equal, performance under heavy loads worsens as the number of trunks per trunk group or the number of paths per route increases (see also Weber⁶). Thus, we might expect the difference in the severity of the poor performance in our networks to be related to these variables. From Table I, which shows the distribution of trunk group sizes for the three nonhierarchical networks, it is clear that network overload performance does not worsen as trunk group size increases. In fact, the 30-node network, which shows the worst overload performance, has an average of only 40.4 trunks per trunk group, while the 25- and 140-node networks have, respectively, 320.7 and 132.6 trunks per trunk group.

However, this should not be interpreted as a contradiction of Krupp's results, since there are other significant differences in the networks. In particular, the networks vary widely in their route sizes (number of paths per route), as shown in Table II. The 30-node network provides on average 6.91 alternate paths per route, the 25-node network provides 2.12, and the 140-node network provides 4.41. Thus, we see a strong relationship between route size and poor overload performance and, we can conclude that route size is a better predictor

of network overload performance than trunk group size. This further substantiates the conclusion that the reason for the poor overload performance of the nonhierarchical networks is the large amount of alternate routing under overloads made possible by the large number of alternate paths.

V. RESULTS WITH TRUNK RESERVATION FOR FIRST-ROUTED TRAFFIC

As we know from the study of hierarchical networks, appropriate network controls are an effective way to mitigate poor network performance under nonengineered conditions. Since the poor network performance seen above is related to inefficient use of trunks for alternate routing, trunk reservation for first-routed traffic appeared to be an effective way to control our test networks. On each trunk group we reserved 5 percent of the trunks, with a minimum of one trunk, for first-routed traffic. Only the effect on network carried load was considered. Other variants of the trunk-reservation control, as well as the impact on point-to-point blockings, have been investigated but are not discussed here.

The effects of trunk reservation on the 30-node networks are typical of all the networks considered. The results under uniform overloads are shown in Fig. 11. By comparing Figs. 6 and 11, we see that trunk reservation has a very beneficial effect with nonhierarchical routing under large overloads. The drop in carried load with increasing offered load without controls disappears with the use of trunk reservation. Trunk reservation limits the number of multilink calls, allowing more efficient use of the trunks. Similarly, the performance of the hierar-

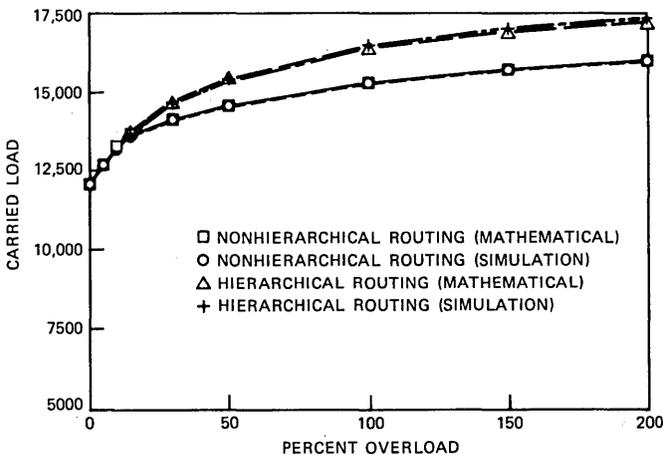


Fig. 11—Performance of the 30-node network under uniform overloads with trunk reservations.

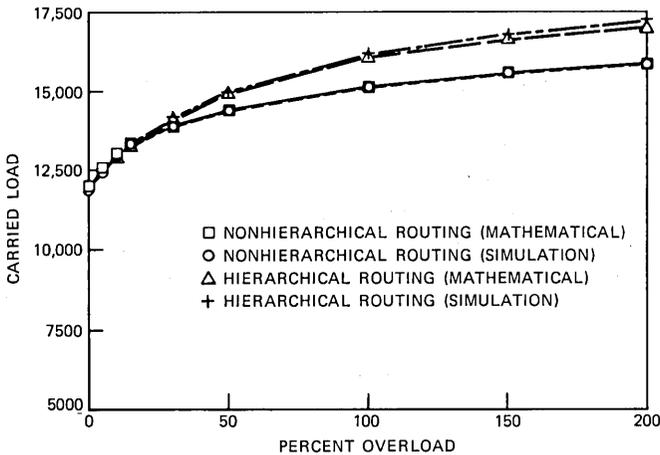


Fig. 12—Performance of the 30-node network under nonuniform overloads with trunk reservation.

chical network under large overloads is improved, although the improvement is not as dramatic as for the nonhierarchical network. At overloads of less than 10 percent, though, two negative effects appear: (1) carried load actually decreases slightly since some alternate-routed calls are prevented from accessing paths that have been engineered to carry them, and (2) some point-to-point blockings are increased, distorting servicing measurements used to schedule trunking augmentations. These effects can be eliminated by using a triggering mechanism to allow the trunk-reservation control only when overloads are large enough to preclude these effects.⁷

Figure 12 shows the trunk reservation results for the 30-node networks under the nonuniform overloads. The effect of the control under large overloads for both the nonhierarchical and hierarchical networks is essentially the same as that seen under uniform overloads. One contrast to the uniform overload case is that now trunk reservation is usually beneficial even at the lighter overloads. Reduction in carried load with trunk reservation is extremely small.

VI. SUMMARY

Analysis of small, symmetric, uniformly loaded nonhierarchical networks has shown the existence of network instabilities for certain networks when only trunking behavior is considered. We have demonstrated by example that such instabilities persist even when the assumptions of symmetry and uniformly distributed loads are removed. Using a mathematical model that we developed, together with a simulation model, we have studied the performance of three realistic

network models under overloads. The 25- and 140-node nonhierarchical networks were designed using the unified algorithm. The 30-node nonhierarchical network was engineered using another design algorithm. In addition, two of the models were studied with hierarchical routing. Results obtained demonstrate that, without controls, the performance of nonhierarchical networks is inferior to that of hierarchical networks under overloads. No instabilities of the type described in Refs. 4 and 5 are seen in any of the engineered nonhierarchical networks, but a drop in carried load consistently occurs at about 10-percent overload because of the tendency of the nonhierarchical networks to alternate-route calls when overloaded. The severity of this drop appears to be correlated with the number of paths per route; nonhierarchical networks with more paths per route exhibit greater throughput degradation under overloads. Such behavior is not seen in the hierarchical networks, which intrinsically can better limit the amount of alternate routing under overloads. An important assumption in our models is the absence of switching system effects. Other studies have shown that, when these effects are included, the performance of networks under large overloads changes significantly from that given by our analysis. However, our results indicate that engineered nonhierarchical networks without controls exhibit unsatisfactory performance even at overloads of 10 to 15 percent, where switching system dynamics are not likely to be an important factor.

We also have applied a control, namely, trunk reservation, for first-routed traffic, to the network models. This control improves the performance of the nonhierarchical and hierarchical networks. By diminishing the amount of alternate routing in the networks under large overloads, trunk reservation permits 1-link calls to use the trunks more efficiently. For the nonhierarchical networks, this results in a continuous increase in carried load with increasing offered load over the entire range of overloads considered. However, at very light overloads, carried load may drop when trunk reservation is activated. This suggests the use of a triggering mechanism to impose trunk reservation only at larger overloads.

REFERENCES

1. C. J. Truitt, "Traffic Engineering Techniques for Determining Trunk Requirements in Alternate Routing Trunk Networks," *B.S.T.J.*, 33, No. 2 (March 1954), p. 277.
2. G. R. Ash et al., "Dynamic, Nonhierarchical Arrangement for Routing Traffic," U.S. Patent 4,345,116, August 17, 1982.
3. G. R. Ash, R. H. Cardwell, and R. P. Murray, "Design and Optimization of Networks With Dynamic Routing," *B.S.T.J.*, 60, No. 8 (October 1981), pp. 1787-1820.
4. R. S. Krupp, "Stabilization of Alternate Routing Networks," *IEEE Int. Commun. Conf.*, Paper No. 3I.2, Philadelphia, 1982.
5. Y. Nakagome and H. Mori, "Flexible Routing in the Global Communication Network," *Seventh Int. Teletraffic Cong.*, Paper No. 426, Stockholm, 1973.
6. J. H. Weber, "A Simulation Study of Routing and Control in Communications Networks," *B.S.T.J.*, 43, No. 6 (November 1964), pp. 2639-76.

7. D. G. Haenschke, D. A. Kettler, and E. Oberer, "DNHR: A New SPC/CCIS Network Challenge," Tenth Int. Teletraffic Cong., Session No. 3.2, Paper No. 5, Montreal, 1983.
8. R. L. Franks and R. W. Rishel, "Overload Model of Telephone Operation," B.S.T.J., 52, No. 9 (November 1973), pp. 1589-1615.

APPENDIX

Mathematical Models

A.1 Models without controls

The model for the nonhierarchical networks without controls requires that a network be specified, together with trunk-group sizes, point-to-point offered loads, and a fixed route for each point-to-point pair. It is assumed that (1) the mixture of traffic offered to a trunk group is Poisson, (2) trunk-group blocking probabilities are independent, and (3) the time needed to make connections (setup time) is small enough, relative to the average holding time of calls, to be ignored. The effect of the first two assumptions can be gauged by comparison with the simulation results, since these assumptions do not occur in the simulator.

We permit the trunk-group sizes and the offered load and the number of paths for each point-to-point pair, as well as the number of trunk groups in a path, to be arbitrary. We also assume that a call blocked on a trunk group of a path can always be cranked back to the originating office so that the call can access the next path in its route.

In the discussion that follows, the term *path* means a set of distinct trunk groups that form a connection between two nodes. A *route* is an ordered collection of paths connecting the same point-to-point pair, specifying the paths used for routing calls between the pair in the order that seizure of the paths is attempted.

Before giving the details of the model, we introduce some notation. Let L^j be the offered load for point-to-point pair j and let $L = \sum_j L^j$ be the total offered load. Let p_i , n_i , and a_i denote, respectively, the blocking probability, trunk-group size, and offered load for trunk group i (in erlangs), and let $q_i = 1 - p_i$. We denote a path by r , a route by R , the route for point-to-point pair j by R^j , and the route formed by the first k paths of R^j by

$$R_k^j = (r_{i_1}^j, \dots, r_{i_k}^j).$$

Finally, we define $D(R)$ to be the probability that route R is blocked, i.e., each path in R has at least one blocked trunk group.

The basic idea of our analysis is to determine the offered load a_i for trunk group i as a function of the trunk-group blockings. For each route containing trunk group i , we determine the contribution that the route makes to the total trunk-group offered load. Suppose trunk group i is in path r^j , the k th path in the route for point-to-point pair

j . The load carried by r_k^j, c_k^j , is given by

$$c_k^j = L^j [D(R_{k-1}^j) - D(R_k^j)], \quad (1)$$

which is the point-to-point load for pair j that overflows the first $k - 1$ paths but not the k th path. The load c_k^j contributes to the carried load for each trunk group $i \in r_k^j$. The total carried load for trunk group i, K_i is obtained by considering all paths containing trunk group i and is given by

$$K_i = \sum_{\substack{j,k \\ i \in r_k^j}} c_k^j.$$

From the relation

$$K_i = a_i q_i$$

for Poisson traffic, we immediately obtain

$$a_i = \sum_{\substack{j,k \\ i \in r_k^j}} c_k^j / q_i. \quad (2)$$

In addition to the relations given by (2), based on our assumption of Poisson trunk-group offered loads, we relate the trunk-groups offered loads to the trunk-group blockings by the Erlang-B formula:

$$p_i = B(n_i, a_i). \quad (3)$$

The equations given by (2) and (3) can be solved iteratively, starting with an initial estimate of the trunk-group blockings, to determine the trunk-group offered loads and blocking probabilities in equilibrium. [The calculation of $D(R)$ is discussed below.] Once a solution has been obtained, several quantities of interest can be calculated. In particular, network blocking, z , is given by

$$z = \frac{\sum_j L^j D(R^j)}{L},$$

and network carried load, C , by

$$C = L(1 - z).$$

We now consider the calculation of $D(R)$, $R = (r_1, \dots, r_k)$. If the paths of R are disjoint, then, by our independence assumption, the blocking probabilities for the paths are independent, so that

$$D(R) = \prod_{t=1}^k \left(1 - \prod_{i \in r_t} q_i \right). \quad (4)$$

For the nonhierarchical networks studied here, the paths have all been restricted to contain either one or two trunk groups. This implies that all paths in a route are disjoint, and $D(R)$ can be calculated using the formula above.

The model used for the hierarchical networks without controls was taken from Ref. 8. We used only the trunk-group portion of the model, omitting the parts dealing with switching system dynamics, retrials, and DABY (don't answer, busy). In fact, this abbreviated model differs from the nonhierarchical model only in the way in which trunk-group carried load is calculated. In the nonhierarchical-network model, this calculation incorporates crankback, while in the hierarchical model, final trunk groups are taken into account.

A.2 Models with trunk reservation for first-routed traffic

Network performance was also modeled with trunk reservation for first-routed traffic. Under this control, a threshold is specified for each trunk group, and alternate-routed calls attempting to seize a trunk on the trunk group are refused if the number of busy trunks on the trunk group has reached the threshold. For the nonhierarchical networks, this control was implemented by subjecting a call to trunk reservation on all legs of an alternate path. In the hierarchical networks, the classification of a call as first-routed or alternate-routed was made at each switching system that the call traversed. Each call was classified as follows: On the first-choice trunk group out of a switching system the call is considered first-routed, whereas on any other trunk group it was considered alternate-routed. All alternate-routed calls offered to a trunk group were subjected to this control. A call overflowing a trunk group because of trunk reservation was offered to the next path in its route. When the first path of a route uses fewer trunk groups than the alternate paths (e.g., the first path is a direct path), then under large loads this control has the effect of decreasing the average number of trunks per call and thus increasing network carried load.

For the trunk reservation model, let a be the total trunk-group offered load, p the trunk group blocking probability, and $q = 1 - p$. Also, let m be the trunk-reservation threshold on the trunk group, \hat{q} the probability that no more than $m - 1$ trunks in the trunk group are busy, \hat{a} the trunk-group offered load that is subject to the trunk-reservation control (i.e., the alternate-routed traffic), \hat{K} the trunk-group carried load subjected to trunk reservation, and $r = \hat{a}/a$. Using a birth-death model for the behavior of n servers with offered load a when less than m servers are busy and offered load $a(1 - r)$ when at least m servers are busy, we obtain the probabilities P_j that exactly j trunks on the trunk group are busy:

$$\begin{aligned}
 P_j &= \frac{a^j}{j!} P_0, \quad j = 0, \dots, m-1, \\
 &= \frac{a^j}{j!} (1-r)^{j-m} P_0, \quad j = m, \dots, n.
 \end{aligned}$$

Here

$$P_0 = \left[\sum_{k=0}^m \frac{a^k}{k!} + \sum_{k=m+1}^n \frac{a^k}{k!} (1-r)^{k-m} \right]^{-1}.$$

It follows that

$$p = \frac{a^n}{n!} (1-r)^{n-m} P_0 \tag{5}$$

$$\hat{q} = \sum_{j=0}^{m-1} \frac{a^j}{j!} P_0. \tag{6}$$

The quantities p and \hat{q} are easily calculated using recursive formulas.

We also need formulas for the calculation of a and \hat{a} , which we obtain by relating these quantities to the corresponding carried loads. The total carried load K is given by

$$\begin{aligned}
 K &= \sum_{j=0}^n j P_j \\
 &= ar\hat{q} + a(1-r)q,
 \end{aligned}$$

so that the offered load that is not subject to trunk reservation is given by

$$a - \hat{a} = \frac{K(1-r)}{r\hat{q} + (1-r)q}. \tag{7}$$

We relate the offered load \hat{a} to the carried load \hat{K} by

$$\hat{a} = \frac{\hat{K}}{\hat{q}}. \tag{8}$$

Carried load K is still calculated as in the previous section with \hat{K} being the portion of K that was subjected to trunk reservation.

The nonhierarchical model with trunk reservation for first-routed traffic is obtained by replacing eq. (2) with eqs. (7) and (8) and replacing eq. (3) with eqs. (5) and (6). The calculation of $D(R)$ in (4) is now given by

$$D(R) = \left(1 - \prod_{i \in r_1} q_i \right) \prod_{t=2}^k \left(1 - \prod_{i \in r_t} \hat{q}_i \right).$$

The hierarchical model with trunk reservation for first-routed traffic is obtained by combining eqs. (5) through (8) with the Franks and Rishel formulas for trunk-group carried load. Separate calculations of total carried load and alternate-routed carried load are made for each trunk group.

AUTHOR

Jacqueline M. Akinpelu, B.A. (Mathematics), 1975, Duke University; Ph.D (Operations Research), 1980, The Johns Hopkins University; AT&T Bell Laboratories, 1980—. Ms. Akinpelu initially worked on modeling overloads in nonhierarchical networks. More recently she has been involved in the development of congestion control procedures and the modeling of overload performance for the common channel signaling network. Member, IEEE.

Methods in Traffic Calculations

By D. L. JAGERMAN*

(Manuscript received February 24, 1984)

This paper presents practical formulae and computer programs for the following traffic-related problems: determination of offered load; determination of number of trunks; equivalent random method and Hayward's method; the three-moment match (construction of interrupted Poisson stream); day-to-day load variation, including both time and call congestions; and the analysis of a multiserver queue in a traffic environment.

I. INTRODUCTION

This paper presents methods of traffic calculations that have been worked out and modified by the author for the typical range of problems encountered in traffic applications. Their efficacy depends, in large part, on improved computations of the Erlang loss function and its derivatives. Practical computer programs are given in the Appendix for immediate application. These programs were originally written by the author in TI EXTENDED BASIC but were transcribed to Fortran by Brian Farrell. All of the programs except for the GT/M/S queue were also written in POCKET BASIC for execution on the TRS-80[†] PC1. The programs are simple and are executed rapidly.

This paper does not fully develop the theories of all of the formulae used. Instead, it gives enough description and explanation to make the development of the computational formulae understandable. Numerical examples show the operation of the methods and programs. These results are intended to serve only as checks on the computer programming; for theoretical accuracy, one should consult the references.

* AT&T Bell Laboratories.

[†] Trademark of Tandy Corporation.

Copyright © 1984 AT&T. Photo reproduction for noncommercial use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

Traffic theory traditionally was concerned only with blocking phenomena. However, modern systems are often analyzed with respect to delays, which arise because of the complex implementation of the communication system and the increasing range of services being provided. For example, delays occurring in systems that yield a response to a query (reservation systems, credit card systems, etc.) are important enough to require careful analysis. In this paper a multi-server delay queue GT/M/S, whose arrival stream is typical of those occurring in traffic theory, is analyzed using the tools of that theory, which are developed here. The waiting time distribution and mean waiting time may be conveniently calculated using the program for the GT/M/S queue given in the Appendix.

II. INTERPOLATION

The Erlang loss function arises in the study of the M/M/n/n queueing problem, that is, a Poisson stream of calls offering a erlangs to a fully available trunk group of n trunks; the Erlang loss function expresses the blocking probability, that is, the probability that an arriving call is rejected because no trunk is available. It usually is stated in the form¹

$$B(n, a) = \frac{a^n}{n!} \bigg/ \sum_{j=0}^n \frac{a^j}{j!} \quad (1)$$

or, equivalently,

$$B(n, a)^{-1} = \sum_{j=0}^n \frac{n!}{j!} a^{-(n-j)}. \quad (2)$$

In terms of the descending factorial, $n^{(j)}$, defined by

$$n^{(0)} = 1, \quad n^{(j)} = n(n-1) \cdots (n-j+1) (j \geq 1), \quad (3)$$

one may also write

$$B(n, a)^{-1} = \sum_{j=0}^n n^{(j)} a^{-j}. \quad (4)$$

The integral representation of Fortet² may be obtained from (4) as follows. From the Eulerian integral

$$a^{-j} = a \int_0^\infty e^{-ay} \frac{y^j}{j!} dy, \quad (5)$$

one has

$$B(n, a)^{-1} = a \int_0^\infty e^{-ay} \sum_{j=0}^n \frac{n^{(j)}}{j!} y^j dy \quad (6)$$

and, hence

$$B(n, a)^{-1} = a \int_0^{\infty} e^{-ay}(1 + y)^n dy. \quad (7)$$

Since the above integral has meaning for nonintegral n , the general definition of $B(x, a)$, in which $x > 0$ is unrestricted and $a > 0$ is (see Ref. 3)

$$B(x, a)^{-1} = a \int_0^{\infty} e^{-ay}(1 + y)^x dy. \quad (8)$$

This relates $B(x, a)$ to the incomplete gamma function,⁴ namely,

$$B(x, a)^{-1} = a^{-x} e^a \int_a^{\infty} e^{-y} y^x dy, \quad (9)$$

$$= a^{-x} e^a \Gamma(x + 1, a). \quad (10)$$

The function $B(x, a)^{-1}$ satisfies an important linear difference equation that will now be obtained. Integration by parts applied to (8) yields

$$B(x, a)^{-1} = 1 + x \int_0^{\infty} e^{-ay}(1 + y)^{x-1} dy. \quad (11)$$

Hence,

$$B(x, a)^{-1} = \frac{x}{a} B(x - 1, a)^{-1} + 1. \quad (12)$$

This is an excellent recursion for the successive computation of $B(x, a)^{-1}$. For integral values of x , the initial value $B(0, a) = 1$ is convenient.

For use in the equivalent random method⁵ and Hayward's approximation,⁶ to be discussed later, it is important to have an easily calculable approximation for $B(x, a)$ when x is nonintegral. For this purpose Newton's interpolation formula will be used.⁷ Define the forward difference operator, Δ , by

$$\Delta f(x) = f(x + 1) - f(x). \quad (13)$$

Then, the powers $\Delta^2, \Delta^3, \dots$, are defined by successive application of Δ , and thus

$$\Delta^2 f(x) = f(x + 2) - 2f(x + 1) + f(x). \quad (14)$$

Newton's interpolation formula is

$$f(x + h) = \sum_{j=0}^{\infty} \binom{h}{j} \Delta^j f(x). \quad (15)$$

Let

$$f(x) = \ln B(x, a), \quad (16)$$

$$n = [x], \quad h = x - n, \quad (17)$$

in which the brackets designate integral part. Newton's formula is now applied up to the second difference to obtain

$$\begin{aligned} \ln B(x, a) \simeq \ln B(n, a) + h\Delta \ln B(n, a) \\ + \frac{1}{2} h(h-1)\Delta^2 \ln B(n, a). \end{aligned} \quad (18)$$

For convenience in writing, the following abbreviations are used

$$B = B(n, a), \quad B_1 = B(n+1, a), \quad B_2 = B(n+2, a). \quad (19)$$

One now obtains, from (18),

$$B(x, a) \simeq B^{1-h} B_1^h \left[\frac{B_1^2}{BB_2} \right]^{\frac{1}{2}h(1-h)}. \quad (20)$$

The accuracy of (20) improves with increasing x . The worst error occurs at $h = 0.5$. Some comparisons are given in Table I.

III. DERIVATIVES

For economic considerations and for iteration formulae for the solution of equations involving $B(x, a)$, as exemplified later in this paper, the derivatives $\partial B(x, a)/\partial a = B_a$ and $\partial B(x, a)/\partial x = B_x$ are needed.⁸ The symbol $B = B(x, a)$ will be used.

From (8) differentiation with respect to a yields

$$-B^{-2}B_a = \int_0^\infty e^{-ay}(1+y)^x dy - a \int_0^\infty e^{-ay}y(1+y)^x dy, \quad (21)$$

$$\begin{aligned} &= \int_0^\infty e^{-ay}(1+y)^x dy - a \int_0^\infty e^{-ay}(1+y)^{x+1} dy \\ &+ a \int_0^\infty e^{-ay}(1+y)^x dy, \end{aligned} \quad (22)$$

Table I—Comparison of interpolation with exact values

x	a	$B=$	$B\simeq$
1.5	0.1	0.02155	0.02132
5.5	2	0.02146	0.02145
10.5	8	0.10011	0.10013
100.5	90	0.02517	0.02517

$$= \left(1 + \frac{1}{a}\right) B^{-1} - B(x+1, a)^{-1}, \quad (23)$$

$$B_a = \left(\frac{x+1}{a} B^{-1} + 1\right) B^2 - \left(1 + \frac{1}{a}\right) B, \quad (24)$$

$$B_a = \left(\frac{x}{a} - 1 + B\right) B. \quad (25)$$

For B_x one has, from (8),

$$B_x = -B^2 a \int_0^\infty e^{-ay} (1+y)^x \ln(1+y) dy. \quad (26)$$

Unfortunately, there is no exact evaluation of (26) in convenient form. Useful, easily calculable approximations may, however, be obtained. First, a crude but useful approximation will be obtained. Let

$$f(y) = Bae^{-ay}(1+y)^x. \quad (27)$$

Then, from (8),

$$f(y) \geq 0, \quad \int_0^\infty f(y) dy = 1. \quad (28)$$

Also, one has

$$\int_0^\infty yf(y) dy = B(x+1)^{-1}B - 1, \quad (29)$$

$$= \frac{x+1}{a} + B - 1. \quad (30)$$

Jensen's inequality⁹ for a random variable (ξ) and a function $g(x)$ convex on the range of (ξ) is

$$Eg(\xi) \geq g(E\xi). \quad (31)$$

Accordingly, let ξ have the density function $f(y)$. Then, (26) may be expressed as

$$B_x/B = -E \ln(1 + \xi). \quad (32)$$

Since $-\ln(1+y)$ is convex on $y \geq 0$, using Jensen's inequality gives

$$B_x/B \geq -\ln\left(\frac{x+1}{a} + B\right), \quad (33)$$

in which $E\xi$ was obtained from (30). It will be convenient to set

$$\alpha = \frac{x+1}{a} + B \quad (34)$$

so that one has

$$B_x/B \geq -\ln \alpha. \quad (35)$$

To obtain a better approximation to B_x than the lower bound of (33), the difference equation (12) will be used. Again, let

$$f(x) = \ln B(x, a).$$

Then, the difference equation becomes

$$f(x+1) - f(x) = -\ln \alpha. \quad (36)$$

The Taylor expansion for $f(x+1) - f(x)$ gives

$$f'(x) \cong -\ln \alpha - \frac{1}{2} f''(x). \quad (37)$$

From $f'(x) \cong -\ln \alpha$, one has $f'' \cong -\frac{\alpha'}{\alpha}$; hence,

$$f'(x) \cong -\ln \alpha + \frac{\alpha'}{2\alpha}. \quad (38)$$

One has

$$f'(x) = \frac{B_x}{B}, \quad (39)$$

$$\alpha' = \frac{1}{a} + B_x. \quad (40)$$

Substituting these values of f' , α' into (38) yields the following approximation for B_x :

$$B_x/B \cong -\frac{\ln \alpha - 1/(2a\alpha)}{1 - B/(2\alpha)}. \quad (41)$$

With (35) designated as bound and (41) designated as approximation, Table II presents comparisons with exact values taken from the table of Akimaru and Nishimura. Throughout the table $B = 0.01$.

IV. DETERMINATION OF OFFERED LOAD

In the equation

$$B(x, a) = P, \quad (42)$$

Table II—Comparisons of derivative values

x	a	$-B_x/B$	Bound	Approximation
5	1.3608	1.4025	1.4860	1.4044
10	4.4612	0.8626	0.9065	0.8630
20	12.0306	0.5406	0.5628	0.5406
50	37.9014	0.2956	0.3042	0.2956

in which x and P are given, a is to be determined. Newton's method of iteration is well suited to the problem. Let a_0 be an initial approximation, and let a_1 be the refined result. Then,

$$a_1 = a_0 - \frac{B_0 - P}{(B_a)_0}, \quad (43)$$

in which the subscript 0 indicates evaluation at $a = a_0$. Using (25) one has

$$a_1 = a_0 - \frac{B_0 - P}{\left(\frac{x}{a_0} - 1 + B_0\right) B_0}. \quad (44)$$

The problem of obtaining a good starting point remains. For this purpose an inequality for B is obtained.

From

$$(1 + y)^x \leq e^{xy} \quad (45)$$

and (8) one gets

$$B(x, a)^{-1} \leq \frac{a}{a - x}, \quad x < a. \quad (46)$$

Using (46) in (12) now yields

$$B(x, a) \geq 1 - \frac{x}{a + 1}, \quad (47)$$

with no restriction on x . Thus, setting $B = P$ in (47), the initial value can be

$$a_0 = \frac{x}{1 - P} - 1. \quad (48)$$

As an example of the convergence rate of (44) starting with (48), consider $x = 20$, $P = 0.01$. The following values were obtained using the program given in the Appendix.

$$\begin{aligned} a_0 &= 19.202, & a_1 &= 14.057, & a_2 &= 12.568, \\ a_3 &= 12.088, & a_4 &= 12.031, & a_5 &= 12.031. \end{aligned}$$

Another important case occurs when the carried load, L , is specified and the offered load is required; the relevant equation is

$$L = a(1 - B(x, a)). \quad (49)$$

Newton's formula in the form

$$a_1 = a_0 - \frac{a_0(1 - B_0) - L}{(L_a)_0} \quad (50)$$

is used. Since

$$L_a = 1 - B - aB_a, \quad (51)$$

one has, using (25) and (5),

$$a_1 = a_0 - \frac{a_0(1 - B_0) - L}{1 - B_0 - (x - a_0 + a_0B_0)B_0}. \quad (52)$$

The inequality

$$a < L \left(1 + \frac{L}{x(x - L)} \right) \quad (53)$$

gives a convenient starting value for (52); thus,

$$a_0 = L \left(1 + \frac{L}{x(x - L)} \right). \quad (54)$$

The convergence rate is the same as in (44).

V. DETERMINATION OF NUMBER OF TRUNKS

The equation

$$B(x, a) = P \quad (55)$$

will now be solved for x given a and P . The Newton iteration formula now reads

$$x_1 = x_0 - \frac{B_0 - P}{(B_x)_0}. \quad (56)$$

The bound for B_x , namely (33), will be used in (56); thus

$$x_1 = x_0 + \frac{B_0 - P}{B_0 \ln \alpha_0}. \quad (57)$$

The starting value for x is again obtained from (47); it is

$$x_0 = (1 - P)(1 + a). \quad (58)$$

One must use (20) to evaluate B , since x need not be an integer. Using the approximate value (33) for B_x does not impair the accuracy of the result. It merely slows down the convergence rate over what Newton's method would provide with the exact derivative.

Using the program of the appendix and $a = 12.031$, $P = 0.01$, the following values are obtained:

$$\begin{array}{llll} x_0 = 12.901, & x_1 = 16.319, & x_2 = 18.352, & x_3 = 19.505, \\ x_4 = 19.932, & x_5 = 19.997, & x_6 = 20.000. & \end{array}$$

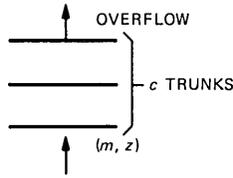


Fig. 1—Common trunk group.

VI. EQUIVALENT RANDOM METHOD AND HAYWARD'S METHOD

Figure 1 schematizes the problem of ascertaining blocking on a common trunk group. The common trunk group is offered a composite stream, which is the superposition of overflow streams.

The input stream is characterized by the offered load m and peakedness z . The equivalent random method considers the stream (m, z) to be the overflow stream of a fictitious trunk group of x trunks and Poisson-offered load a . This is shown in Fig. 2.

If the stream (m, z) is not in fact the overflow of a single trunk group, then x need not be an integer. The Kosten formulae for overflow and peakedness are

$$m = aB(x, a), \quad (59)$$

$$z = 1 - m + \frac{a}{x + m + 1 - a}. \quad (60)$$

These can be arranged in the following form:

$$m = aB \left[a \frac{m + z}{m + z - 1} - m - 1, a \right], \quad (61)$$

$$x = a \frac{m + z}{m + z - 1} - m - 1. \quad (62)$$

The problem is to determine the equivalent random parameters x and

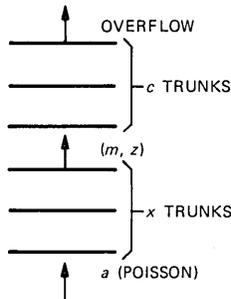


Fig. 2—Common and fictitious trunk groups.

a ; that is, to solve (61) for a and then obtain x from (62). When x and a are known then, of course, the required blocking probability, P , is given by

$$P = \frac{B(x+c, a)}{B(x, a)}. \quad (63)$$

Newton's method will be used to solve for a . Thus,

$$a_1 = a_0 - \frac{aB_0 - m}{\frac{d}{da}(aB)_0}, \quad (64)$$

in which $B = B\left(a \frac{m+z}{m+z-1} - m - 1, a\right)$.

One has

$$\frac{d}{da}(aB) = B + (x+m+1)B_x + aB_a, \quad (65)$$

and using (25) and (33),

$$\frac{d}{da}(aB) \cong [x+m+1-a-(x+m+1)\ln \alpha]B. \quad (66)$$

Newton's eq. (64) now becomes

$$a_1 = a_0 - \frac{a_0B_0 - m}{[x_0+m+a-a_0-(x_0+m+1)\ln \alpha_0]B_0}. \quad (67)$$

The formula of Rapp,¹⁰ namely,

$$a_0 = mz + 3z(z-1), \quad (68)$$

will start the iteration.

The following test case was used on the program in the Appendix. For $x = 10$, $a = 8$ one finds, from (59) and (60), that $m = 0.97329$, $z = 2.04016$. The results of the run are

$$\begin{array}{lll} x_0 = 10.527, & x_1 = 10.181, & x_2 = 10.067, \\ a_0 = 8.352, & a_1 = 8.121, & a_2 = 8.045, \\ x_3 = 10.025, & x_4 = 10.010, & x_5 = 10.004, \\ a_3 = 8.017, & a_4 = 8.007, & a_5 = 8.003. \end{array}$$

The equivalent random method is usually used only when the service distribution on the common group is exponential, since the Kosten formula was derived for that distribution. However, in this regard, see Ref. 11 for a discussion of the constant service time case. To consider

blocking on a common group with other service distributions, the approximation of W. S. Hayward is used. (See Ref. 6 for example.) In Fig. 1 the approximation is

$$P_h = B \left(\frac{c}{z}, \frac{m}{z} \right). \quad (69)$$

To use (69), the peakedness, z , must be referred to the service distribution that is considered. This will now be discussed.¹² For a service distribution, $F(x)$, with service rate μ , let

$$F_0(x) = F(x/\mu), \quad (70)$$

that is, the distribution scaled to unit rate, and further

$$F_0^c(x) = 1 - F(x),$$

$$F_0^{c(2)}(y) = \int_0^\infty F_0^c(x)F_0^c(x+y)dx. \quad (71)$$

The notation $z(F_0^c; \mu)$ will be used to show the dependence of z on μ as a function and on F as a functional. Since z is usually known relative to some distribution, it would be useful to be able to transform z to other distributions. The Mellin transform will accomplish this.

For a function, $f(x)$, defined on $(0, \infty)$, the function $\bar{f}(s)$ defined by

$$\bar{f}(s) = \int_0^\infty x^{s-1}f(x)dx \quad (72)$$

is called the Mellin transform of $f(x)$. Let

$$f(\mu) = z(F_0^c; \mu) - 1, \quad (73)$$

$$g(\mu) = z(G_0^c; \mu) - 1. \quad (74)$$

Then, the required transformation formula for z is

$$\bar{g}(s) = \frac{\bar{G}_0^{c(2)}(s)}{\bar{F}_0^{c(2)}(s)} \bar{f}(s). \quad (75)$$

For a renewal stream with renewal density $m(\tau)$, the peakedness may be calculated directly by

$$z(F_0^c; \mu) = 1 - \frac{\lambda}{\mu} + 2\mu \int_0^\infty F_0^{c(2)}(u)m(u)du. \quad (76)$$

As an example, consider the stream given by

$$m(\tau) = \lambda + Ae^{-a\tau}. \quad (77)$$

By (76), z relative to exponential service is

$$z(e^{-x}; \mu) = 1 + \frac{A}{\alpha + \mu}. \quad (78)$$

It is desired to transform this to the service distribution

$$G(x) = 1 - (x + 1)e^{-x}, \quad (79)$$

for which $\mu = 1/2$. One has

$$\bar{F}_0^{c,(2)}(s) = \frac{1}{2} \Gamma(s), \quad (80)$$

$$\bar{G}_0^{c,(2)}(s) = \frac{5}{8} 2^{-s} \Gamma(s) + \frac{3}{4} 2^{-s-1} \Gamma(s + 1). \quad (81)$$

Hence,

$$\frac{\bar{G}_0^{c,(2)}(s)}{\bar{F}_0^{c,(2)}(s)} = \frac{5}{4} 2^{-s} + \frac{3}{4} 2^{-s}s. \quad (82)$$

Since

$$\bar{f}(s) = \frac{\pi}{\sin \pi s} A \alpha^{s-1}, \quad (83)$$

one now obtains from (75) and (82),

$$\bar{g}(s) = \frac{\pi}{\sin \pi s} A \left[\frac{5}{8} \left(\frac{\alpha}{2} \right)^{s-1} + \frac{3}{8} \left(\frac{\alpha}{2} \right)^{s-1} s \right]. \quad (84)$$

Hence,

$$z(G_0^c; \mu) = 1 + \frac{A}{8} \left[\frac{10}{\alpha + 2\mu} + \frac{12\mu}{(\alpha + 2\mu)^2} \right]. \quad (85)$$

Since $\mu = 1/2$, one finally has

$$z = 1 + \frac{A}{8} \left[\frac{10}{\alpha + 1} + \frac{6}{(\alpha + 1)^2} \right]. \quad (86)$$

Consider the following numerical examples. Let the common group have exponential service distribution, and let $c = 15$, $m = 10$, $z = 3$. Then, the equivalent random parameters are $x = 39.615$, $a = 46.721$. Thus, one has

$$P = \frac{B(54.615, 46.721)}{B(39.615, 46.721)} = 0.151, \quad (87)$$

$$P_h = B(5, 3.333) = 0.139. \quad (88)$$

Let the service rate for this example be $\mu = 1$, and let the arrival stream be that defined by (77) with $\lambda = 10$, $A = 4$, $\alpha = 1$. This is

consistent with $m = 10$, $z = 3$. Now let the service distribution be that of (79). Then, by (86)

$$z = 4.25 \tag{89}$$

and

$$P_h = B \left(\frac{15}{4.25}, \frac{10}{4.25} \right) = 0.187. \tag{90}$$

VII. THE THREE-MOMENT MATCH

The interrupted Poisson process is on (flows) for an exponential period of time whose mean duration is γ^{-1} , and is off (stopped) for an exponential period whose mean duration is ω^{-1} . This may be thought of as a Poisson process of rate λ entering a switch that is alternately closed and opened. The output of the switch is the interrupted Poisson process with a rate λ' . This is shown in Fig. 3.

A stream that is the overflow of a single trunk group with Poisson offered load will be called on 0-stream. The interrupted Poisson process provides a useful approximation to an 0-stream.¹³ The technique of approximation offers the 0-stream to an infinite server group and offers the interrupted Poisson stream to another infinite server group. In each group the distribution of the number of busy servers at any instant of time is obtained. The first three moments of these distributions are then equated. They give equations that define the interarrival time distribution of the interrupted Poisson stream. Since this stream is renewal, it is now completely defined.

Let the 0-stream be the overflow from a trunk group with x trunks and offered load a , and let

$$B = B(x, a), \quad B_1 = B(x + 1, a), \quad B_2 = (x + 2, a). \tag{91}$$

Then, the required equations for λ , λ' , and the switch parameters, γ , ω , are

$$\delta_0 = B, \quad \delta_1 = \frac{1}{a} \frac{\delta_0^{-1}}{B_1^{-1} - B^{-1}} \tag{92}$$

$$\delta_2 = \frac{2}{a^2} \frac{\delta_0^{-1} \delta_1^{-1}}{B_2^{-1} - 2B_1^{-1} + B^{-1}} \tag{93}$$

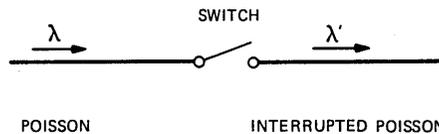


Fig. 3—Generation of interrupted Poisson process.

$$\lambda = a \frac{\delta_2(\delta_1 - \delta_0) - \delta_0(\delta_2 - \delta_1)}{2\delta_1 - \delta_0 - \delta_2}, \quad (94)$$

$$\omega = \frac{\delta_0 \lambda - a\delta_1}{\lambda \delta_1 - \delta_0} \quad (95)$$

$$\gamma = \frac{\omega \lambda - a\delta_0}{a \delta_0}, \quad (96)$$

$$\lambda' = \lambda \frac{\omega}{\gamma + \omega}. \quad (97)$$

The Laplace transform, $\tilde{f}(s)$, of the interarrival time density function, $f(x)$, is

$$\tilde{f}(s) = \frac{\lambda s + \lambda \omega}{s^2 + (\lambda + \gamma + \omega)s + \lambda \omega}, \quad (98)$$

and the peakedness relative to exponential service is

$$z(\mu) = 1 + \frac{\lambda \gamma}{\gamma + \omega} \frac{1}{\gamma + \omega + \mu}. \quad (99)$$

The equations for the three-moment match were set up in Ref. 13, taking $\mu = 1$. Thus, (99) yields the exact peakedness of the 0-stream for $\mu = 1$. A program for the evaluation of λ , λ' , γ , ω , is given in the Appendix.

The following numerical example was evaluated. For $x = 10$, $a = 8$, one finds $\lambda = 5.4405$, $\gamma = 2.7054$, $\omega = 0.5894$, $\lambda' = 0.9733$, and $z = 2.0402$.

The interrupted Poisson process is also used to construct a stream with given parameters (m, z) ($z > 1$). The equivalent random parameters x, a are first found, then the above equations are used, even when x is not an integer, to obtain the parameters for the interrupted Poisson stream. As an example, consider the case used earlier for which $m = 10$, $z = 3$, with unit service rate. Using $x = 39.6148$, $a = 46.7214$, one finds $\lambda = 25.9889$, $\gamma = 4.3033$, $\omega = 2.6912$.

An interrupted Poisson stream may be used in simulation studies. It may also be used to obtain the blocking probability of Fig. 1, instead of the equivalent random method or Hayward's method. The blocking probability, P_{ip} , is given by¹⁴

$$P_{ip} = 1 + \sum_{j=1}^c C^{(j)} \lambda^{-j} \prod_{i=1}^j \frac{\gamma + \omega + i}{\omega + i}. \quad (100)$$

Since this process is constructed to be a good approximation to an 0-stream, it is to be expected that P_{ip} in (100) should be in close

agreement with (63), as given by the equivalent random method. The following numerical examples of Table III show this.

The interrupted Poisson process is used when analyzing the delay queue considered in Section IX.

VIII. DAY-TO-DAY LOAD VARIATION

Observation of traffic on a consistent hour basis from day to day shows variation. To partially account for this variation,⁵ one often assumes that the offered load, a , is random with a gamma density $f(x)$. Thus,

$$f(x) = e^{-x/\gamma} \frac{x^{\alpha-1}}{\Gamma(\gamma)\gamma^\alpha}, \quad (101)$$

$$\alpha\gamma = m, \quad \alpha\gamma^2 = \sigma^2, \quad (102)$$

in which m, σ^2 are the mean and variance, respectively, of a . Any load-dependent statistic such as $B(x, a), aB(x, a), z$, etc. is replaced by its respective means. Let $g(a)$ be such a statistic; then it is necessary to evaluate $Eg(a)$. One can construct an approximation by using the Gauss-Laguerre quadrature theory.¹⁵ Then the offered load can be considered to consist of two Poisson streams whose offered loads are a_1, a_2 and to occur with probabilities P_1, P_2 , respectively. The quantities P_1, a_1, P_2, a_2 in terms of m and σ are

$$P_1 = \frac{1}{2} - \frac{1}{2} \frac{\sigma}{\sqrt{m^2 + \sigma^2}}, \quad (103)$$

$$a_1 = \frac{m^2 + \sigma^2 + \sigma \sqrt{m^2 + \sigma^2}}{m}, \quad (104)$$

$$P_2 = \frac{1}{2} + \frac{1}{2} \frac{\sigma}{\sqrt{m^2 + \sigma^2}}, \quad (105)$$

$$a_2 = \frac{m^2 + \sigma^2 - \sigma \sqrt{m^2 + \sigma^2}}{m}. \quad (106)$$

Thus, one has

$$Eg(a) \cong P_1g(a_1) + P_2g(a_2). \quad (107)$$

Table III—Comparison of interrupted Poisson method with equivalent random method

c	10	10	15
m	8	15	10
z	10	3	3
P_{ip}	0.5273	0.4910	0.1492
P	0.5278	0.4912	0.1507

For example, to compute day-to-day time congestion, one uses $EB(x, a)$.

Using the symbol \bar{B} for this quantity, one has, from (107),

$$\bar{B} \cong P_1 B(x, a_1) + P_2 B(x, a_2). \quad (108)$$

Of possibly greater importance is the probability, P_B , that a call is blocked. To obtain this, let

$$0(x, a) = aB(x, a), \quad (109)$$

that is, $0(x, a)$ is the overflow traffic, and let $E0(x, a) = \bar{0}$, then

$$P_B = \frac{\bar{0}}{m}. \quad (110)$$

Thus, one has approximately

$$P_B \cong \frac{1}{m} [P_1 a_1 B(x, a_1) + P_2 a_2 B(x, a_2)]. \quad (111)$$

Wilkinson found empirically that one may often relate σ^2 to m through

$$\sigma^2 = 0.13m^{1.5}, \quad 0.13m^{1.7}, \quad 0.13m^{1.84} \quad (112)$$

for low, medium, and high variation, respectively.⁵ Programs for evaluating \bar{B} and P_B are given in the Appendix. A numerical example is $x = 20$, $m = 15$, and (112) for medium variation. ($\sigma^2 = 12.9807$.) One finds $P_B \cong 0.0795$ and $\bar{B} \cong 0.0635$.

The behavior of \bar{B} is somewhat counter-intuitive, since, while \bar{B} may at first increase with increasing σ , ultimately it decreases to zero.¹⁶ The behavior of P_B is more satisfactory since it increases monotonically to one for $\sigma \rightarrow \infty$. One also has $P_B \geq \bar{B}$. An example for which $\bar{B} < B(x, m)$ is given by $B(3, 2) = 0.2105$, while $\bar{B} = 0.1988$ when $\sigma = 1$.

IX. GT/M/S QUEUE

Until now no delay queues were considered; however, it is becoming more and more important to estimate delays occurring in traffic systems. The approximations developed in this paper serve this purpose. In particular, the interrupted Poisson stream can be used as input to a GT/M/S queue for which an exact solution can be obtained. The peculiarity of the problem considered here is that the input stream is defined through specification of (m, z) rather than the usual specifications used in queueing theory (distribution of time between arrivals, etc.), hence the designation T for traffic. Considered from this point of view, a queue may be a queue in a traffic environment. Figure 4 shows this. For a discussion of this approach see Heffes.¹⁷

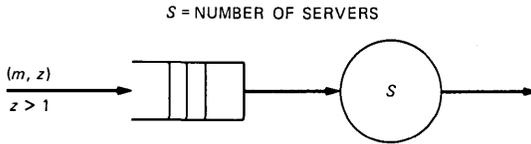


Fig. 4—Traffic-type queueing system.

The solution of the GT/M/S system with interrupted Poisson stream (obtained from three-moment match) as input may be expected to provide a close approximation when the input to the GT/M/S is a superposition of 0-streams. But even in more general situations the approximation is good.

The servers are assumed to have exponential service-time distribution with unit mean service rate, and to be identical and independent. The paradigm for the solution is as follows:

$$(m, z) \rightarrow (x, a) \rightarrow (\lambda, \gamma, \omega) \rightarrow w(t). \quad (113)$$

Thus, from the traffic, m , and peakedness, z , one obtains the equivalent random parameters x (number of trunks) and a (offered load), which give the interrupted Poisson parameters λ , γ , and ω . The interrupted Poisson stream is now used as input to a GI/M/S queue from which the exact waiting time distribution is computed. To accomplish the last step, the following formulae are used.¹⁸ The function $\tilde{f}(s)$ is given in (98). Define r to be the root of

$$r = \tilde{f}[(1 - r)S], \quad (114)$$

where S is the number of servers, satisfying $0 < r < 1$, and $\delta = 1 - r$. Thus,

$$\delta = \frac{1}{2S^2} [\sqrt{\{(\lambda + \gamma + \omega)S - S^2\}^2 + 4S^2\{(\gamma + \omega)S - \lambda\}} - \{(\lambda + \gamma + \omega)S - S^2\}]. \quad (115)$$

Define the renewal density, $\tilde{m}(s)$, by

$$\tilde{m}(s) = \frac{1}{1 - \tilde{f}(s)} - 1. \quad (116)$$

Then, the following quantities are calculated:

$$C_j = \prod_{i=1}^j \tilde{m}(i), \quad (117)$$

$$A^{-1} = \frac{1}{\delta} + \sum_{j=1}^S \frac{\binom{S}{j}}{C_j[1 - \tilde{f}(j)]} \frac{S[1 - \tilde{f}(j)] - j}{S\delta - j}, \quad (118)$$

$$P = \frac{A}{\delta}. \quad (119)$$

One now has

$$P[w > 0] = P, \quad (120)$$

$$P[w > t] = Pe^{-\delta St}, \quad (121)$$

$$EW = \frac{P}{\delta D}. \quad (122)$$

The above computation of delay is very sensitive to errors in the determination of $\hat{f}(s)$, which, in turn, rests on an accurate evaluation of $B(x, a)$. Accordingly, another method of computing $B(x, a)$, which is more accurate than (20), will be used. The formula is³

$$B(x, a)^{-1} \cong \sum_{j=0}^{[x+a]} x^{(j)} a^{-j}. \quad (123)$$

Let

$$U_j = x^{(j)} a^{-j}. \quad (124)$$

Then, the relative error, ϵ , in calculating $B(x, a)$ is

$$\epsilon = -B(x, a)U_{[x+a]}. \quad (125)$$

A computer program employing (123) is given in the Appendix. This program combines all computations needed. It accepts m, z, S and yields r, P, EW . A test based on $U_{[x+a]}$ is made: If $U_{[x+a]} \leq 10^{-7}$, the computations are accepted and the results printed; this also reduces the time of computation by reducing the number of times a particular loop is used. If, however $U_{[x+a]} > 10^{-7}$, then the computation is considered inaccurate and the message **not accurate** is printed. This situation occurs when the offered load, m , is not large and the peakedness is near one. For example, $m = 5, z = 1.1, S = 7$ prints **not accurate**. Formula (123) is not as robust as (20) and therefore was not used in previous computations of this paper. If greater accuracy is needed in those computations, then (123) may be used in place of (20) with the required test for accuracy. Table IV gives some sample calculations.

X. SUMMARY

The mathematical methods and algorithms presented here enabled the construction of convenient computer programs that rapidly evaluate many traffic-related problems. These represent some of the main approaches and approximations used in this area of traffic theory.

Some of the routines used in the individual programs are common.

Table IV—Examples of GT/M/S computations

<i>m</i>	<i>z</i>	<i>s</i>	<i>r</i>	<i>P</i>	<i>EW</i>
12	1.1	15	0.8160	0.3444	0.1248
5	3	8	0.8750	0.5132	0.5132
2	5	5	0.8882	0.6091	1.0894
2	3	4	0.8418	0.5814	0.9186
1	4	3	0.8468	0.6392	1.3911
1	3	2	0.8505	0.7446	2.4896
0.7	4	1	0.9354	0.9354	14.4851

For this reason, a main index-driven program containing all of the programs given would shorten coding and enable the user to select one program after another as the need arises.

REFERENCES

1. B. V. Gnedenko and I. N. Kovalenko, *Introduction to Queueing Theory*, Jerusalem: S. Monson, 1968, Chapter 1.
2. R. Syski, *Introduction to Congestion Theory in Telephone Systems*, Edinburgh: Oliver and Boyd, 1959.
3. D. L. Jagerman, "Some Properties of the Erlang Loss Function," *B.S.T.J.*, 53, No. 3 (March 1974), pp. 525-51.
4. A. Erdelyi, *Bateman Manuscript Project*, Vol. 1, New York: McGraw-Hill, 1953.
5. R. I. Wilkinson, "Theories of Toll Traffic Engineering in the U.S.A.," *B.S.T.J.*, 35, No. 2 (March 1956), pp. 421-514.
6. A. A. Fredericks, "Congestion in Blocking Systems—A Simple Approximation Technique," *B.S.T.J.*, 59, No. 6 (July-August 1980), pp. 805-27.
7. C. Froberg, *Introduction to Numerical Analysis*, Second Edition, Reading, MA: Addison-Wesley, 1965.
8. H. Akimaru and T. Nishimura, "The Derivatives of Erlang's B Formula," *Rev. Elec. Commun. Lab*, 11, No. 9-10, 1963.
9. Hardy, Littlewood, and Polya, *Inequalities*, New York: Cambridge University Press, 1934.
10. L. Y. Rapp, "Planning of Junction Networks in a Multi-Exchange Area-Part 1," *Ericson Technics*, 20, 1964, pp. 77-130.
11. P. J. Burke, "The Overflow Distribution for Constant Holding Time," *B.S.T.J.*, 50, No. 10 (December 1971), pp. 3195-210.
12. A. E. Eckberg, unpublished work.
13. A. Kuczura, "The Interrupted Poisson Process as an Overflow Process," *B.S.T.J.*, 52, No. 3 (March 1973), pp. 437-48.
14. L. Takačs, *Introduction to the Theory of Queues*, New York: Oxford, 1962.
15. V. D. Krylov, *Approximate Calculation of Integrals*, New York: MacMillan Co., 1962.
16. A. Kashper, S. M. Rocklin, and C. R. Szelag, "Effects of Day-to-Day Load Variation on Trunk Group Blocking," *B.S.T.J.*, 61, No. 2 (February 1982), pp. 123-35.
17. H. Heffes, "Analysis of First-Come-First-Served Queueing Systems With Peaked Inputs," *B.S.T.J.*, 52, No. 7 (September 1973), pp. 1215-28.
18. R. B. Cooper, *Introduction To Queueing Theory*, Second Edition, North Holland, 1981.

APPENDIX

Fortran Programs for Main Calculations in Text

Erlang Loss Function

```

IMPLICIT REAL *8 (A-H,O-Z)
REAL*8 MO, M1
WRITE (6, 10)
10      FORMAT (/ , 1X, 4X, 'ERLANG LOSS FUNCTION' ,
2        // , 1X, 'NO. OF TRUNKS = ')
    
```

```

20      READ (5,20) X
        FORMAT (F7.4)
        WRITE (6,30)
30      FORMAT (1X, 'OFFERED LOAD = ')
        READ (5,20) A
        N = IDINT (X)
        H = X - N
        B = 1.
        DO 50 J = 1, N
        B = J/A * B + 1
50      CONTINUE
        B1 = (N+1)/A * B + 1.
        B2 = (N+2)/A * B1 + 1.
        Y = B** (1.-H) * B1**H
        Y = Y * (B1*B1/B/B2)** (H*(1.-H)/2.)
        B = 1./Y
        MO = A*B
        ZO = 1. - MO + A/(X+1.+MO-A).
        M1 = IDINT (100000.*MO+.5)/100000.
        Z1 = IDINT (100000.*ZO+.5)/100000.
        C = IDINT (100000.*B+.5)/100000.
        WRITE (6,70) X,A,B,M1,Z1
70      FORMAT (/ ,1X, 'B(', F8.4, ', ', ' ',
2         F8.4, ') = ', F7.5, / ,1X,
3         ' OVERFLOW = ', F8.4, / ,1X,
4         ' PEAKEDNESS = ', F8.4, / / / /)
        STOP
        END

```

Offered Load From Blocking

```

        IMPLICIT REAL*8 (A-H,O-Z)
        WRITE (6,10)
10      FORMAT (/ ,1X, 2, 'OFFERED LOAD FROM',
2         ' BLOCKING', //1X, 'TRUNKS = ')
        READ (5,20) N
20      FORMAT (I3)
        WRITE (6,30)
30      FORMAT (1X, 'BLOCKING PROB. = ')
        READ (5,40) P
40      FORMAT (F4.4)
        A = N/(1.-P) - 1.
        DO 60 K=1, 5
        B = 1.
        DO 50 J = 1, N
        B = J/A * B + 1.
50      CONTINUE
        F = 1.-P*B
        F1 = N/A - 1. + 1./B
        A = A - F/F1
60      CONTINUE
        C = IDINT (10000.*A+.5)/10000.
        WRITE (6,70) C

```

```

70          FORMAT (/ , 1X, 'OFFERED LOAD = ' , F6.4 )
           STOP
           END

```

Offered Load From Carried Load

```

           IMPLICIT REAL*8 (A-H, O-Z)
           REAL*8 L
           WRITE ( 6 , 10 )
10          FORMAT (/ , 1X, 2X, 'OFF. LOAD FROM' ,
2           'CAR. LOAD' , // , 1X, 'TRUNKS = ' )
           READ ( 5 , 20 ) N
20          FORMAT ( I3 )
           WRITE ( 6 , 30 )
30          FORMAT ( 1X, 'CAR LOAD = ' )
           READ ( 5 , 40 ) L
40          FORMAT ( F7.4 )
           A = L * ( 1. + L / N / ( N - L ) )
           DO 60 K = 1 , 5
           B = 1.
           DO 50 J = 1 , N
           B = J / A * B + 1.
50          CONTINUE
           B = 1. / B
           F = A * ( 1. - B ) - L
           F1 = 1. - B - ( N - A + A * B ) * B
           A = A - F / F1
60          CONTINUE
           C = IDINT ( 10000. * A + .5 ) / 10000.
           WRITE ( 6 , 70 ) C
70          FORMAT (/ , 1X, 'OFFERED LOAD = ' ,
2           F8.4 , /// )
           STOP
           END

```

Trunks From Blocking

```

           IMPLICIT REAL*8 (A-H, O-Z)
           WRITE ( 6 , 10 )
10          FORMAT (/ , 1X, 4X, 'TRUNKS FROM' ,
2           'BLOCKING' , // , 1X, 'OFFERED LOAD = ' )
           READ ( 5 , 20 ) A
20          FORMAT ( F7.4 )
           WRITE ( 6 , 30 )
30          FORMAT ( 1X, 'BLOCKING PROB. = ' )
           READ ( 5 , 40 ) P
40          FORMAT ( F4.4 )
           X = ( 1. - P ) * ( 1. + A )
           DO 60 K = 1 , 10
           N = IDINT ( X )
           H = X - N
           B = 1.

```

```

DO 50 J=1,N
B = J/A*B+1
50 CONTINUE
B1 = (N+1)/A*B+1.
B2 = (N+2)/A*B1+1.
Y = B**(1.-H)*B1**H
Y = Y*(B1*B1/B/B2)**(H*(1.-H)/2)
G = (X+1.)/A+1./Y
X = X + (1.-P*Y)/DLOG(G)
60 CONTINUE
C = IDINT(10000.*X+.5)/10000.
WRITE(6,70) C
70 FORMAT(/,1X,'NO. OF TRUNKS = ',
2 F8.4)
STOP
END

```

Equivalent Random Parameters

```

IMPLICIT REAL*8(A-H,O-Z)
WRITE(6,10)
10 FORMAT(/,1X,5X,'EQUIV. RAND. ',
2 'PARA. ',//,1X,'LOAD = ')
READ(5,20) M
20 FORMAT(F7.4)
WRITE(6,30)
30 FORMAT(1X,'PEAKEDNESS = ')
READ(5,40) Z
40 FORMAT(F7.4)
A = M*Z+3.*Z*(Z-1.)
X = A*(M+Z)/(M+Z-1.) - M. - 1.
DO 70 K=1,30
N=IDINT(X)
H=X-N
B=1.
DO 50 J =1,N
B=J/A*B+1.
50 CONTINUE
B1=(N+1.)/A*B+1.
B2=(N+2.)/A*B1+1.
Y=B**(1.-H)*B1**H
Y=Y*(B1*B1/B/B2)**(H*(1.-H)/2.)
B=1./Y
G=(X+1.)/A + B
R=A*B-M
D=B*(X+M+1.-A-(X+M+1.)*DLOG(G))
A=A-R/D
70 X=A*(M+Z)/(M+Z-1.)-M-1
CONTINUE
X1=IDINT(10000.*X+.5)/10000.
A1=IDINT(10000.*A+.5)/10000.
WRITE(6,80)X1,A1
80 FORMAT(/,1X,'EQUIV. TRUNKS = ',

```

```

2      F8.4,/, 1X, 'EQUIV. LOAD = ', F8.4)
      STOP
      END

```

Blocking E.R.M.

```

      IMPLICIT REAL*8 (A-H, O-Z)
      REAL*8 M2, M2, M
      WRITE (6, 10)
10     FORMAT (/, 1X, 6X, 'BLOCKING E.R.M.',
2       //, 1X, 'TRUNKS = ')
      READ (5, 20) C
20     FORMAT (F7.4)
      WRITE (6, 30)
30     FORMAT( 1X, 'OFFERED LOAD = ')
      READ (5, 20) M
      WRITE (6, 40)
40     FORMAT ( 1X, 'PEAKEDNESS = ')
      READ (5, 50) Z
50     FORMAT (F7.4)
      A = M*Z + 3.*Z*(Z-1)
      X = A*(M+Z)/(M+Z-1.) - M-1.
      DO 80 K=1, 25
      CALL ERLNG (A, X, B)
      G = (X+1.)/A + B
      A = A - (A*B-M)/B/(X+M+1.-A-
2      (X+M+1.)*DLOG(G))
      X = A*(M+Z)/(M+Z-1.) - M-1.
80     CONTINUE
      X = X + C
      CALL ERLNG2 (A, X, B)
      P1 = B
      MO = A*B
      ZO = 1.-MO+A/(X+1.+MO-A)
      M2 = IDINT (10000.*MO+.5)/10000.
      Z2 = IDINT (10000.*ZO+.5)/10000.
      X=X-C
      CALL ERLNG(A, X, B)
      P = P1/B
      P2 = IDINT (10000.*P+.5)/10000.
      WRITE (6, 100) P2, M2, Z2
100    FORMAT (/, 1X, 'BLOCKING E.R.M. = ',
2      F6.4,/, 1X, 'OVERFLOW TRAFFIC = ',
3      F8.4,/, 1X, 'OVERFLOW PEAKEDNESS = ',
4      F8.4)
      STOP
      END

```

Three-Moment Match

```

      IMPLICIT REAL*8 (A-H, O-Z)
      REAL*8 L, L1, M, M1
      DIMENSION D(3)

```

```

WRITE (6,10)
10      2  FORMAT (/ ,1X,5X, 'THREE MOMENT',
           'MATCH',//,1X, 'EQUIV. TRUCKS = ',)
READ (5,20) X
20      FORMAT (F7.4)
WRITE (6,30)
30      FORMAT (1X, 'EQUIV. LOAD = ')
READ (5,20) A
B=1.
N=IDINT (X)
H=X-N
DO 50 J=1,N
B=J/A*B+1.
50      CONTINUE
B1 = (N+1)/A*B+1.
B2 = (N+2.)/A*B1+1.
Y = B**(1.-H)*B1**H
Y = Y*(B1*B1/B/B2)*(H*(1.-H)/2.)
D(1) = 1./Y
B3=(X+1.)/A*Y + 1.
D(2)=Y/A/(B3-Y)
B4=(X+2.)/A*B3 + 1.
D(3)=2./A/A*Y/D(2)/(B4-2*B3+Y)
L= D(3)*(D(2)-D(1))-D(1)*(D(3)-D(2))
L= L/(2.*D(2)-D(1)-D(3))*A
W= D(1)/L*(L-A*D(2))/(D(2)-D(1))
G= W/A*(L-A*D(1))/D(1)
L1= IDINT(10000.*L+.5)/10000.
W1= IDINT(10000.*W+.5)/10000.
G1= IDINT(10000.*G+.5)/10000.
M= L*W/(G+W)
M1= IDINT(10000.*M+.5)/10000.
Z= 1.+L*G/(G+W)/(1.+G+W)
Z1= IDINT(10000.*Z+.5)/10000.
80      WRITE (6,80) L1, W1, G1, M1, Z1
           2  FORMAT (/ ,1X, 'L=', F7.4,/, 1X,
           3  'W=', F7.4,/, 1X, 'G=', F7.4,/, 1X,
           'M=', F7.4,/, 1X, 'Z=', F7.4)
STOP
END

```

Blocking Day-to-Day Time

```

IMPLICIT REAL*8 (A-H,O-Z)
REAL*8 M,MB
WRITE (6,10)
10      2  FORMAT (/ ,1X,4X, 'BLOCKING-DAY TO',
           'DAY TIME',//,1X, 'TRUNKS = ')
READ (5,20) X
20      FORMAT (F7.4)
WRITE (6,30)
30      FORMAT (1X, 'MEAN LOAD = ')
READ (5,40) M

```

```

40      FORMAT (F7.4)
      WRITE (6,50) V
50      FORMAT (1X, 'VARIANCE = ')
      READ (5,60) V
60      FORMAT (F7.4)
      W1 = (1-DSQRT(V/(M*M+V)))/2.
      A = (M*M+V+DSQRT(V*(M*M+V)))/M
      A1 = A
      CALL ERLNG (A,X,B)
      MB = W1*B
      W2 = 1. - W1
      A = (M*M+V-DSQRT(V*(M*M+V)))/M
      A2 = A
      CALL ERLNG (A,X,B)
      MB = MB + W2*B
      C = IDINT (10000.*MB+.5)/10000.
      WRITE (6,90) C
90      FORMAT (/,1X, 'MEAN BLOCKING = ',
2         F6.4, ///)
      STOP
      END

```

Blocking Day-to-Day Call

```

      IMPLICIT REAL*8 (A-H,O-Z)
      REAL*8 M,MB
      WRITE (6,10)
10      2   FORMAT (/,1X,4X, 'BLOCKING-DAY TO',
      'DAY CALL',//,1X, 'TRUNKS = ')
      READ (5,20) X
20      FORMAT (F7.4)
      WRITE (6,30)
30      FORMAT (1X, 'MEAN LOAD = ')
      READ (5,40) M
40      FORMAT (F7.4)
      WRITE (6,50)
50      FORMAT (1X, 'VARIANCE = ')
      READ (5,60)
60      FORMAT (F7.4)
      W1 = (1.-DSQRT(V/(M*M+V)))/2.
      A = (M*M+V+DSQRT(V*(M*M+V)))/M
      A1 = A
      CALL ERLNG (A,X,B)
      MG = W1*A*B
      W2 = 1.-W1
      A = (M*M+V-DSQRT(V*(M*M+V)))/M
      A2 = A
      CALL ERLNG (A,X,B)
      MB = MB+W2*A*B
      C = IDINT (10000.*MB/M+.5)/10000.
      WRITE (6,90) C
90      FORMAT (/,1X, 'MEAN BLOCKING = ',

```

```

2      F6.4,///)
      STOP
      END

      SUBROUTINE ERLNG (A,X,B)
      IMPLICIT REAL*8 (A-H, O-Z)
      H=X-N
      B = 1.
      DO 30 J = 1,N
      B = J/A*B+1.
30     CONTINUE
      B1 = (N+1.)/A*B+1.
      B2 = (N+2.)/A*B1+1.
      Y = B** (1.-H)*B1**H
      Y = Y*(B1*B1/B/B2)**(H*(1.-H)/2.)
      B = 1./Y
      RETURN
      END

```

Waiting Time in GT/M/S Queue

```

      IMPLICIT REAL*8 (A-H, O-Z)
      REAL*8 L,M,N1
      INTEGER S
      DIMENSION D(3)
      WRITE (6,10)
10     FORMAT (/ ,1x, 'WAITING TIME IN GT/M/S',
2       'QUEUE',//, 1x, 'OFFERED LOAD = ')
      READ(5,20) M
20     FORMAT(F7.4)
      WRITE(6,30)
30     FORMAT(1x, 'PEAKEDNESS = ')
      READ(5,40) Z
40     FORMAT(F7.4)
      WRITE(6,50)
50     FORMAT(1X, 'NO. OF SERVERS = ')
      READ(5,60) S
60     FORMAT(I3)
      A = M*Z+3.*Z*(Z-1.)
      X = A*(M+Z)/(M+Z-1.)-M-1.
      DO 100 K=1,30
      N = X+A
      U2 = 1.
      Y = 1.
      DO 80 J = 1,N
      U2 = U2*(X-J+1)/A
      Y = Y+U2
      IF (DABS(U2).LE.1E-7) GO TO 90
80     CONTINUE
90     B = 1./Y
      G = (X+1.)/A+B
      N1 = A*B-M
      D1 = B*(X+M+1.-A-(X+M+1.)*DLOG(G).

```

```

A = A-N1/D1
X = A*(M+Z)/(M+Z-1.)-M-1.
100 CONTINUE
IF (DABS(U2) .GT. 1E-7) GO TO 200
D(1) = B
B3 = (X+1.) /A*Y+1.
D(2) = Y/A/(B3-Y)
B4 = (X+2.) /A*B3+1.
D(3) = 2. /A/A*Y/D(2)/(B4-2.*B3+Y)
L = D(3)*(D(2)-D(1))-D(1)*(D(3)-D(2))
L = L/(2.*D(2)-D(1)-D(3))*A
W = D(11)/L*(L-A*D(2))/(D(2)-D(1))
G = W/A*(L-A*D(1))/D(1)
V = (L+G+W)*S-S*S
V1 = (G+W)*S-L*W
D2 = (DSQRT(V*V+4.*V1*S*S)-V)
2 /2./S/S
C = 1.
E = 1.
U1 = 1./D2
DO 150 J = 1, S
E = E*(S-J+1.)/J
C = C*(1./F(J, L, W, G) - 1.)
U = E/C/F(J, L, W, G)*(S*F(J, L, W, G)
2 -J)/(S*D2-J)
U1 = U1 + U
150 CONTINUE
R = 1.-D2
P = 1./U1/D2
W1 = P/D2/S
WRITE(6, 170)R, P, W1
170 FORMAT(/, 1X, 'ROOT = ', F8.4, /,
2 1X, 'P(W>0) = ', F6.4, /,
3 1X, 'MEAN W = ', F8.4)
STOP
200 WRITE(6, 210)
210 FORMAT(/, 1X, 'NOT ACCURATE')
STOP
END
REAL FUNCTION F*8(J, L, W, G)
IMPLICIT REAL*8(A-H, O-Z)
REAL*8 L
F = (J*J+(G+W)*J)/(J*J+
2 (L+G+W)*J+L*W)
RETURN
END

```

AUTHOR

David L. Jagerman, B.E.E. (Electrical Engineering), 1949, Cooper Union; M.S., and Ph.D. (Mathematics), 1954 and 1962, respectively, New York University; AT&T Bell Laboratories, 1963—. Mr. Jagerman has been engaged in mathematical research on quadrature, interpolation, and approximation

theory, especially related to the theory of widths and metrical entropy, with application to the storage and transmission of information. For the past several years, he has worked on the theory of difference equations and queueing, especially with reference to traffic theory and computers. He is currently preparing a text on difference equations with application to stochastic models.

On the Capacity of Sticky Storage Devices

By H. S. WITSENHAUSEN*

(Manuscript received December 26, 1983)

When required to make a transition to a new state, a memory cell may, with a probability dependent on the state, refuse to do so (i.e., "stick"). Assuming that error correcting codes can be used at each write-read cycle, one seeks the maximum error-free (in the Shannon sense), long-term average capacity per cell and cycle. This problem is solved here for binary cells with either unilateral or symmetric stickiness. The methods used apply to more general cases as well. In the Appendix, some essential inequalities of dynamic programming are demonstrated.

I. INTRODUCTION

Information theoretic studies of storage devices have been mostly concerned with overcoming the existence of subsets of permanently defective cells.¹ We are concerned instead with the case of identical cells with the deficiency that use of a cell in one write-read cycle affects the cell's behavior in the next cycle. An extreme example of this is "write-once" memory,^{2,5,9} such as punched paper tape or optical disks, where the long-term average throughput per cell and cycle is, of course, zero. A previous paper⁸ considers a deterministic cell model in which the aftereffect of usage is of limited duration, permitting positive average rates. The present paper similarly considers the perhaps more realistic case of stochastic cells.

In this model, the store has N cells, each of which can be viewed as an input-output automaton. The store is used for T successive cycles.

* AT&T Bell Laboratories.

Copyright © 1984 AT&T. Photo reproduction for noncommercial use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

At each cycle, fresh data from a source S_t , independent of all earlier sources, are encoded into an N -vector \mathbf{X}_t of inputs applied to the cells. The resulting N -vector \mathbf{Y}_t of cell outputs goes to a decoder that must reproduce the source output with a probability of error approaching 0 as N increases. Note that if cell states can only change on inputs, not spontaneously, then the reading operation may be repeated any number of times in the same cycle with the same results. At each cycle only the probability distribution of the state of the store is known. This distribution, together with the cell model, defines a channel relating \mathbf{Y}_t to \mathbf{X}_t . The difficulty of the problem is that operating at or near the capacity of that channel at one cycle may leave an unfavorable state distribution for the next cycle. Thus, the largest possible number of bits per cell and cycle that are accurately recoverable can only be determined by dynamic programming. To make such a program manageable, a theorem is first proven that permits one to obtain the limit for large N at each cycle by considerations involving a single cell. This theorem applies to any cell for which state and output are one and the same.

Results are obtained for two types of binary cells. For the first type, only one of the two states is "sticky": when the cell is in that state and the input requires transition to the other state, there is a probability ϵ that the transition does not take place. For the second type, both states are assumed sticky in the same sense. The computations reveal that for both models the long-term policy is steady rather than periodic. For the symmetric case, this implies that the maximum throughput per cell and cycle is just the capacity of the binary symmetric channel with crossover probability one-half the sticking probability.

II. INFORMATION-THEORETIC BOUNDS

Let \mathcal{X} , \mathcal{Y} , and \mathcal{W} be finite alphabets, and let $X_t^i \in \mathcal{X}$ denote the input and $Y_t^i \in \mathcal{Y}$ the output of cell i at cycle t . Let $W_t^i \in \mathcal{W}$ be independent random variables representing the internal random effects of cell i at cycle t .

The initial state of cell i is represented by Y_0^i in \mathcal{Y} . We use the notation $\mathbf{X}_t = (X_t^1, \dots, X_t^N)$, $\mathbf{Y}_t = (Y_t^1, \dots, Y_t^N)$, $\mathbf{W}_t = (W_t^1, \dots, W_t^N)$. The quantities $\mathbf{X}_1, \dots, \mathbf{X}_T, Y_0^1, \dots, Y_0^N, W_1^1, \dots, W_T^N$ (a total of T vectors in \mathcal{X}^N , N variables in \mathcal{Y} , and NT variables in \mathcal{W}) are jointly independent.

The distributions of the Y_0^i and W_t^i are given, together with the equation describing the operation of the cells

$$Y_t^i = f_t^i(X_t^i, Y_{t-1}^i, W_t^i). \quad (1)$$

The distribution of the \mathbf{X}_t is dependent upon the choice of encoding, and the source statistics.

The case of interest is the "homogeneous" one where:

1. f_t^i is the same function f for all i and t .
2. The Y_0^i have the same distribution as a generic Y_0 for all i .
3. The W_t^i have the same distribution as a generic W for all i and t .

The more general case is allowed in the model because the main theorem below is valid without the homogeneity assumptions.

At each cycle t , the encoder and decoder are designed with knowledge of the distribution of past events but not of their realizations. Thus, neither encoder nor decoder knows the exact state of the memory (i.e., the vector \mathbf{Y}_{t-1}) at the end of the previous cycle. As \mathbf{W}_t is also known only in distribution, the memory appears as a noisy channel with input \mathbf{X}_t and output \mathbf{Y}_t . Then the entropy of the message that can be reconstructed with negligible error at cycle t is bounded, in view of the data-processing theorem,* by

$$I(\mathbf{X}_t; \mathbf{Y}_t),$$

where I is the mutual information. For the total throughput, one thus has the upper bound

$$\sum_{t=1}^T I(\mathbf{X}_t; \mathbf{Y}_t). \quad (2)$$

Theorem 1:

$$\sum_{t=1}^T I(\mathbf{X}_t; \mathbf{Y}_t) \leq \sum_{i=1}^N \sum_{t=1}^T I(X_t^i; Y_t^i). \quad (3)$$

Proof: By the independence of the initial states, the internal disturbances, and the sources, one has (H denoting entropy)

$$H(\mathbf{Y}_0) = \sum_{i=1}^N H(Y_0^i), \quad (4)$$

$$H(\mathbf{W}_t) = \sum_{i=1}^N H(W_t^i) \quad t = 1, \dots, T, \quad (5)$$

and also

$$H(X_t^i Y_t^i W_t^i Y_{t-1}^i) = H(X_t^i) + H(Y_{t-1}^i) + H(W_t^i), \quad (6)$$

since Y_t^i is determined by (1), and Y_{t-1}^i depends only on the initial state, the earlier source encodings, and the earlier disturbances, all independent of \mathbf{X}_t , hence of X_t^i . And for the same reason

* If X and Z are conditionally independent given Y , then $I(X; Z) \leq I(X; Y)$.

$$H(\mathbf{X}_t \mathbf{Y}_t \mathbf{W}_t \mathbf{Y}_{t-1}) = H(\mathbf{X}_t) + H(\mathbf{Y}_{t-1}) + H(\mathbf{W}_t). \quad (7)$$

The following four relations are standard properties of entropy:

$$H(\mathbf{X}_t \mathbf{Y}_t \mathbf{W}_t \mathbf{Y}_{t-1}) - H(\mathbf{X}_t \mathbf{Y}_t) = H(\mathbf{Y}_{t-1} \mathbf{W}_t | \mathbf{X}_t \mathbf{Y}_t) \quad (8)$$

$$H(X_t^i Y_t^i W_t^i Y_{t-1}^i) - H(X_t^i Y_t^i) = H(Y_{t-1}^i W_t^i | X_t^i Y_t^i) \quad (9)$$

$$H(\mathbf{Y}_{t-1} \mathbf{W}_t | \mathbf{X}_t \mathbf{Y}_t) \leq \sum_{i=1}^N H(Y_{t-1}^i W_t^i | X_t^i Y_t^i) \quad (10)$$

$$H(\mathbf{Y}_T) \leq \sum_{i=1}^N H(Y_T^i). \quad (11)$$

Using these facts one has

$$\begin{aligned} & \sum_{t=1}^T I(\mathbf{X}_t, \mathbf{Y}_t) \\ &= \sum_{t=1}^T H(\mathbf{X}_t) + H(\mathbf{Y}_t) - H(\mathbf{X}_t \mathbf{Y}_t) \text{ (by definition)} \\ &= \sum_{t=1}^T H(\mathbf{Y}_t) + H(\mathbf{X}_t \mathbf{Y}_t \mathbf{W}_t \mathbf{Y}_{t-1}) - H(\mathbf{X}_t \mathbf{Y}_t) - H(\mathbf{Y}_{t-1}) \\ &\quad - H(\mathbf{W}_t) \text{ [by (7)]} \\ &= H(\mathbf{Y}_T) - H(\mathbf{Y}_0) + \sum_{t=1}^T [H(\mathbf{Y}_{t-1} \mathbf{W}_t | \mathbf{X}_t \mathbf{Y}_t) - H(\mathbf{W}_t)] \\ &\quad \text{[by summation and (8)]} \\ &\leq \sum_{i=1}^N -[H(Y_T^i) - H(Y_0^i)] + \sum_{t=1}^T [H(Y_{t-1}^i W_t^i | X_t^i Y_t^i) - H(W_t^i)] \\ &\quad \text{[by (11), (4), (10), and (5)]} \\ &= \sum_{i=1}^N \sum_{t=1}^T (H(Y_t^i) - H(Y_{t-1}^i) - H(W_t^i) + H(Y_{t-1}^i W_t^i | X_t^i Y_t^i) - H(X_t^i Y_t^i)) \\ &\quad \text{[by summation and (9)]} \\ &= \sum_{i=1}^N \sum_{t=1}^T H(Y_t^i) + H(X_t^i) - H(X_t^i Y_t^i) \text{ [by (6)]} \\ &= \sum_{i=1}^N \sum_{t=1}^T I(X_t^i, Y_t^i) \text{ (by definition),} \end{aligned}$$

which was to be proved.

In the homogeneous case, the problem of finding the maximum M of $\sum_{t=1}^T I(X_t^i; Y_t^i)$ over all distributions of the independent variables

X_1^i, \dots, X_T^i subject to (1) is the same for all i . Having found M by solving this generic single-cell problem, a bound of NM is established for the total throughput. On the other hand, a throughput per cell, arbitrarily close to M , can be achieved for large enough N by choosing encoders with code words picked at random with each X_t^i independently having the maximizing distribution of the generic problem at stage t . Thus, the bound is sharp, asymptotically in N .

The bound is obtained, for any given cell model, by solving a T -stage, finite horizon, dynamic program. For large T it is the maximum long-term average per cycle (and cell) that is of prime interest. Some of the dynamic programming issues are discussed in the Appendix.

III. UNILATERAL STICKINESS

It is assumed that a binary cell acquires its input as new state when its previous state was 0, but when the previous state was 1 and the input is 0 the cell remains stuck at 1 with probability ϵ .

This is modeled, with $X_t, Y_t, W_t \in \{0, 1\}$ by letting $W_t = 1$ with probability ϵ and 0 with $1 - \epsilon$, and in ordinary arithmetic,

$$Y_t = X_t + (1 - X_t)Y_{t-1}W_t \quad t = 1, \dots, T. \quad (12)$$

Let

$$p_t = \Pr\{X_t = 0\}, \quad (12a)$$

$$s_t = \Pr\{Y_t = 0\}, \quad (12b)$$

and

$$\epsilon = \Pr\{W_t = 1\}. \quad (12c)$$

Then (12) implies

$$s_t = p_t(1 - \epsilon(1 - s_{t-1})), \quad (13)$$

and one obtains, with h the binary entropy function,

$$\begin{aligned} I(X_t; Y_t) &= h(p_t(1 - \epsilon(1 - s_{t-1})) - p_t h(\epsilon(1 - s_{t-1}))) \\ &= h(s_t) - p_t h\left(\frac{s_t}{p_t}\right). \end{aligned} \quad (14)$$

This leads to the dynamic program (where τ denotes the number of cycles to go)

$$\begin{aligned} V_\tau(s) &= \max_p h(p(1 - \epsilon + \epsilon s)) \\ &\quad - p h(\epsilon(1 - s)) + V_{\tau-1}(p(1 - \epsilon + \epsilon s)) \end{aligned} \quad (15)$$

with $V_0(s) = 0$.

This is an easy task for a computer, and the results show that p and s soon stabilize around steady-state values. Furthermore, as the Appendix shows, using the value functions found in the finite horizon solution, one can derive both upper and lower bounds on the optimal long-term average per cycle. In this problem these bounds soon agree to many decimals. The computer results are used only to conclude that the long-term optimum is steady. (This cannot be taken for granted, as often such problems have periodic solutions, crop rotation being the most ancient example of this.)

In a steady state, the constant values of p and s must satisfy, by (13), the relation

$$s = p(1 - \epsilon(1 - s)) \tag{16}$$

and by (14) the optimal throughput per cell and cycle is the maximum of

$$h(s) - ph \left(\frac{s}{p} \right) \tag{17}$$

subject to (16).

This amounts to maximizing a transcendental function on the unit interval. While no closed-form solution is known, the maximum, and the corresponding p and s are easily computed, and they are given as a function of ϵ in Table I. (The maximum is given in base 2, i.e., bits per cell and cycle.) It is important to note that for fixed T , the dynamic program (15) defines a total throughput which can only be approached from below as N increases. On the other hand, the long-range average is approached from above, because for finite T the transient effect of an initially clear memory will permit a slightly higher total.

As the case of small ϵ is of greatest interest we note the expansions [following from (16) and (17)]: at the maximum

$$s = \frac{1}{2} - \frac{\epsilon}{8} + o(\epsilon), \tag{18}$$

and

$$p = \frac{1}{2} + \frac{\epsilon}{8} + o(\epsilon), \tag{19}$$

and the value of the maximum is

$$\log 2 - \frac{1}{2} h \left(\frac{\epsilon}{2} \right) + o(\epsilon). \tag{20}$$

Table I—Results for unilateral sticking

Epsilon	Maximum	Input p	State s
0.00	1.00000000	0.50000000	0.50000000
0.01	0.97718297	0.50125629	0.49874371
0.02	0.95921284	0.50252532	0.49747469
0.03	0.94300294	0.50380733	0.49619267
0.04	0.92790483	0.50510257	0.49489743
0.05	0.91361310	0.50641131	0.49358869
0.06	0.89994925	0.50773381	0.49226619
0.07	0.88679616	0.50907034	0.49092966
0.08	0.87407111	0.51042119	0.48957881
0.09	0.86171265	0.51178665	0.48821335
0.10	0.84967335	0.51316702	0.48683298
0.15	0.79315764	0.52030370	0.47969631
0.20	0.74099706	0.52786405	0.47213596
0.25	0.69176025	0.53589839	0.46410162
0.30	0.64460887	0.54446658	0.45553343
0.35	0.59898428	0.55364065	0.44635936
0.40	0.55447912	0.56350833	0.43649168
0.45	0.51077418	0.57417812	0.42582189
0.50	0.46760281	0.58578644	0.41421357
0.55	0.42472834	0.59850838	0.40149163
0.60	0.38192756	0.61257412	0.38742589
0.65	0.33897650	0.62829543	0.37170459
0.70	0.29563534	0.64611064	0.35388938
0.75	0.25162917	0.66666668	0.33333334
0.80	0.20661826	0.69098302	0.30901701
0.85	0.16014419	0.72082550	0.27917453
0.90	0.11151237	0.75974695	0.24025309
0.95	0.05945393	0.81725604	0.18274403
0.96	0.04841568	0.83333337	0.16666671
0.97	0.03707783	0.85236594	0.14763415
0.98	0.02536387	0.87610072	0.12389940
0.99	0.01313736	0.90909100	0.09090918

IV. SYMMETRIC STICKINESS

Suppose that, when its input would require a change of state, a cell remains, with probability ϵ , in its former state.

This is modeled, with $X_t, Y_t, W_t \in \{0, 1\}$ by letting $W_t = 1$ with probability ϵ and

$$Y_t = (1 - W_t)X_t + W_t Y_{t-1}. \quad (21)$$

Let

$$p_t = \Pr\{X_t = 0\}, \quad (22a)$$

$$s_t = \Pr\{Y_t = 0\}, \quad (22b)$$

$$\epsilon = \Pr\{W_t = 1\}. \quad (22c)$$

Then (21) implies

$$s_t = (1 - \epsilon)p_t + \epsilon s_{t-1} \quad (23)$$

and

$$I(X_t; Y_t) = h((1 - \epsilon)p_t + \epsilon s_{t-1}) - p_t h(\epsilon(1 - s_{t-1})) + (1 - p_t)h(\epsilon s_{t-1}). \quad (24)$$

Note that, if both states are equally likely, then at the next cycle one faces a binary symmetric channel with crossover probability $\epsilon/2$. The capacity of this channel

$$\log 2 - h(\epsilon/2) \quad (25)$$

is attained by choosing symmetrically distributed input and this will lead to a symmetric distribution of the next state, so that this situation perpetuates itself. So (25) is an achievable long-term average, and it will turn out to be the best possible.

Assuming only that the optimum is time-invariant, one has, by (23), the condition

$$s = p, \quad (26)$$

and this reduces the problem to showing that the maximum, over $0 \leq s \leq 1$, of

$$h(s) - sh(\epsilon(1 - s)) - (1 - s)h(\epsilon s) \quad (27)$$

is at $s = 1/2$, where its value is $\log 2 - h(\epsilon/2)$. When $s = 0$ or 1 , (27) vanishes. For fixed s in $(0, 1)$ let

$$F(\epsilon) = \log 2 - h(\epsilon/2) - h(s) + (1 - s)h(\epsilon s) + sh(\epsilon(1 - s)). \quad (28)$$

It suffices to show that $F(\epsilon) \geq 0$ for $0 < \epsilon < 1$. One has $F(1) = F'(1) = 0$ and

$$F''(\epsilon) = \frac{(1 - \epsilon)(1 - 4s(1 - s))}{\epsilon(2 - \epsilon)(1 - \epsilon s)(1 - \epsilon(1 - s))} \geq 0. \quad (29)$$

So F is convex and tangent to the ϵ axis at $\epsilon = 1$, hence nonnegative, as required.

To confirm the steady-state solution, one runs the finite horizon dynamic program, with τ stages to go:

$$V_\tau(s) = \max_p h((1 - \epsilon)p + \epsilon s) - ph(\epsilon(1 - s)) - (1 - p)h(\epsilon s) + V_{\tau-1}((1 - \epsilon)p + \epsilon s) \quad (30)$$

with $V_0(s) \equiv 0$.

Using a grid of probabilities including $p = s = 1/2$, the upper and lower bounds on the long-term average (see Appendix), derived from the value functions, converge towards each other. In Table II the

Table II—Upper and lower bounds on optimal long-term average per cycle

Epsilon	Bounds	Formula
0.05	[0.83133907, 0.83133907]	0.8313390685
0.10	[0.71360304, 0.71360304]	0.7136030429
0.15	[0.61568846, 0.61568846]	0.6156884559
0.20	[0.53100441, 0.53100441]	0.5310044064
0.25	[0.45643556, 0.45643556]	0.4564355568
0.30	[0.39015970, 0.39015970]	0.3901596953
0.35	[0.33098416, 0.33098416]	0.3309841649
0.40	[0.27807191, 0.27807191]	0.2780719051
0.45	[0.23080717, 0.23080717]	0.2308071710
0.50	[0.18872188, 0.18872188]	0.1887218755
0.55	[0.15145182, 0.15145182]	0.1514518217
0.60	[0.11870910, 0.11870910]	0.1187091008
0.65	[0.09026388, 0.09026388]	0.0902638775
0.70	[0.06593194, 0.06593194]	0.0659319446
0.75	[0.04556600, 0.04556600]	0.0455659971
0.80	[0.02904941, 0.02904941]	0.0290494055
0.85	[0.01629174, 0.01629174]	0.0162917374
0.90	[0.00722555, 0.00722555]	0.0072255460
0.95	[0.00180412, 0.00180412]	0.0018041210

bounds shown have been obtained by iterating until the difference is below $5 \cdot 10^{-9}$ and they are compared with formula (25), all logarithms are in base 2.

V. CONCLUSIONS

While complex coding may be impractical, the value of the above results is that they provide precise bounds on what is achievable. One notes, comparing (20) with (25), that two-sided stickiness is, for the same small sticking probability ϵ , twice as damaging as one-sided stickiness. The case of two different positive sticking probabilities can be handled by the same techniques.

Note that the problem can be considered in another light by interchanging time and space. It then becomes a special case of the general interference channel, as shown in an earlier paper.⁸

REFERENCES

1. C. Heegard and A. A. El Gamal, "On the Capacity of Computer Memory With Defects" I.E.E.E. Trans. Inform. Theory, *IT-29*, No. 5 (September 1983), pp. 731-9.
2. C. Heegard, "On the Capacity of Permanent Memory," 1983 Conf. on Inform. Sciences and Systems, Johns Hopkins University (March 1983).
3. A. R. Odoni, "On Finding the Maximal Gain for Markov Decision Processes," Oper. Res., *17*, No. 5 (September-October 1969), pp. 857-60.
4. M. L. Puterman, ed., *Dynamic Programming and Its Applications*, New York: Academic Press, 1978.
5. R. L. Rivest and A. Shamir, "How to Reuse a 'Write-Once' Memory," Inform. Contr., *55*, No. 1-3 (October-November-December 1982), pp. 1-19.

6. D. J. White, *Finite Dynamic Programming*, New York: John Wiley and Sons, 1978.
7. P. Whittle, *Optimization Over Time*, Vol. I., New York: John Wiley and Sons, 1982.
8. H. S. Witsenhausen and A. D. Wyner, "On Storage Media With Aftereffects," *Inform. Contr.*, 56, No. 3 (March 1983), pp. 199-211.
9. J. K. Wolf et al., "Coding for a Write-Once Memory," *AT&T Bell. Lab. Tech. J.*, 63, No. 6, Part 2 (July-August 1984), pp. 1089-112.
10. N. A. J. Hastings, "Bounds on the Gain of a Markov Decision Process," *Oper. Res.*, 19, No. 1 (January-February 1971), pp. 240-4.

APPENDIX

Remarks on Dynamic Programming

A.1 Introduction

The literature on dynamic programming (e.g., Refs. 4, 6, 7) is primarily devoted to the important case of stochastic systems (controlled Markov chains). Deterministic dynamic programs are covered as a special case of degenerate probabilities. As a result, the theorems proven in most texts and papers make assumptions that are not necessary for the most important deterministic results we need. Among such assumptions the following two are particularly bothersome:

1. The state space must be finite.
2. All policies must be ergodic.

The first assumption rules out even the case of a real interval as state space (which is our case). The second is essentially never true in a deterministic setup. This is our justification for giving the very simple proof of the results we need in this paper.

We are concerned with the optimal long-term average return per stage. It has been claimed that this case is only academic because either the number of stages is small, and then a finite horizon treatment is appropriate, or it is large, and then the time span involved is such that the time value of utility must be included by using the discounted model instead. In this paper, as millions of memory cycles can take place per second, we have a strong counterexample to this viewpoint: the undiscounted, long-term average is the appropriate quantity to study.

A.2 The deterministic case

The deterministic programs are defined by two sets, S , U , and two functions: $r: S \times U \rightarrow R$, $f: S \times U \rightarrow S$. Here S is the state space, U the control variable set, $r(s, u)$ the return using $s \in S$, $u \in U$ and $f(s, u)$ is the next state. The finite horizon program is thus written

$$V_{\tau}(s) = \sup_{u \in U} r(s, u) + V_{\tau-1}(f(s, u)), \quad (31)$$

where τ is the number of stages to go and $V_0(s) \equiv 0$.

It has long been known that one can eliminate U , r , and f in favor

of a relation on states and a function defined on this relation. Specifically, let $\rho(s) \subset S$ be defined by

$$\rho(s) = f(s, U), \quad (32)$$

then the pairs $\{(s, s') \mid s' \in \rho(s)\}$ form a subset ρ of $S \times S$. Also, for $s' \in \rho(s)$, let

$$k(s, s') \equiv \sup_u \{r(s, u) \mid f(s, u) = s'\}. \quad (33)$$

In words, one need only know (i) which states one can go to next, and (ii) what is the optimal value of going there. The finite horizon problem then becomes

$$V_\tau(s) = \sup_{s' \in \rho(s)} k(s, s') + V_{\tau-1}(s'). \quad (34)$$

This reformulation (which is not possible for stochastic problems) has two advantages:

(i) Theoretical analysis is simplified.

(ii) In computation, when an infinite S is approximated by a finite subset S_1 , no requantization is needed, avoiding this cause of error accumulation. One solves the problem for the subset, with ρ and k restricted to $S_1 \times S_1$.

A.3 Bellman's equation for the long-term average

If λ denotes the optimal long-term average return per stage, then it should satisfy the equation, already given by Bellman

$$\lambda + W(s) = \sup_{s' \in \rho(s)} k(s, s') + W(s'). \quad (35)$$

for all $s \in S$.

The unknown constant λ and unknown function W (which matters only modulo the addition of an arbitrary constant) are reminiscent of an eigenvalue equation. However, the maximum operator, while non-linear, has a more favorable numerical behavior. Of course (35) is stated on the assumption that the required limit exists: it assumes that a policy for which the lim sup is as large as possible actually gives a limit. This depends on the structure of relation ρ , but we will not pursue this question, as our need is rather for inequalities.

A.4 Performance inequalities

Let V be any bounded real function on the state space S . For instance, V could be a value function obtained from a finite horizon program, an average of several such functions (in periodic cases), or just a plain wild guess. Then define \bar{V} by

(ii) By (36), for any s ,

$$\bar{V}(s) - V(s) \leq M \quad (47)$$

which, by (36), gives, for all s and $s' \in \rho(s)$,

$$k(s, s') + V(s') - V(s) \leq M. \quad (48)$$

Thus if s_0, s_1, s_2, \dots is any admissible sequences [i.e., $s_{t+1} \in \rho(s_t)$ for all t], then

$$\frac{k(s_0, s_1) + V(s_1) - V(s_0) \leq M,}{k(s_{t-1}, s_t) + V(s_t) - V(s_{t-1}) \leq M.} \quad (49)$$

By addition

$$\sum_{i=1}^t k(s_{i-1}, s_i) + V(s_t) - V(s_0) \leq Mt. \quad (50)$$

Thus by (40), dividing by t

$$\frac{1}{t} \sum_{i=1}^t k(s_{i-1}, s_i) \leq M + \frac{c}{t} \quad (51)$$

so that

$$\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t k(s_{i-1}, s_i) \leq M, \quad (52)$$

which was to be proved.

These inequalities have been known at least since the publication of Refs. 3 and 10, although their proof is submerged under unnecessary assumptions. The inequalities are crucial to any computer treatment of the long-run average. In addition, the above theorem proves the sufficiency of Bellman's equation (35), for if this equation holds with a bounded V and some λ , then for this V one obtains $m = M = \lambda$, and thus λ is optimal and achievable.

AUTHOR

Hans S. Witsenhausen, Ph.D. (Electrical Engineering), 1966, The Massachusetts Institute of Technology; AT&T Bell Laboratories, 1966—. Mr. Witsenhausen has been associated with the Université Libre de Bruxelles, with the European Computation Center, Brussels, and with the Princeton Research Division of Electronic Associates, Inc. At MIT he was with the Electronic Systems Laboratory as a Lincoln Laboratory Associate and a Hertz Fellow. At AT&T Bell Laboratories, he is associated with the Mathematics and Statistics Research Center. He was a Senior Fellow at the Imperial College of Science and Technology, London, in 1972, a Visiting Professor at MIT in 1973, and Vinton Hayes Senior Fellow at Harvard University in 1975-76. He has worked on problems of hybrid computation, control theory, optimization, geometric inequalities, and other applied mathematical fields.

Velocity-Saturated Characteristics of Short-Channel MOSFETs

By G. W. TAYLOR*

(Manuscript received October 3, 1984)

A theory is developed for the I-V characteristics of metal oxide semiconductor field-effect transistors (MOSFETs) when the channel fields are sufficiently high to cause appreciable saturation of the carrier drift velocity. The full velocity-field curve for bulk silicon is used with the base value adjusted to account for surface scattering effects. Use of this form gave the best fit to experimental data. Using some simple expansions to reduce the rather complex integral produces a useful analytic result, which gives a continuous description from the square law results for long-channel devices throughout the whole range of velocity-saturated operation in short-channel devices. For the first time the electron temperature has been introduced as the parameter, which increases the channel charge at pinch-off, decreases the saturation voltage, and increases the channel field at the pinch-off point as the current (and hence bias voltages) is increased. The effects of series resistance and surface roughness scattering are incorporated into the analytic formulation. We compare the results with experimental submicron devices and find excellent agreement.

I. INTRODUCTION

As the metal oxide semiconductor field-effect transistor (MOSFET) evolves towards submicron channel dimensions it is found that, below gate lengths of $5 \mu\text{m}$, its performance is modified by the effects of hot-electron scattering on carrier transport. When a device reaches a gate dimension of $0.5 \mu\text{m}$, its complete "on" region is dominated by velocity

* AT&T Bell Laboratories.

Copyright © 1984 AT&T. Photo reproduction for noncommercial use is permitted without payment of royalty provided that each reproduction is done without alteration and that the Journal reference and copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free by computer-based and other information-service systems without further permission. Permission to reproduce or republish any other portion of this paper must be obtained from the Editor.

saturation. Almost all descriptions of these phenomena in the literature are based on two-dimensional numerical solutions of total device operation.¹⁻⁴ Such an approach provides insight into device operation and the nature of its physics, but it is of little practical use from a designer's point of view, since it can realistically consider only a single device at a time. Generally, circuit simulation programs must use empirical fitting routines to generate device characteristics in the velocity-saturated mode because there is no analytical solution of this behavior. Some estimates of velocity-saturated conduction have been based on the simplified model of a constant mobility up until the saturation velocity is achieved,^{5,6} although some of these results have been useful over certain particular voltage ranges with parameter fitting, the interpretation of the physics has been lacking. In fact, academically, a more rigorous treatment of the problem greatly helps us visualize the ultimate limits of device miniaturization. A proper description of device performance in velocity saturation must be established as a natural extension of the device characteristics before velocity saturation becomes significant. However, since an analytical solution can be found only for certain special cases, one is usually forced to assume a two-section model as just described. Such an approach unavoidably leads to discontinuities in the incremental parameters and, particularly, in regions of crossover. In addition, it precludes prediction of the pinch-off field, which is fixed at $\mathcal{E} = \mathcal{E}_c$ in this model. The other missing feature in previous work has been the electron temperature. Since the carrier transport is due to electrons moving with their saturated drift velocity and hence at elevated temperatures, the electron temperature must be an integral part of the representation.

In this paper we use the velocity-field relationship that most nearly fits the experimental data as a starting point to develop a model. The charge transport description is derived from basic principles and incorporates the electron temperature as a natural part of the solution. It is shown in detail how the velocity-saturated characteristics are a natural extension of the constant mobility case. We will consider the case of both the triode and the saturation regions; for the saturation case the details of channel length modulation under velocity-saturated conditions are derived in Appendix A. The effects of series contact resistance are incorporated into the theory, and the results are compared with data over a wide range of parameters.

II. THEORY

2.1 *The velocity-field relationship*

When the drift velocity can no longer be considered a linear function of the electric field, modifications to the metal oxide semiconductor

(MOS) theory must be introduced. Many functional forms have been used to characterize the decrease in mobility at higher electric-field strengths.⁷⁻¹³ For the case of bulk Si, Conwell has shown that the assumption of a Maxwellian distribution with electron temperature, T_e , for the symmetrical part of the electron distribution function yields predictions that agree with the experiment.¹⁴ The result for the energy-loss rate, $B(T_e)$, for phonon scattering is

$$B(T_e) = q\mu_0 \left(\frac{T_0}{T_e}\right)^{1/2} \left(\frac{v_s}{\mu_0}\right)^2 \frac{(T_e - T_0)}{T_0}, \quad (1)$$

where μ_0 is the zero field mobility, T_0 is the lattice temperature, v_s is the saturation velocity of electrons, and q is the electronic charge. Now it has been shown that the general momentum and energy relations are

$$j = qn\mu(T_e)\mathcal{E} + q \frac{d}{dx} [nD(T_e)] \quad (2a)$$

$$j\mathcal{E} = nB(T_e) + \frac{j}{q} \delta(T_e)k \frac{dT_e}{dx}, \quad (2b)$$

respectively, where \mathcal{E} is the electric-field strength; $\mu(T_e)$ is the temperature, or field-dependent mobility; j is the electronic current density; k is Boltzmann's constant; n is the electron density; and $\delta(T_e)kT_e$ is the average kinetic energy transported per electron.¹⁵ Since we are dealing with the MOS surface along which the scattering mechanisms are not well known, (1) may not be appropriate; we will continue to use it, however, to expedite the analysis. The second term on the right-hand side of (2b) is equivalent to $dS(T_e)/dx$. Here $S(T_e)$ is the flux of energy in the positive x direction; we have taken $S = -(j/q)\delta(T_e)kT_e$ only and ignored the small contribution $-K(T_e)[(dT_e)/(dx)]$ because of the thermal conductivity of the electrons. Note that δ typically is approximately a constant for a particular relaxation time relation and for acoustic phonon scattering $\delta \approx 2$.¹⁵ Equations (2a) and (2b) have three unknowns— \mathcal{E} , n , and T_e —so that the third relation of Poisson's equation for the MOS channel region is required to obtain a complete solution. The most familiar form of this equation is the one-dimensional charge equation (the gradual channel approximation) for the MOS channel, which is

$$q \int_0^\infty n \, dy = C_o(V_{GS} - V_T - V), \quad (2c)$$

where V is related to \mathcal{E} by

$$\int_0^x \mathcal{E} \, dx' = V$$

and C_o is the oxide capacitance. These equations are solved by simplification of (2a). The right-hand side of (2a) is the sum of the drift and diffusion components of the device current, the drift component being characterized by a drift velocity,

$$v = \mu(T_e) \mathcal{E} \quad (3)$$

It is well known for conduction above the threshold voltage that over most of the MOS channel, drift is the dominant component, and hence, only the first term on the right-hand side of (2a) needs to be retained. We can use this result in (2b); then by neglecting the second term on the right-hand side of (2b) and substituting from (1) for $B(T_e)$, we obtain the result

$$q\mu_0 \left(\frac{T_0}{T_e}\right)^{1/2} \left(\frac{T_e}{T_0} - 1\right) \left(\frac{v_s}{\mu_0}\right)^2 = q \mathcal{E}^2 \mu.$$

Solving for T_e we find

$$T_e = T_0 \left\{ \frac{(\mathcal{E}/\mathcal{E}_c)^2}{2} \frac{\mu}{\mu_0} + \left[\frac{(\mathcal{E}/\mathcal{E}_c)^4}{4} \left(\frac{\mu}{\mu_0}\right)^2 + 1 \right]^{1/2} \right\}^2, \quad (4)$$

where \mathcal{E}_c is the critical field parameter,

$$\mathcal{E}_c = v_s/\mu_0. \quad (5)$$

We can then use this result to evaluate the relative importance of the neglected term in (2b). Using (2c) in (2a) we can write the current-field relation for the MOS channel, which is

$$I = \mu(\mathcal{E})C_o(V_{GS} - V_T - V) \mathcal{E} \quad (6a)$$

If we then use a typical mobility

$$\mu = \mu_0(T_0/T_e)^{1/2} \quad (6b)$$

such as one might find in a bulk crystal, we can use (4) and (6a) to evaluate the relative importance of the terms in (2b), and we find

$$\frac{j}{q} \frac{\delta(T_e)k}{\delta} \frac{dT_e}{dx} \approx 2\delta(T_e) \frac{\frac{kT_0}{q} \left[1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c}\right)^2 \right]}{(V_{GS} - V_T - V)} \left(\frac{\mathcal{E}}{\mathcal{E}_c}\right)^2. \quad (6c)$$

This ratio is small until we are close to pinch-off in the channel or unless $\mathcal{E} \geq \mathcal{E}_c$. The restriction on this term comes from neglecting the diffusion term in (2a). We will assume that (4) is valid throughout the MOS channel.

Equation (4) has been derived without reference to a specific function, $\mu(\mathcal{E})$. We therefore can use (4) to determine T_e , and we must still determine $\mu(\mathcal{E})$. To determine $\mu(\mathcal{E})$, we could just assume the

bulk relation (6b), which would then yield the results

$$\mu = \frac{\mu_0}{\sqrt{1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c}\right)^2}} \quad (7a)$$

and

$$T_e = T_0[1 + (\mathcal{E}/\mathcal{E}_c)^2]. \quad (7b)$$

However, there has been much speculation that this form is inadequate for the Si surface mobility,⁷⁻¹⁴ and a variation of this form that has been used fairly widely is

$$\mu(\mathcal{E}) = \frac{\mu_0}{\left[1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c}\right)^B\right]^{1/B}}, \quad (7c)$$

where B was determined by Caughey and Thomas¹³ to be $B \approx 1.1$. To illustrate the variation of $\mu(\mathcal{E})$, plots were generated and are shown in Fig. 1 for values of B ranging from 1 to ∞ . The important parameters— μ_0 , \mathcal{E}_c , and v_s —are identified in the figure.

Because the determination of B at this stage in our understanding of MOS surface physics is unavoidably experimental, we must, at the outset, choose a value for B that will allow a physical solution to be found. This is necessary because the use of an undetermined B leads to a mathematically hopeless situation. The vindication of this approach ultimately will come from an unambiguous determination of

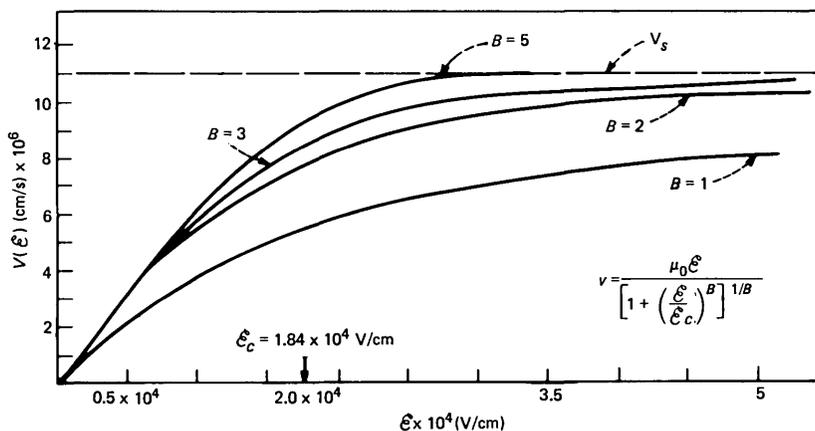


Fig. 1—Variation of $v(\mathcal{E})$ for the formula (7c) with values of B ranging from 1 to ∞ . A value of $\mu_0 = 650 \text{ cm}^2/\text{V-s}$ was used here simply to clarify the variation of v with electric field.

the parameters μ_0 , \mathcal{E}_c , v_s , and B from the experimental data. Since it is suspected that $B = 2$ for N channel devices (which is of main interest here), we will give the greatest attention to the treatment of this form. In our final comparison with experiment we will demonstrate that the value of $B \simeq 2$ is a good one.

In the remainder of the paper we will discuss the effects of a field-dependent mobility, $\mu(\mathcal{E})$, on the MOS I-V characteristic and the resultant modifications to the conventional device laws ($B = \infty$) for $B = 2$ generally and with some mention of the case $B = 1$. The theory will show, as is well known in short-channel devices, that when the drift velocity of carriers approaches the saturated value, the current from the conventional case, is reduced substantially.

The theory will also show the well-known experimental result that the drain saturation voltage for the velocity-saturated case is substantially less than the conventional value. It is shown that the decrease in voltage is caused by the increase in the charge in the channel at pinch-off under hot-electron conditions resulting from the elevated electron temperature.

2.2 Device characteristics

2.2.1 Triode region

Figure 2 gives a device cross section, which shows terminal voltages and possible series resistances in the source and drain leads. When the general mobility relationship (7) is used, only for the special case of $B = 1$ can a closed form solution for the current-voltage characteristic be obtained. If we use (5) and (7b) for $B = 1$, the result shown by Hoeneisen⁸ is

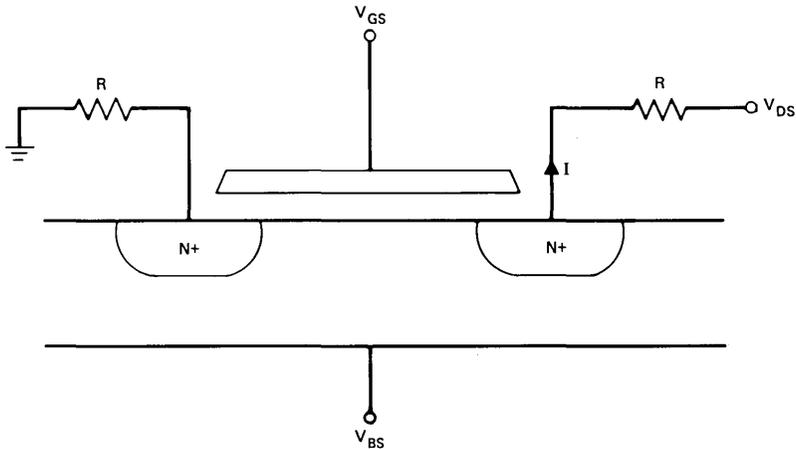


Fig. 2—Device cross section showing terminal voltages and possible parasitic series resistors.

$$I = \frac{\mu_0 C_o W}{L} \left[\frac{(V_{GS} - V_T) V_{DS} - V_{DS}^2}{1 + V_{DS}/\mathcal{L}_c L} \right]. \quad (8)$$

For the case of $B = 2$, the MOS device current, when approximated by the drift mechanism only, yields

$$I = \frac{\mu_0 W}{\left[1 + \left(\frac{\mathcal{L}}{\mathcal{L}_c} \right)^2 \right]^{1/2}} C_o (V_{GS} - V_T - V) \mathcal{L}. \quad (9a)$$

By rearranging terms and changing variables, we obtain

$$\frac{IL}{\mu_0 WC_o} = \int_{V_{GS} - V_T - V_{DS}}^{V_{GS} - V_T} (z^2 - I'^2)^{1/2} dz, \quad (9b)$$

where

$$I' = \frac{I}{v_s WC_o}, \quad z = V_{GS} - V_T - V.$$

Performing the integration, we find

$$I = \frac{\mu_0 WC_o}{2L} \left\{ (V_{GS} - V_T) \sqrt{(V_{GS} - V_T)^2 - \left(\frac{I}{Wv_s C_o} \right)^2} \right. \\ - (V_{GS} - V_T - V_{DS}) \sqrt{(V_{GS} - V_T - V_{DS})^2 - \left(\frac{I}{Wv_s C_o} \right)^2} \\ \left. - \left(\frac{I}{Wv_s C_o} \right)^2 \right. \\ \left. \cdot \ln \left[\frac{V_{GS} - V_T + \sqrt{(V_{GS} - V_T)^2 - \left(\frac{I}{Wv_s C_o} \right)^2}}{V_{GS} - V_T - V_{DS} + \sqrt{(V_{GS} - V_T - V_{DS})^2 - \left(\frac{I}{Wv_s C_o} \right)^2}} \right] \right\}. \quad (10a)$$

Previously, it has been shown that (for the constant mobility case) it is possible to incorporate the effects of the bulk charge variation into V_T .¹⁶ As the gate voltage increases into the range where velocity saturation becomes significant, the threshold voltage modulation by the drain voltage becomes relatively unimportant. We will therefore assume that we can continue to use the drain-voltage dependence of V_T that was developed for the constant mobility case. Equation (10a) describes the device characteristic exactly up until pinch-off has been achieved but, because of its implicit nature, it is of relatively little use. However, we produce useful results with some simple approximations. Consider the terms under the second square root sign. This quantity

must be positive to have a physically meaningful result, so that

$$\left(\frac{I}{Wv_s C_o}\right)^2 \leq (V_{GS} - V_T - V_{DS})^2.$$

Now the current on the left-hand side is a maximum and the right-hand side is a minimum at the pinch-off condition, and if the equality sign were obeyed at this point we would have

$$I = C_o W v_s \bar{Q}. \quad (10b)$$

In writing (10b) we have used the definition of \bar{Q} as the channel charge at the pinch-off point, as we develop in Appendix A [see (83)]. From this result we see that the inequality rather than the equality is always obeyed, since (10b) requires that all of the charge be moving with the saturated velocity at the pinch-off point. This condition can never be reached because of the contribution of diffusion to the current flow. We therefore note that under all square root signs in (10) we will have positive quantities and also that the voltage term is generally much larger than the term I' . We may therefore make expansions of all square root quantities, and by retaining terms to the first order, we obtain

$$I \approx \frac{1}{2} \mu_0 \frac{W}{L} C_o \left\{ (V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2 - I'^2 \ln \left[\frac{2(V_{GS} - V_T) - \frac{I'^2}{V_{GS} - V_T}}{2(V_{GS} - V_T - V_{DS}) - \frac{I'^2}{V_{GS} - V_T - V_{DS}}} \right] \right\}. \quad (10c)$$

For the moment we will neglect terms in I' in the argument of the logarithm without incurring significant error, and we may then solve the resulting quadratic equation for I to obtain

$$I = W v_s C_o \cdot \left\{ \left[\left(\frac{\mathcal{L}_c L}{a} \right)^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a} \right]^{1/2} - \frac{\mathcal{L}_c L}{a} \right\}, \quad (11a)$$

where

$$a \equiv \ln \left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}} \right). \quad (11b)$$

This expression for the current is very useful because it allows us to predict the two limiting forms of conduction that are always observed in a short-channel device. For small gate (drain voltages) the term in square brackets is small, and we may expand under the square

root sign to obtain

$$I \approx \frac{\mu_0 C_o W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right], \quad (12)$$

which is precisely the form of the constant mobility case that we expect to obtain when the effects of velocity saturation can be ignored. On the other hand, when this term becomes large, only this term needs to be retained, and we find

$$I \approx \frac{v_s C_o W}{\sqrt{a}} \sqrt{(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2}}. \quad (13)$$

This equation applies to only a limited range, since typically there is only a small range of drain voltage in which velocity saturation is dominant before pinch-off occurs. This will become evident in our comparison of the exact and approximate solutions.

2.2.2 Saturation region

Traditionally, in MOSFET physics the terms operation in pinch-off and operation in saturation were interchangeable because it was generally thought that pinch-off was necessary for saturation to occur. For some time now it has been recognized that diffusion becomes important near the pinch-off point.¹⁷ It is shown in Appendix A [see (79) through (83)] that the saturation voltage is reached when a particular ratio, \bar{R} , of drift to diffusion conduction is reached in the channel. \bar{R} is derived in Appendix B [see (200) and (201)] in terms of the basic device parameters. The transfer of the conduction mechanism from drift to diffusion is the extent of any pinching effect, since beyond this point in the channel the drift component increases once again, the oxide field reverses in direction, and the current moves away from the surface into the bulk. Saturation of current occurs basically because of the occurrence of field reversal at some drain voltage. The saturation point and the field-reversal point are separated in potential in the channel by a potential, \bar{Q}/C_o , where \bar{Q} is the channel charge (mobile charge) at the saturation point. For the constant mobility case the charge, \bar{Q} , is small, so that the potential difference between the two points is small in that case. Hence, there really was no need to discriminate between them. As velocity saturation becomes pronounced, the charge, \bar{Q} (and thus voltage, \bar{Q}/C_o), grows considerably, leading to a substantial change in V_{SAT} (the drain saturation voltage) from its conventional value. Thus, the potential difference between the saturation point and the field-reversal point continues to grow, as does the physical separation between the two points, although the potential will increase more since the field is also increasing. From

this point on we will refer to the point in the channel where the diffusion current is $\bar{R}I$ as the saturation point or the pinch-off point, interchangeably. We will distinguish between it and the field-reversal point.

Under conditions of hot-electron flow (as in Appendix A) we will continue to use the Einstein relation

$$D(\mathcal{E}) = \frac{kT_e}{q} \mu(\mathcal{E}) \quad (14)$$

to relate the mobility and diffusion coefficients through the effective electron temperature, T_e . This has recently been shown to be a reasonable assumption if it is recognized that, for mobilities significantly below μ_0 , the low field value, the electron temperature will be higher than the lattice temperature.¹⁸ Also, according to (2a) we will consider the term

$$D(\mathcal{E})dQ/dx \gg Q \frac{dD(\mathcal{E})}{dx},$$

i.e., the diffusion constant is not a strong function of the electric field.

The result for the saturation voltage obtained in Appendix A [cf. (82)] is

$$\bar{V} \equiv V_{\text{SAT}} = V_{\text{GS}} - V_T - \left(\frac{1 - \bar{R}}{\bar{R}} \right) \frac{kT_e}{q} \frac{(C_o + \bar{C}_s F)}{C_o}, \quad (15)$$

where F is the short-channel factor¹⁶ at the threshold condition and \bar{C}_s is the semiconductor depletion capacitance at the pinch-off point. Each quantity denoted by a bar will signify its value at the pinch-off point. From this result we can see that the charge at the pinch-off point now continues to grow as we move into the hot-electron regime, linearly with the electron temperature. Since we are using $B = 2$ and have shown for this case that T_e is given by (7b), then we have

$$\bar{Q} = \frac{(1 - \bar{R})}{\bar{R}} (C_o + \bar{C}_s F) \frac{kT_0}{q} \left[1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 \right]. \quad (16)$$

The value of \mathcal{E} , the electric field at the pinch-off position, is determined from the drift relation

$$\mathcal{E} = \frac{I(1 - \bar{R})}{W\mu\bar{Q}}. \quad (17a)$$

Using (17a) and (7a) in (16) we obtain

$$\left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 \left[1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 \right] = \left[\frac{(1 - \bar{R})I}{W\mathcal{E}_c\mu_0\bar{Q}_0} \right]^2, \quad (17b)$$

which has the solution

$$\frac{\mathcal{E}}{\mathcal{E}_c} = \left(\left\{ \frac{1}{4} + \left[\frac{I(1 - \bar{R})}{Wv_s\bar{Q}_0} \right]^2 \right\}^{1/2} - \frac{1}{2} \right)^{1/2}, \quad (17c)$$

where

$$\bar{Q}_0 = \frac{(1 - \bar{R})}{\bar{R}} \frac{kT_0}{q} (C_o + \bar{C}_s F)$$

is the value of charge at the pinch-off point in the absence of hot-electron effects.¹⁹ This is an interesting result because it encompasses all modes of operation from the region of low currents without velocity saturation through the regime of high fields dominated by velocity saturation. For low currents we can expand the square root and (keeping only the first term) find that, without velocity saturation, we have

$$\mathcal{E} \approx \frac{I(1 - \bar{R})}{W\mu_0\bar{Q}_0}. \quad (17d)$$

If the current (and therefore the channel field) is high (velocity saturation becomes dominant), then we have

$$\mathcal{E} \approx \sqrt{\frac{I(1 - \bar{R})\mathcal{E}_c}{W\mu_0\bar{Q}_0}}, \quad (17e)$$

so that the field changes from a linear to a square root dependence upon the device current. We note that from (17d) and (17e) the transition from one to the other takes place at $\mathcal{E} = \mathcal{E}_c$, as we would expect. We can also determine from (17c) that the device current that flows when $\mathcal{E} = \mathcal{E}_c$ is

$$I = \sqrt{2} \frac{Wv_s\bar{Q}_0}{(1 - \bar{R})}.$$

The interesting feature of this result is the independence of channel length. We should, thus, expect devices of all gate lengths to change to velocity-saturated behavior at the same value of channel current. If we now use (17c) in (16) we obtain

$$\bar{Q} = \frac{(1 - \bar{R})}{\bar{R}} (C_o + \bar{C}_s F) \frac{kT_0}{q} \left[\frac{1}{2} + \sqrt{\frac{1}{4} + \frac{I^2(1 - \bar{R})^2}{(Wv_s\bar{Q}_0)^2}} \right], \quad (18a)$$

and again we can obtain the limiting cases

$$\bar{Q} = \bar{Q}_0$$

and

$$\bar{Q} = \frac{I(1 - \bar{R})}{Wv_s} \quad (18b)$$

for low and high currents, respectively. A useful approximation to (18a) is found to be

$$\bar{Q} = \bar{Q}_0 \sqrt{1 + \left[\frac{I(1 - \bar{R})}{Wv_s \bar{Q}_0} \right]^2}, \quad (18c)$$

or

$$\bar{Q} = \bar{Q}_0 + \frac{I(1 - \bar{R})}{Wv_s} \quad (18d)$$

expressions, which are mathematically more convenient. The approximation will introduce some error when the two terms under the root sign are equal, but it yields the correct value in the extremes. Using the saturation voltage (15), we may now determine the saturation current; using (8) and (11a) we obtain for the cases of $B = 1$ and $B = 2$, respectively,

$$I_{SAT} = \frac{\mu_0 C_o W}{2\bar{y}} \left[\frac{(V_{GS} - V_T)^2 - (\bar{Q}/C_o)^2}{1 + \frac{V_{GS} - V_T - \bar{Q}/C_o}{\mathcal{E}_c \bar{y}}} \right], \quad B = 1 \quad (19a)$$

and

$$I_{SAT} = Wv_s C_o \cdot \left[\sqrt{\left(\frac{\mathcal{E}_c \bar{y}}{a} \right)^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}/C_o)^2}{a}} - \frac{\mathcal{E}_c \bar{y}}{a} \right], \quad B = 2, \quad (19b)$$

where \bar{y} is the position of the pinch-off point in the channel. Since we do not have the equivalent to (6b) for the case of $B = 1$, we cannot determine T_e or, therefore, \bar{Q}/C_o ; thus, we can proceed no further with that case.

Consider the case of $B = 2$. By substituting (18c) into (19b) and rearranging, we find

$$I^2 + 2 \frac{(\mathcal{E}_c \bar{y} v_s C_o W)}{(1 + a)} I - \frac{(v_s C_o W)^2}{(1 + a)} \left[(V_{GS} - V_T)^2 - \left(\frac{\bar{Q}_0}{C_o} \right) \right] = 0, \quad (20a)$$

which has the solution

$$I_{SAT} = Wv_s C_o \cdot \left\{ \left[\left(\frac{\mathcal{E}_c \bar{y}}{a^*} \right)^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a^*} \right]^{1/2} - \frac{\mathcal{E}_c \bar{y}}{a^*} \right\}, \quad (20b)$$

where

$$a^* \equiv a + 1. \quad (20c)$$

By comparison with (11a) it can be seen that the expressions for the current in the triode region and saturation region are very similar. The currents are identical at the pinch-off position because the change from the parameter a to a^* exactly compensates for the change from \bar{Q}/C_o to \bar{Q}_o/C_o . In using (18c) in (19b) we have omitted the factor $(1 - \bar{R})$. We must do this to obtain consistent results because (19b) was derived on the basis of drift alone.

As in the case of the triode region, (10b) can predict two limiting cases. For small gate voltages, the appropriate expansion of the square root yields

$$I \simeq \frac{\mu_o C_o W}{2\bar{y}} \left[(V_{GS} - V_T)^2 - \left(\frac{\bar{Q}_o}{C_o} \right)^2 \right], \quad (21a)$$

as we would expect in the absence of hot-electron effects. On the other hand, in the extreme of large gate voltages we obtain the result

$$I = \frac{v_s C_o W}{\sqrt{a^*}} (V_{GS} - V_T). \quad (21b)$$

An interesting feature of this equation is the absence of \bar{y} and therefore the absence of channel-length modulation effects. The interpretation is that for high enough gate voltages and/or short enough gate lengths, the velocity of carriers at the source approaches the saturation velocity. In this situation the current is determined only by mobile charge at the source, which is independent of gate length and depends only on the gate voltage. Although we can approach this situation, we could not achieve it in practice because of the onset of breakdown and punchthrough effects.

Returning to (15) and using our approximate value of \bar{Q} from (18c), we find the saturation voltage to be

$$\bar{V} \equiv V_{SAT} = V_{GS} - V_T - \frac{\bar{Q}_o}{C_o} \sqrt{1 + \left(\frac{I}{Wv_s \bar{Q}_o} \right)^2}. \quad (22)$$

Using (20b) in (22), we then find

$$V_{SAT} = V_{GS} - V_T - \left[\left(\frac{\bar{Q}_o}{C_o} \right)^2 \left(1 - \frac{1}{a^*} \right) + 2 \left(\frac{\mathcal{E}_c \bar{y}}{a^*} \right)^2 + \frac{1}{a^*} \right. \\ \left. \cdot (V_{GS} - V_T)^2 - \frac{2 \mathcal{E}_c \bar{y}}{a^*} \sqrt{\left(\frac{\mathcal{E}_c \bar{y}}{a^*} \right)^2 + \frac{1}{a^*} (V_{GS} - V_T)^2 \left(\frac{\bar{Q}_o}{C_o} \right)^2} \right]^{1/2}. \quad (23)$$

This result is almost the same as one obtained in the literature using a piecewise continuous constant mobility and velocity-saturated model; i.e., from the continuity of current at a field \mathcal{E}_c in the channel one has

$$\mu_0 \frac{W}{\bar{y}} C_o \left[(V_{GS} - V_T) V_{SAT} - \frac{V_{SAT}^2}{2} \right] = v_s C_o W (V_{GS} - V_T - V_{SAT}), \quad (24)$$

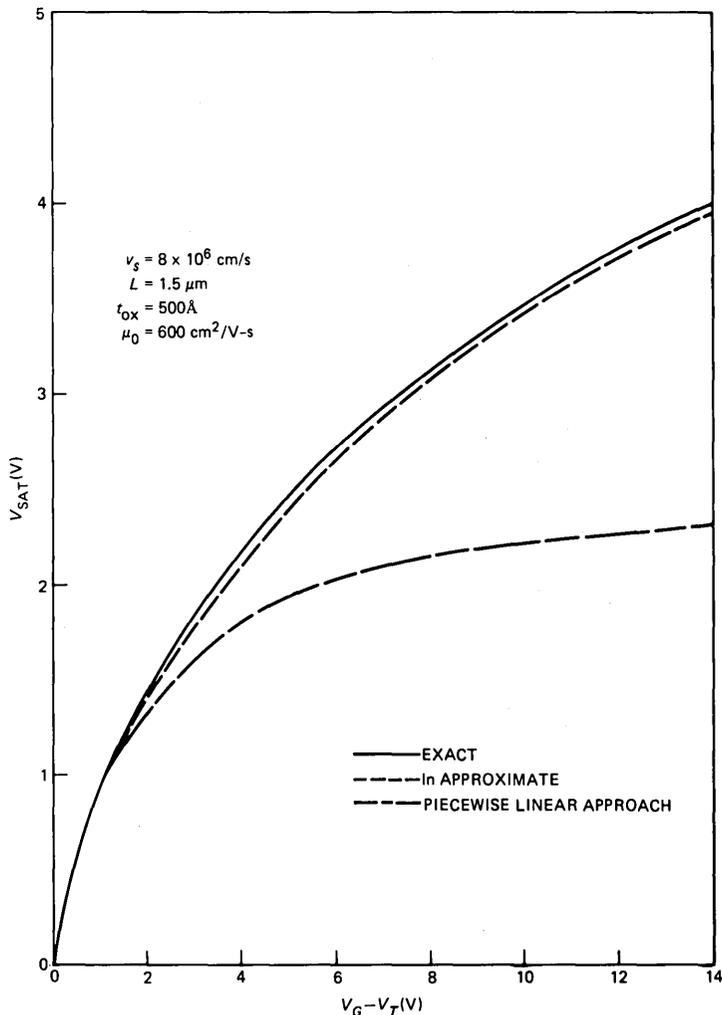


Fig. 3—Variation of V_{SAT} with gate voltage showing the complete solution using $B = 2$, the approximate solution using $B = 2$, and the piecewise linear constant mobility result.

which may be solved to yield

$$V_{\text{SAT}} = V_{\text{GS}} - V_T + \mathcal{E}_c \bar{y} - \sqrt{(V_{\text{GS}} - V_T)^2 + (\mathcal{E}_c \bar{y})^2}. \quad (25)$$

These two results are compared in Fig. 3 and it is seen that over much of the range of velocity saturation, (25) is rather inaccurate. The use of (25) to predict the current can then lead to errors of about 20 percent, as we show later in Fig. 9. By using (23) we obtain a good approximation to the exact result [the form used for a is described later by (33a)]. The differences between these two curves arise simply from the errors introduced in making the expansions to solve the original current equation (10a). The only drawback to (23) is that some iteration is required at the transition from the triode to the saturation region since a^* is a weak function of V_{SAT} through the logarithmic term. As we can see from (10b), this voltage enters only weakly into the current itself through a^* , but is necessary to determine where the saturation and triode regions meet and to provide an accurate merging of the two. In the calculation of I_{SAT} it should be noted that it is not necessary to calculate (23). Since the current is given explicitly by (20b), then (22) can be used to find V_{SAT} . It would also be possible to treat a as a constant parameter to be determined so that no iteration is required; however, it is shown later that the variation in a is important in achieving continuity of g_m .

In line with the reasoning that led to (24) it is of some interest to consider the field \mathcal{E}_c as a kind of dividing line in the channel between the non-velocity-saturated and the velocity-saturated portions. Using (9a)

$$V_{\text{GS}} - V_T - V_C = \frac{\sqrt{2}I}{v_s C_o W} \quad (26)$$

is the condition in the channel when a field of $\mathcal{E} = \mathcal{E}_c$ is reached at a channel potential of $V = V_C$. We can then use (10b) to give V_C in the triode region, which is

$$V_C = V_{\text{GS}} - V_T + \frac{\sqrt{2} \mathcal{E}_c L}{a} - \sqrt{\frac{2}{a} (V_{\text{GS}} - V_T)^2 - \frac{2}{a} (V_{\text{GS}} - V_T - V_{\text{DS}})^2 + \left(\frac{\sqrt{2} \mathcal{E}_c L}{a}\right)^2}, \quad (27)$$

where a is given by (11b) and is a known function of V_{GS} and V_{DS} . We can then use this result in our expression (11a) for the current to find the position in the channel at which this field is achieved; i.e., using (26) in (11a) yields

$$y_c = \sqrt{2} \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_C)^2 - \frac{(V_{GS} - V_T - V_C)^2}{2} a}{\mathcal{E}_c(V_{GS} - V_T - V_C)}, \quad (28)$$

where V_C is given by (27). If we follow the same procedure for operation in saturation then we obtain from (20b) and (26)

$$V_{GS} - V_T - V_C = \left\{ \left(\frac{\sqrt{2} \mathcal{E}_c \bar{y}}{a^*} \right)^2 + 2 \left[(V_{GS} - V_T)^2 - \left(\frac{\bar{Q}_0}{C_o} \right)^2 \right]^{1/2} - \frac{\sqrt{2} \mathcal{E}_c \bar{y}}{a^*} \right\}, \quad (29)$$

which is similar in form to (27) but now depends on \bar{y} . The solution (28) applies to saturation or triode operation. Consider the situation $y_c = 0$, which implies either that the carriers are subjected to a field $> \mathcal{E}_c$ at all points in the channel or that they are moving with a velocity of $v_s/\sqrt{2}$ at the source end of the channel. Using (28) to find $V_{GS} - V_T - V_C$ and (11a) for I we find

$$I = \frac{W v_s C_o}{\sqrt{a^* + 1}} (V_{GS} - V_T),$$

i.e., we are very close to the velocity-saturated limit of (21b). It is then of some interest to determine the applied voltage for which this condition is achieved as a function gate length. Using (20b) we find

$$(V_{GS} - V_T) = 2 \mathcal{E}_c L \sqrt{a^* + 1}. \quad (30)$$

This voltage value is plotted as a function of gate length in Fig. 4 using the values of a^* as determined by (33a) and (33b) (and shown plotted in Fig. 9). For $L = 0.5 \mu\text{m}$ this condition is not achieved until 3V are applied to the device. These results simply point to the fact that the use of L/v_s to determine transit times usually gives values that are too small by a factor of about 2.

2.3 Discussion of results

2.3.1 I-V curves and current saturation

The interpretation of the phenomenon of the saturation of the drain current presented here is the following. Classical saturation of the current occurs when the charge in the channel has been reduced to the minimum value given by (16). This condition always occurs in the channel just before inversion in sign of the transverse oxide field occurs; i.e., pinch-off occurs at a channel potential given by (15), and when this potential has increased by another \bar{Q}/C_o volts, we have zero voltage across the oxide, resulting in field inversion. Therefore, pinch-

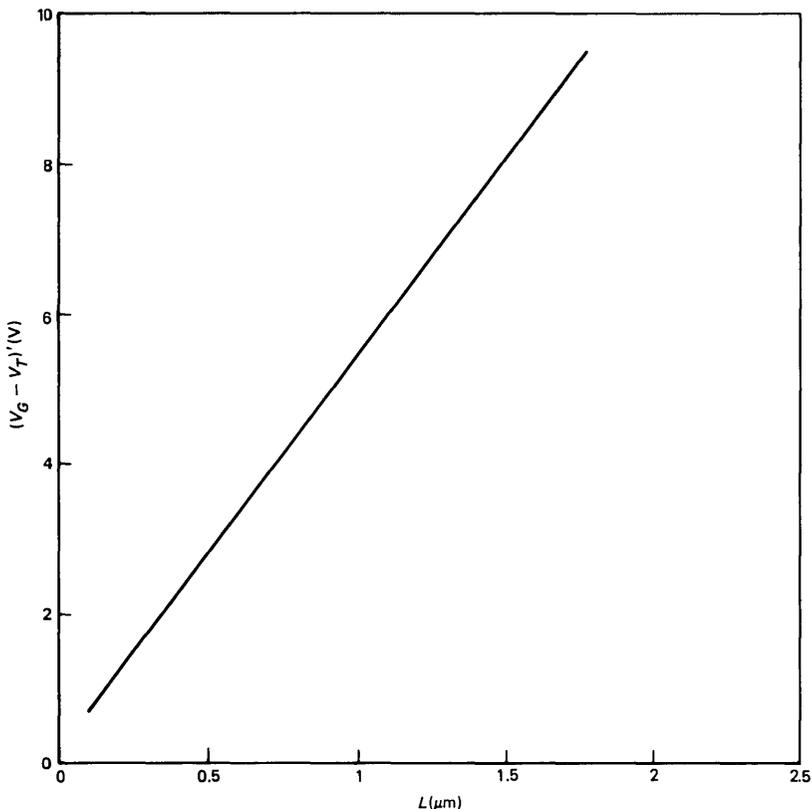


Fig. 4—Variation of $(V_{GS} - V_T)'$, the gate voltage for which the source velocity is $v_s/\sqrt{2}$, with gate length.

off in the channel will always be associated with a field inversion in the oxide. Now in the case without velocity saturation the voltage \bar{Q}/C_o is typically 0.05V, which is dropped over a fairly short distance (of the order of 0.1 μm) in the channel, so that it is clear that pinch-off and field inversion occur at almost the same point in the channel. When velocity saturation becomes dominant, \bar{Q}/C_o becomes a sizeable term compared to $V_{GS} - V_T$ because of the dependence of \bar{Q} upon the electron temperature. This is the reason for the apparently lower drain saturation voltage in the velocity-saturated regime, as we can see from (15) and the plots in Fig. 3. Under these conditions the voltage drop in the channel between the pinch-off point and the field inversion point, and the distance between these two points may grow considerably from the constant mobility case.

The parameter a requires more attention for the computation of accurate results. From the results (12) and (21a) it is clear that a is

an important parameter only in the limit of velocity saturation. Consider the value of a at the saturation voltage for which case it has its greatest effect. Taking a in its complete form from (10c) we have

$$a \cong \ln \left[\frac{V_{GS} - V_T + (V_{GS} - V_T) \sqrt{1 - \frac{1}{1+a}}}{\bar{Q}/C_o + \bar{Q}_o/C_o} \right], \quad (31)$$

where we have used (18c) in the denominator and also the fact that in the limit of velocity saturation we have from (20b) or (11a)

$$\frac{I}{Wu_s C_o} \simeq \frac{V_{GS} - V_T}{\sqrt{1+a}}. \quad (32)$$

From this result we see that in the limit of a saturated drift velocity if $a < 1$, then we should be able to express a as a function of only $(V_{GS} - V_T)/(\bar{Q}/C_o)$. The result we obtain should be useful over the whole range of operation because, although determined from velocity-saturated conditions, a disappears from the expressions for the current when velocity saturation becomes unimportant. We have found by iteration that the result

$$a = 1.24 \ln \left(\frac{V_{GS} - V_T}{V_{GS} - V_T - V_{DS}} \right) \quad (33a)$$

in the triode region, which goes to

$$a = 1.24 \ln \left(\frac{V_{GS} - V_T}{\bar{Q}/C_o} \right) \quad (33b)$$

at the saturation voltage, gives excellent agreement over a wide range of parameters. The factor 1.24 was chosen to match the approximate and exact formulas for the current well into velocity-saturated operation both in the triode and saturation current regions. It was not chosen to match the exact and approximate forms of (31) itself so that it could be used in a more general way to compensate for all errors involved in the expansions in (10b). The corresponding value of a^* is found by using (33b) in (20c).

These calculations are shown for some typical devices in Figs. 5, 6, and 7 over a wide range of gate voltages, channel lengths, and gate oxide thicknesses. The approximate and exact predictions of the saturation voltage are also shown. It is noted that the logarithmic dependence of a upon drain voltage in (33) must be included to obtain an accurate result. The current therefore is a totally explicit function of the device voltages; the only iteration required is in the calculation of V_{SAT} , as we mentioned earlier. As the figure indicates, the approximations are very good.

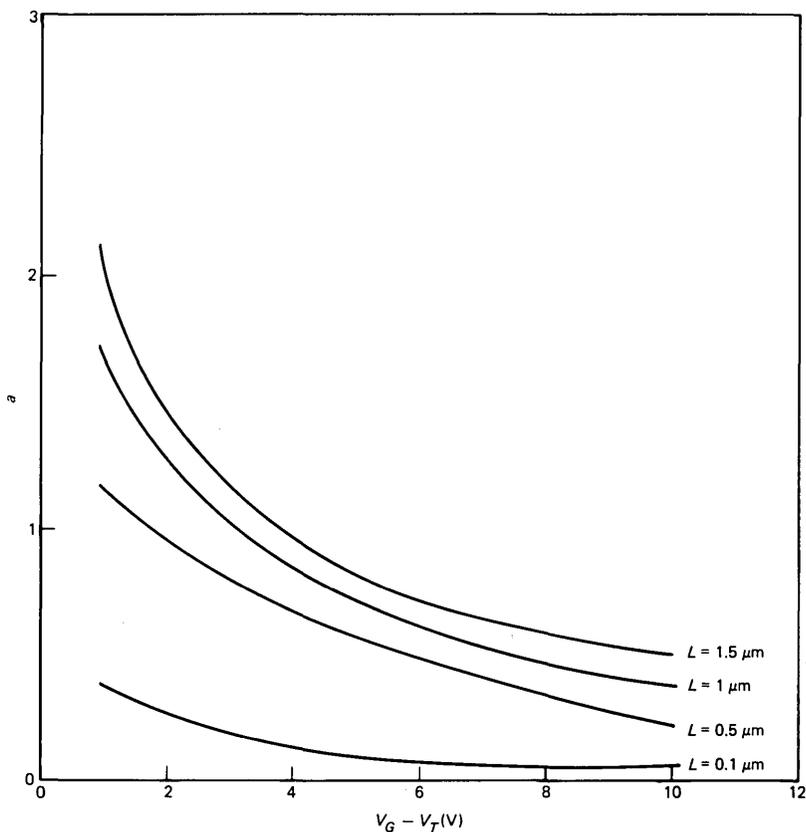


Fig. 5—Variation of the voltage parameter a at saturation with gate voltage for various gate lengths.

Figure 8 shows the conventional square law result. For the 1.5- μm device, for example, the effects of velocity saturation are relatively unimportant for $V_{GS} = 2$ but become progressively more significant as the gate voltage is increased, so that for $V_{GS} = 6$ there is a great difference between both the current and the saturation voltage. It is clear how the effects of a saturated drift velocity have reduced the current available from the device for a given supply voltage. It is for this reason that increasing the supply voltage has very little effect on gate propagation delay beyond some particular value of voltage. From (20b) we see that there is a significant departure from the square law when

$$V_{DD} \approx V_T + \frac{\mathcal{L}_c L}{\sqrt{2a}} \quad \text{or} \quad \approx V_T + \mathcal{L}_c L \quad (34)$$

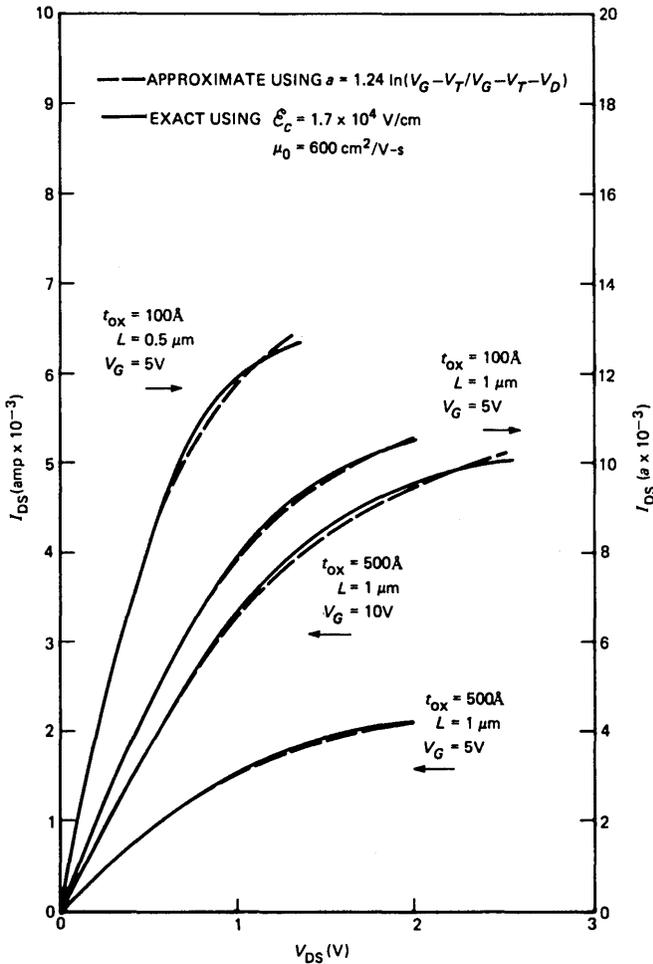


Fig. 6—Comparison of the rigorous and approximate solutions for a wide range of oxide thickness gate length and gate voltage. The voltage V_{SAT} is shown by the arrows that indicate the end of the line.

for typical values of a . We should, therefore, like to design for a supply voltage of approximately (34) since we would then have the maximum speed and the minimum power dissipation for a given channel length of the driver in an inverter for example. However, in practical applications of short-channel devices the supply voltage will be higher than this value to ensure adequate noise margins and a sufficient ratio of $V_{DD}:V_T$ due to processing tolerances on V_T . This will definitely be the case if V_T is increased intentionally in order to suppress subthreshold leakage or if the supply voltage must be held arbitrarily at 5V to provide transistor-transistor logic (TTL) compatibility.

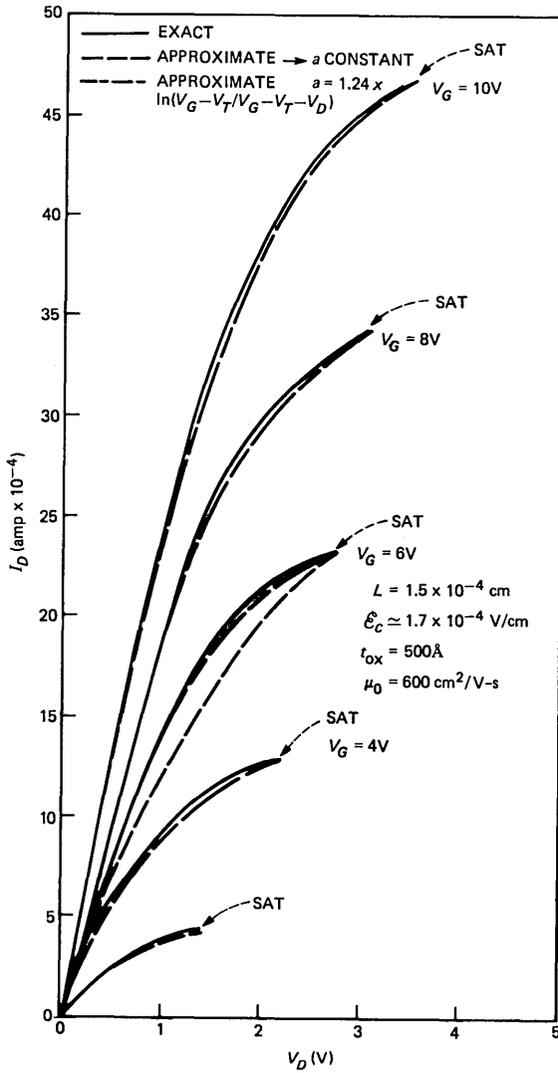


Fig. 7—Comparison of device characteristics for $L = 1.5 \mu\text{m}$ using rigorous and approximate forms. Also, shown for $V_G = 6\text{V}$ is the result of treating the \ln term as constant.

The other point of interest in Fig. 8 is the comparison of the results obtained using the piecewise-linear model of the velocity-field characteristic and the more physical model presented here. The discrepancy (indicated by the arrows in the figure) grows at first quickly and then less rapidly as the gate voltage is increased; i.e., for lower gate voltages such that $V_{GS} - V_T \sim V_{SAT}$ the percent error in the current will be greater than that for $V_{GS} - V_T \gg V_{SAT}$. This is shown by the

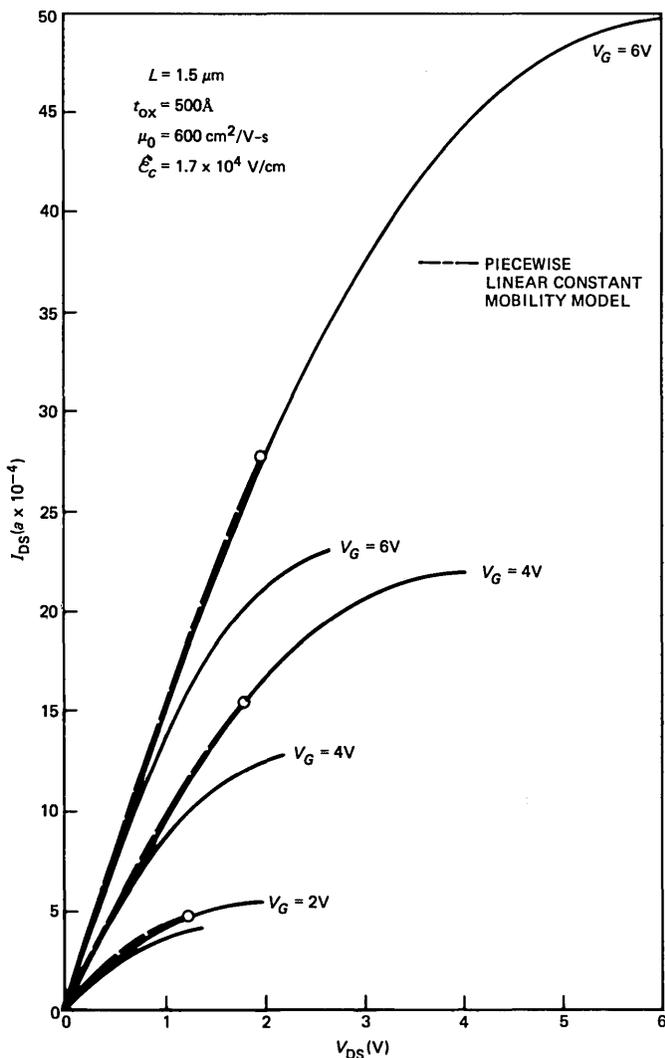


Fig. 8—Comparison of velocity-saturated model for $B = 2$ with constant mobility model. Also shown is the piecewise constant mobility model denoted by circles at saturation.

plot in Fig. 9 for the device in Fig. 8. The maximum in this curve is expected since the errors should be worst when the field at the drain is \mathcal{E}_c because then the differences in the models are at maximum.

2.3.2 Substrate bias dependence

In the triode region the substrate bias totally enters through the threshold voltage as the body effect. It is also known that the threshold

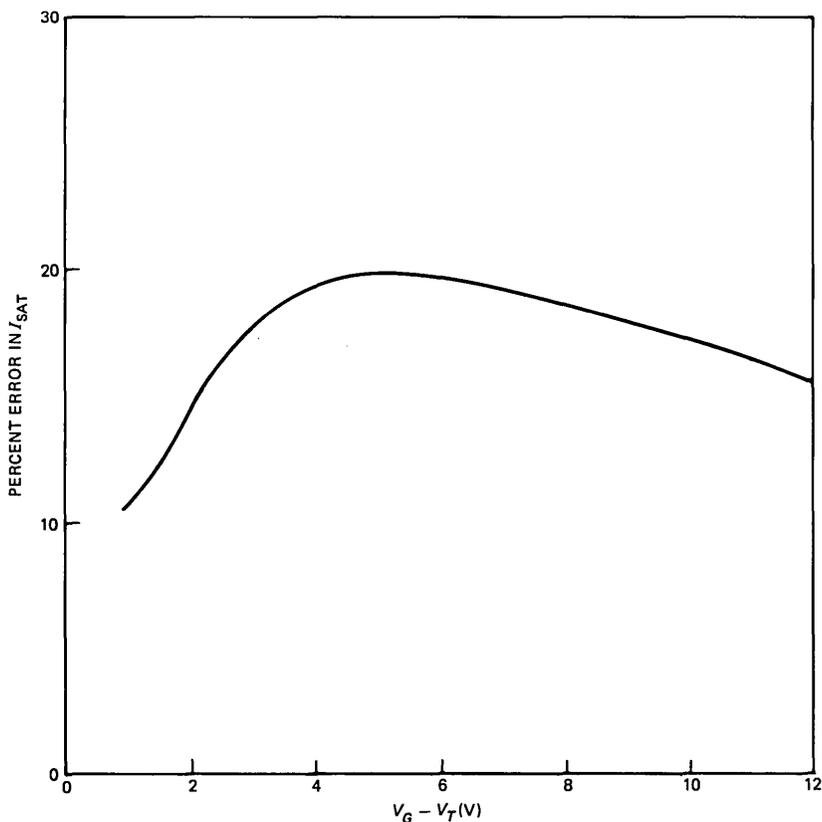


Fig. 9—Error incurred using the piecewise constant mobility model as a function of gate voltage.

voltage is dependent on drain voltage in a short-channel device and this dependence is increased by the application of substrate bias. These effects will be considered in the discussion of experimental results.

2.3.3 Effects of contact resistance and a gate-voltage-dependent mobility

In a real short-channel device, unavoidable series contact resistances can introduce great differences between actual and expected device currents. Also the mobility parameter, μ_0 , exhibits a gate voltage dependence due to the increase in surface scattering with gate bias.²⁰ We will consider both of the effects here since, although unrelated, they affect the device current in the same way. We will consider a series resistance, R , in the drain and the source leads as shown in Fig. 2, and a mobility dependence of

$$\mu_0 = \frac{\mu_{00}}{1 + \theta(V_{GS} - V_T)}. \quad (35)$$

The parameter θ is an empirical constant that determines the dependence of the mobility upon the normal channel field²⁰ and is thus expected to have some substrate bias dependence. With these additions, one can show from (10c) that the triode region result is modified to

$$\frac{I_{TR}}{Wv_s C_o} = \sqrt{\left[\frac{\mathcal{L}'_c L}{a} + \frac{2R'}{a} \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) \right]^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a}} - \left[\frac{\mathcal{L}'_c L}{a} + \frac{2R'}{a} \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right) \right], \quad (36)$$

and the saturation region result is modified to

$$\frac{I_{SAT}}{Wv_s C_o} = \sqrt{\left[\frac{\mathcal{L}'_c \bar{y}}{a'} + \frac{2R'}{a'} \left(\frac{\bar{Q}_0}{C_o} + \frac{V_{GS} - V_T}{2} \right) \right]^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a'}} - \left[\frac{\mathcal{L}'_c \bar{y}}{a'} + \frac{2R'}{a'} \left(\frac{\bar{Q}_0}{C_o} + \frac{V_{GS} - V_T}{2} \right) \right], \quad (37a)$$

where

$$R' = RWv_s C_o \quad (37b)$$

$$\mathcal{L}'_c = \mathcal{L}_c [1 + \theta(V_{GS} - V_T)] \quad (37c)$$

and

$$a' = 1 + a + 2RWv_s C_o. \quad (37d)$$

In writing (37a) we have used both (18c) and (18d) to express the linear and square root terms in \bar{Q}/C_o that appear when (22) is substituted in (36). Therefore, we cannot expect a perfect match between (36) and (37a) at the saturation point, although it will be close. To achieve identical values we would have to use (18c) only and then iterate (37a) to determine I_{SAT} .

The effects of the series resistance and the mobility reduction are similar since they both act to increase the first term under the square root sign in (36) and (37a). From (36d) it is seen that the mobility degradation can be interpreted simply as a movement of \mathcal{L}_c to higher values on the velocity field curve, as shown in Fig. 10. This has been

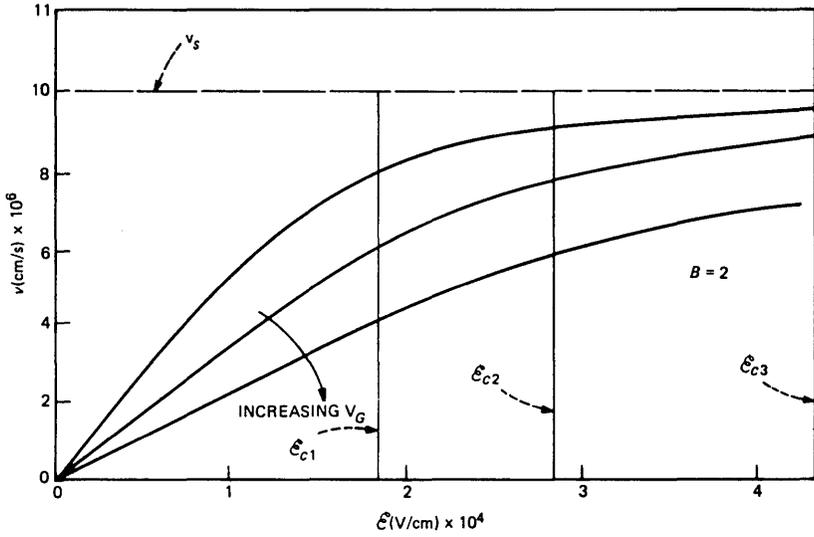


Fig. 10—Variation of \mathcal{L}_c as gate voltage is increased because of the reduction in μ shifting the point of velocity saturation to higher \mathcal{L}_c .

noted elsewhere for the constant mobility case⁶ and can be generalized to any velocity-field relation.

It is interesting to consider the limiting forms of these two results. From (36) by expanding the square root term in the limit of small gate and drain voltages, we obtain

$$I \approx \frac{1}{2} \left[\frac{2(V_{GS} - V_T)V_{DS} - V_{DS}^2}{\mathcal{L}_c L + 2R' \left(V_{GS} - V_T - \frac{V_{DS}}{2} \right)} \right]. \quad (38)$$

If we now consider the case of small drain voltages, which describes the region in which we normally assess the dependence of μ_0 upon $V_{GS} - V_T$, we obtain

$$I = \frac{\mu_{00} C_o W}{L} \frac{(V_{GS} - V_T)V_{DS}}{1 + \left(\theta + 2R \frac{\mu_{00} C_o W}{L} \right) (V_{GS} - V_T)}. \quad (39)$$

From this result it is clear that one must be careful when extracting a physically meaningful value of θ from experimental data since an accurate value of R must first be known. This can be shown clearly by the data in Fig. 11 of the linear region current of three devices that are identical except for the gate length. Each curve is characterized by a section at lower gate voltages, which is linear, and a section at higher gate voltages, which is nonlinear in gate voltage. The gate voltage at

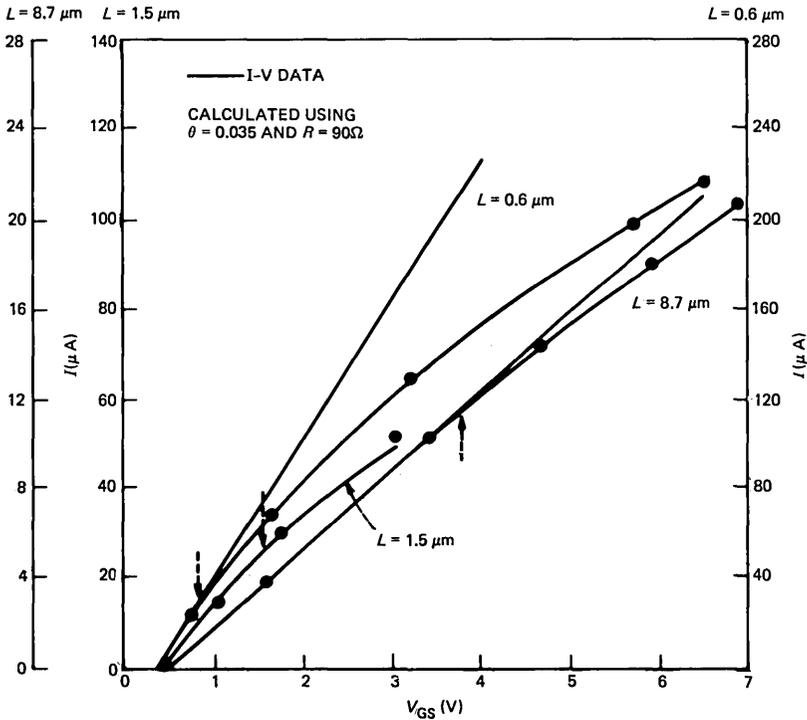


Fig. 11—Linear region current data showing the effect of R and θ . The experimental device parameters are $t_{ox} = 500\text{\AA}$, $N_A = 1 \times 10^{16} \text{ cm}^{-3}$, and $\mu_{00} \approx 560 \text{ cm}^2/\text{V}\cdot\text{s}$.

which the sections join is marked in the figure by an arrow. The most important feature of this voltage is that it moves to lower and lower values for decreasing L . From (35) we see that this breakpoint should be fixed in gate voltage—the reason for the movement to lower values is the increasing importance of the term in R in (38) for decreasing L . If we use $\theta = 0.035$ and a value of $R = 90\Omega$, then we can fit the curves as indicated. To give some idea of the importance of R , for the long-channel device the choice of R is immaterial within limits since θ is dominant. For the smallest length ($L \approx 0.5 \mu\text{m}$) the choice of θ is not important because the R term is dominant. Thus, as channel lengths are reduced it becomes increasingly important to reduce the series R . Both the contact resistance and the mobility reduction will cause V_{SAT} to increase for a given V_G .

Figure 12 shows the effects of a contact resistance of 50Ω on the device current for a typical short-channel device. To write down a closed form for I , we have ignored the dependence of a upon R in arriving at (36) and (37a). The difference this makes is indicated in Fig. 12.

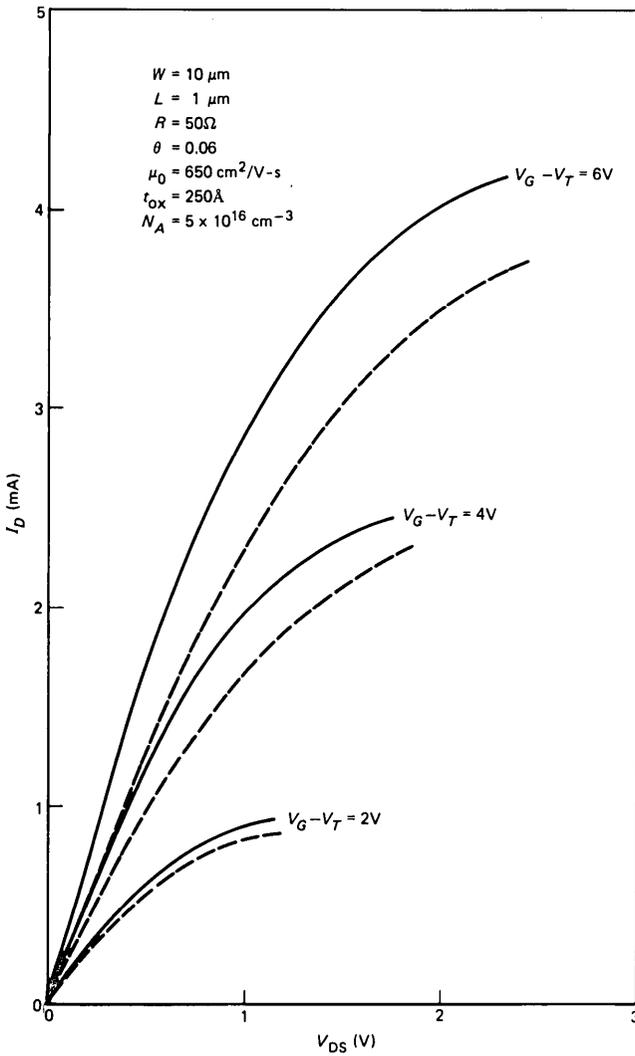


Fig. 12—Effect of series resistance on drain characteristics of a short-channel device. The curves are calculated using (36) with $R = 0$ (solid line) and $R = 505\Omega$ (dashed line). The second dashed line shows the effects of neglecting any effect of a upon R .

2.3.4 Incremental parameters and the continuity of derivatives

2.3.4.1 Current-saturated operation. Since the current as derived in (20b) is a continuous function of gate voltage in moving from non-velocity-saturated to velocity-saturated operation, then the same is true for transconductance, and we obtain

$$\begin{aligned}
 g_m = \frac{dI}{dV_{GS}} = Wv_s C_o \left(\left\{ - \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right)^2 \frac{1}{a^*} \frac{da^*}{dV_{GS}} + \frac{(V_{GS} - V_T)}{a^*} \right. \right. \\
 \left. \left. - \frac{1}{2} \left[\frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a^*} \right] \frac{1}{a^*} \frac{da^*}{dV_{GS}} + \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right)^2 \frac{1}{\bar{y}} \frac{d\bar{y}}{dV_{GS}} \right\} \right. \\
 \left. \div \left[\left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right)^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a^*} \right]^{1/2} \right. \\
 \left. + \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right) \frac{1}{a^*} \frac{da^*}{dV_{GS}} - \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right) \frac{1}{\bar{y}} \frac{d\bar{y}}{dV_{GS}} \right). \quad (40)
 \end{aligned}$$

Similarly, the drain conductance is

$$\begin{aligned}
 g_d = \frac{dI}{dV_{DS}} = Wv_s C_o \left(\left\{ - \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right)^2 \frac{1}{a^*} \frac{da^*}{dV_{DS}} \right. \right. \\
 \left. \left. - \frac{1}{2} \left[\frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a^*} \right] \frac{1}{a^*} \frac{da^*}{dV_{DS}} - \frac{(V_{GS} - V_T)}{a^*} \frac{dV_T}{dV_{DS}} \right. \right. \\
 \left. \left. + \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right)^2 \frac{1}{\bar{y}} \frac{d\bar{y}}{dV_{DS}} \right\} \div \left[\left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right)^2 + \frac{(V_{GS} - V_T)^2 - (\bar{Q}_0/C_o)^2}{a^*} \right]^{1/2} \right. \\
 \left. + \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right) \frac{1}{a^*} \frac{da^*}{dV_{DS}} - \left(\frac{\mathcal{L}_c \bar{y}}{a^*} \right) \frac{1}{\bar{y}} \frac{d\bar{y}}{dV_{DS}} \right). \quad (41)
 \end{aligned}$$

We could also write these equations with a^* everywhere replaced by a and \bar{Q}_0/C_o everywhere replaced by \bar{Q}/C_o . In that case, in the numerator of the square root terms we require the additional terms

$$- \frac{2}{a} \bar{Q}/C_o \frac{d}{dV_{GS}} \left(\frac{\bar{Q}}{C_o} \right) \quad \text{and} \quad - \frac{2}{a} \bar{Q}/C_o \frac{d}{dV_{DS}} \left(\frac{\bar{Q}}{C_o} \right)$$

in g_m and g_d , respectively.

In these equations, \bar{y} is a function of both the drain and gate voltages. The parameter a^* is related to a by (20c), and in the pinch-off condition [using (22) and (11b) in 20(c)] a^* is

$$a^* = 1 + \ln \left[\frac{V_{GS} - V_T}{\frac{\bar{Q}_0}{C_o} + \frac{I}{Wv_s C_o}} \right]. \quad (42)$$

Now the value of a and its variations are unimportant for low values of V_{GS} because the conventional device equations (i.e., without velocity saturation) apply and the forms for g_m and g_d are well known. On the other hand, for operation in velocity saturation we have shown that [cf. (32)] I' is a linear function of gate voltage and so (42) will be

relatively independent of gate and drain voltage. We can, therefore, neglect the a^* terms in (40) and (41) and use the value of a^* that we obtain at the pinch-off point to obtain a fairly good approximation. However, in comparing theory with experiment for accurate results, as we shall see, we must keep the dependence of I in (42) and calculate the current by one or two iterations. Some interesting features of (40) and (41) are worth noting. In the limit of large gate voltages (i.e., velocity saturation) so that $V_{GS} - V_T \gg \mathcal{E}_c \bar{y}$, (40) predicts that

$$g_m = \frac{C_o W v_s}{\sqrt{a^*}}, \quad (43)$$

and (41) predicts that

$$g_d = C_o W v_s \cdot \left[\frac{(\mathcal{E}_c \bar{y})^2}{(V_{GS} - V_T)} \frac{1}{a^{*3/2}} \frac{1}{\bar{y}} \frac{d\bar{y}}{dV_{DS}} - \frac{1}{2} \frac{(V_G - V_T)}{a^{*3/2}} \frac{da^*}{dV_{DS}} - \frac{1}{a^{*1/2}} \frac{dV_T}{dV_{DS}} \right]. \quad (44)$$

2.3.4.2 Triode region. In the triode region the corresponding expressions for the gate transconductance and the drain conductance are

$$g_m = W v_s C_o \left\{ \frac{\left(\frac{\mathcal{E}_c L}{a} \right)^2 \frac{\dot{a}_G}{a} + \frac{(V_{GS} - V_T) - (V_{GS} - V_T - V_{DS})}{a} - \frac{[(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2]}{a^2} \dot{a}_G}{\sqrt{\left(\frac{\mathcal{E}_c L}{a} \right)^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a}}} + \frac{\mathcal{E}_c L}{a^2} \dot{a}_G \right\} \quad (45)$$

and

$$g_d = W v_s C_o \left\{ \frac{\left(\frac{\mathcal{E}_c L}{a} \right)^2 \frac{\dot{a}_D}{a} + \frac{V_{GS} - V_T - V_{DS}}{a} - \frac{[(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2] \dot{a}_D}{a^2} - \frac{V_{DS}}{a} \frac{dV_T}{dV_{DS}}}{\sqrt{\left(\frac{\mathcal{E}_c L}{a} \right)^2 + \frac{(V_{GS} - V_T)^2 - (V_{GS} - V_T - V_{DS})^2}{a}}} + \frac{\mathcal{E}_c L}{a^2} \dot{a}_D \right\} \quad (46)$$

In this case the terms in \dot{a} are necessary to obtain a reasonable solution and so from (33a) we have

$$\frac{da}{dV_{DS}} = \dot{a}_D = \frac{1.24}{V_{GS} - V_T - V_{DS}},$$

$$\frac{da}{dV_{GS}} = \dot{a}_G = \frac{1.24V_{DS}}{(V_{GS} - V_T)(V_{GS} - V_T - V_{DS})}. \quad (47)$$

Again, as in the case of the saturation region, the formulas revert to their conventional forms for small values of gate and drain voltages. The main interest in considering the incremental conductance parameters is to show that continuity exists in making the transition from triode to saturation regions. We would like to show equivalence between (45) and (40) and between (46) and (41) at pinch-off conditions. By comparing (46) and (41) (written in terms of a and \bar{Q}/C_o) we find

$$\frac{(V_{GS} - V_T - V_{DS})}{a} = -\frac{\bar{Q}}{aC_o} \frac{d}{dV_{DS}} \left(\frac{\bar{Q}}{C_o} \right) - \frac{d\bar{y}}{dV_{DS}} \frac{1}{a} \left(\frac{I}{Wv_s C_o} \right)$$

or

$$\bar{Q}/C_o \left[1 + \frac{d}{dV_{DS}} \left(\bar{Q}/C_o \right) \right] = -\frac{d\bar{y}}{dV_{DS}} \left(\frac{I}{Wv_s C_o} \right). \quad (48)$$

Since the derivative on the left-hand side of (48) $\ll 1$, then (48) can be written as (173). As we show in Appendix A from (174) through (177) the derivative continuity equation (48) supplies the other condition needed together with (177) that allows determination of the parameters A_2 and h .

To illustrate these results, the drain conductance has been plotted in Fig. 13 with drain voltage as a parameter. The accuracy is quite good and continuity of the derivative is preserved in moving from the triode to the saturation regions.

2.3.5 Inclusion of the V_{DS} dependence of V_T

2.3.5.1 Triode region. In the triode region the charge-sharing formulation¹⁶ is used in which V_T is expressed

$$V_T = V_{FB} + 2\phi_F + \frac{2\sqrt{2\epsilon q N_A}}{3C_o V_{DS}} \cdot [(V_{DS} + V_{BS} + 2\phi_F)^{3/2} - (V_{BS} + 2\phi_F)^{3/2}] F_{ON},$$

where F_{ON} is the charge-sharing factor in the triode region. Although this result may be written approximately by expanding (for small V_{DS})

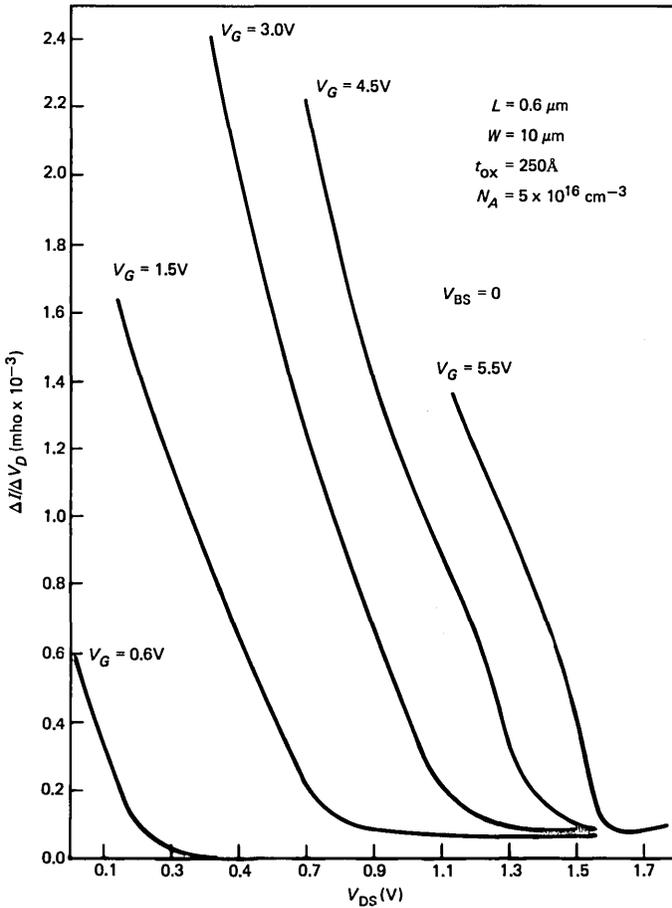


Fig. 13—Comparison of approximate and rigorous forms of the drain conductance.

as

$$V_T = V_{FB} + 2\phi_F + \frac{1}{C_o} \sqrt{2\epsilon q N_A \left(V_{BS} + \frac{V_{DS}}{2} + 2\phi_F \right) F_{ON}},$$

it is important to note that the complete form must always be used for circuit simulation purposes where discontinuities may not be allowed between the triode and saturation regions. The complete form provides a smooth transition for all voltages, whereas the approximate form introduces small glitches for some values.

2.3.5.2 Saturation region. In saturation operation the complete and approximate equations for V_T become

$$V_T = V_{FB} + 2\phi_F + \frac{2}{3} \frac{\sqrt{2\epsilon q N_A}}{C_o V_{SAT}} \cdot [(V_{SAT} + V_{BS} + 2\phi_F)^{3/2} - (V_{BS} + 2\phi_F)^{3/2}]_{SAT} \quad (49)$$

and

$$V_T = V_{FB} + 2\phi_F + \frac{1}{C_o} \sqrt{2\epsilon q N_A (V_{SAT} + V_{BS} + 2\phi_F) F_{SAT}}.$$

As in the triode region the complete form (49) must be used to avoid glitches at the transition regions. This result was developed previously.¹⁶ The difference in this case is that V_{SAT} also depends upon the charge at pinch-off and thus the current through (22). Therefore (22), (19b), and (49) must be solved together to determine the current. To obtain results that are smooth through all transitions, iteration is required.

2.3.6 Effective value of N_A

Almost all practical devices have ion-implanted channels and hence nonuniform doping profiles. Since the short-channel formulation describes the device in terms of averages over the source-drain distance, a reasonable approach is to use an average value of $N_A = \bar{N}_A$ over the same distance. We can calculate this value by conserving charge in the manner

$$\bar{N}_A \bar{x}_d = \int_0^{\bar{x}_d} N_A(x) dx, \quad (50)$$

where \bar{x}_d is the average depletion width between the drain and the source and has been shown to be¹⁶

$$\bar{x}_d \approx \sqrt{\frac{2\epsilon}{q\bar{N}_A}} \sqrt{V_{BS} + 2\phi_F + \frac{V_{DS}}{2}}. \quad (51)$$

We can see from this result that the effective value of \bar{N}_A is in general a function of both drain and substrate bias.

To determine \bar{N}_A we will assume that any ion-implanted profile may be described by the parameters N_p , σ , and R_p ; i.e., even after thermal processing it is assumed that N_p , σ , and an R_p may be found to give a best fit to the measured impurity profile so that

$$N_A(x) = (N_p - N_{AB}) \exp \left[-\left(\frac{x - R_p}{2\sigma} \right)^2 \right] + N_{AB}, \quad (52)$$

where N_{AB} is the background impurity doping. Using (52) and (51) in (50) we then have

$$\sqrt{\bar{N}_A} - \frac{N_{AB}}{\sqrt{\bar{N}_A}} = \left[\frac{q}{2\epsilon \left(V_{BS} + \frac{V_{DS}}{2} + 2\phi_F \right)} \right]^{1/2} \sqrt{\frac{\pi}{2}} N_p \sigma \cdot \left(\operatorname{erf} \left(\frac{R_p}{\sqrt{2}\sigma} \right) + \operatorname{erf} \left\{ \frac{\left[\frac{2\epsilon}{q\bar{N}_A} \left(V_{BS} + 2\phi_F + \frac{V_{DS}}{2} \right) - R_p \right]^{1/2}}{\sqrt{2}\sigma} \right\} \right) \quad (53)$$

for the triode region and

$$\sqrt{\bar{N}_A} - \frac{N_{AB}}{\sqrt{\bar{N}_A}} = \left[\frac{q}{2\epsilon \left(V_{BS} + \frac{V_{SAT}}{2} + 2\phi_F \right)} \right]^{1/2} \sqrt{\frac{\pi}{2}} N_p \sigma \cdot \left(\operatorname{erf} \left(\frac{R_p}{\sqrt{2}\sigma} \right) + \operatorname{erf} \left\{ \frac{\left[\frac{2\epsilon}{q\bar{N}_A} \left(V_{BS} + 2\phi_F + \frac{V_{SAT}}{2} - R_p \right) \right]^{1/2}}{\sqrt{2}\sigma} \right\} \right)$$

for the saturation region. If we use a polynomial expansion for the erf term to allow numerical computation, only a few iterations are required to obtain \bar{N}_A . For the most accurate results, then, a value of $\bar{N}_A(V_{DS}, V_{BS})$ is required at each bias point, although in practice only the V_{BS} dependence is very important, so an average value of V_{DS} could be used for all computation.

III. COMPARISON WITH EXPERIMENT

To establish the validity of the parameter $B = 2$, the \sqrt{I} versus V_{GS} data were plotted in Fig. 14 with drain voltage $V_{DS} = V_{GS}$ for the long- and short-channel devices discussed in Fig. 11. Using all of the parameters determined from the linear region data, theoretical curves were generated and are also plotted. The drain-voltage dependence of V_T was included using the charge-sharing technique,¹⁶ and the variation of \bar{y} was taken from the drain-voltage data in order to verify the triode region model alone. It is evident that the agreement is very good. To compare with the other models, the result of using $B = 1$, and the piecewise continuous model (i.e., constant mobility) are also shown. Both are in error, one underestimating and the other overestimating the actual current.

One can identify two sections in this curve. At low voltages the \sqrt{I} curve is linear in V_{GS} , indicating long-channel behavior; for higher gate voltages the \sqrt{I} curve is sublinear in I , indicating the onset of the effects of velocity saturation. The change from one behavior to the

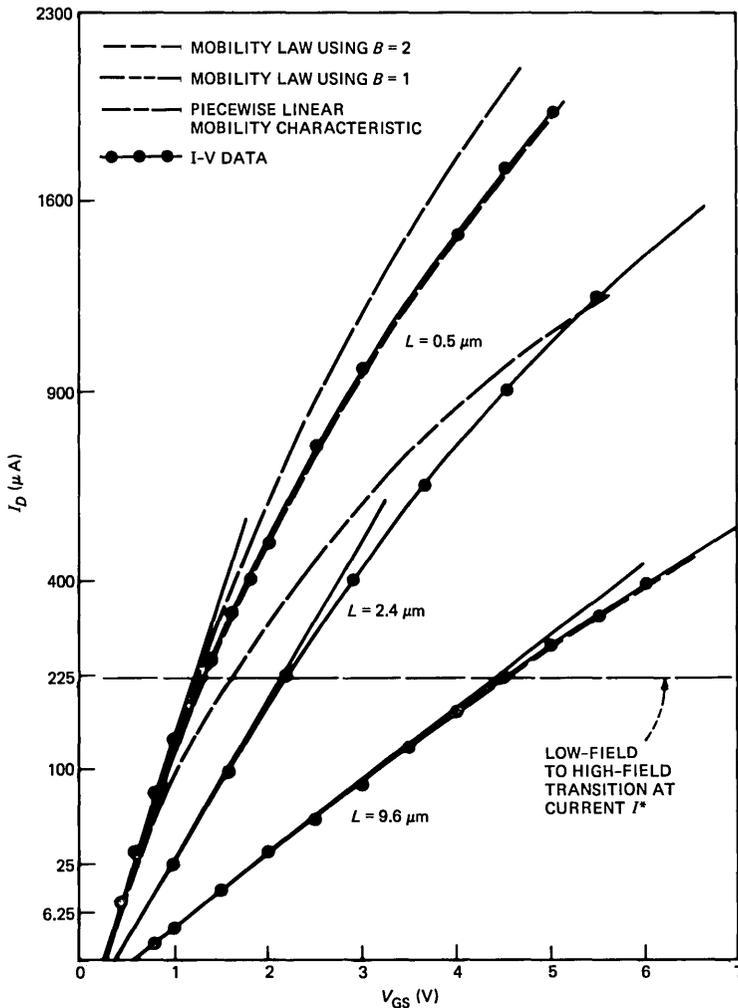


Fig. 14—Saturation region data for long- and short-channel devices plotted on a \sqrt{I} scale for the condition $V_{DS} = V_{GS}$.

other is gradual and takes place when the terms under the root sign in (20b) are approximately equal; i.e., at a current level of about

$$I = v_s WC_o \frac{(V_{GS} - V_T)}{\sqrt{a^*}}.$$

This result should be independent of gate length, as we can see from (17c), which predicts the transition from $\mathcal{E} < \mathcal{E}_c$ to $\mathcal{E} > \mathcal{E}_c$ to occur for a unique value of I , which is

$$I^* = \frac{v_s W \bar{Q}_0}{(1 - \bar{R})}$$

This behavior is observed in Fig. 14, although there is some decrease for the very shortest gate length.

As a final demonstration of the model the drain current data for two devices are shown in Figs. 15a and b. The channel length modulation term can be calculated using (169). However, this description is only valid for large gate voltages, i.e., when \mathcal{E} is $\geq 10^4$ V/cm. As we approach the threshold condition, the derivation breaks down and $L - \bar{y}$ becomes anomalously large. However, near the threshold condition and in the subthreshold region, we know from other work that the simple form of

$$L - \bar{y} = \sqrt{\frac{2\epsilon}{qN_A} (V_{DS} - V_{SAT})}$$

works well. We will, therefore, combine these two results to obtain a continuous solution

$$L - \bar{y} = \frac{A_1 (V_{DS} - V_{SAT})^{3/4}}{\mathcal{E} + A_1 (V_{DS} - \bar{V})^{1/4} \sqrt{\frac{qN_A}{2\epsilon}}}$$

In Fig. 15a the fabrication parameters are $L = 0.32 \mu\text{m}$ (1- μm coded length and 0.68- μm compensation as determined from $1/g_m$ measurements), $W = 30 \mu\text{m}$, $t_{ox} = 230 \text{\AA}$, $r_j = 0.26 \mu\text{m}$, and $N_{AB} = 8 \times 10^{15} \text{cm}^{-3}$. The implanted doping parameters were determined by simulation to be $N_p = 5 \times 10^{16} \text{cm}^{-3}$, $R_p = 0.5 \mu\text{m}$, and $\sigma = 0.1 \mu\text{m}$. Other parameters determined from low voltage data are $R = 15$ through 20Ω , $\mu_0 = 650 \text{cm}^2/\text{V}\cdot\text{s}$ and $\theta = 0.02 \text{V}^{-1}$. The agreement between theory and experiment is very good for $V_{GS} > 1 \text{V}$. For smaller gate voltages the characteristic is dominated by uncontrolled source-drain punchthrough, which is not described by these models.

In Fig. 15b the parameters are the same but $L = 0.85 \mu\text{m}$, so in this case the comparison is extended down to the threshold region (because source-drain punchthrough is not a limiting factor).

IV. SUMMARY

A model has been developed to describe the velocity-saturated characteristics of short-channel MOSFETs. The model has been based upon the velocity-field relation that most nearly fits the experimental data, and it is found that the bulk relationship based upon optical phonon emission is most appropriate if the zero field (parallel to the

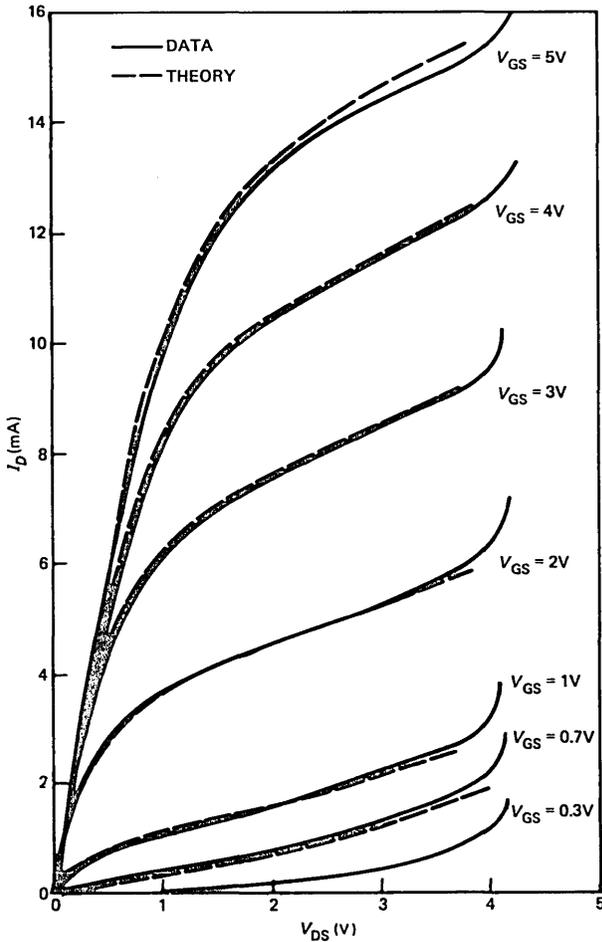


Fig. 15(a)—Drain-voltage versus drain-current data for a short-channel MOSFET showing comparison with theory. $L = 0.32 \mu\text{m}$.

surface) mobility is simply modified to take into account the effects of increased scattering in the potential well at the semiconductor-insulator interface. By using this approach, the electron temperature has been brought into the problem and becomes the necessary ingredient that allows a smooth transition from the long-channel to the short-channel behavior. The model is found to fit well with experimental data.

V. ACKNOWLEDGMENTS

The author would like to acknowledge helpful discussions with W. Fichtner, S. M. Sze, and G. E. Smith.

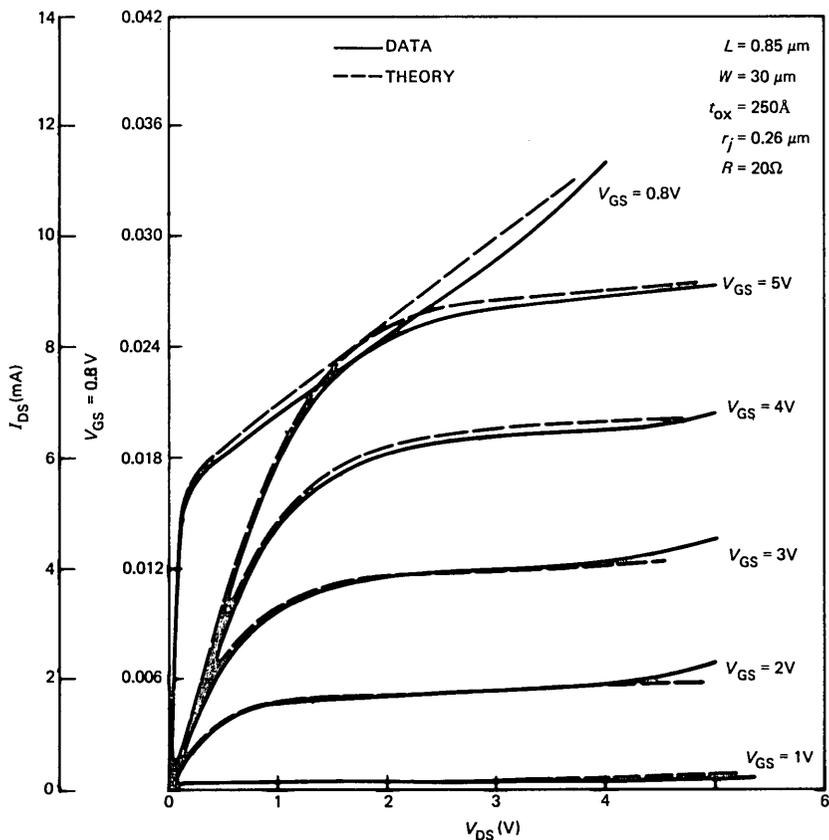


Fig. 15(b)—Drain-current versus drain-voltage data for a short-channel MOSFET showing comparison with theory. $L = 0.85 \mu\text{m}$.

REFERENCES

1. K. Yamaguchi, "Field-Dependent Mobility Model for Two-Dimensional Numerical Analysis of MOSFETs," *IEEE Trans. Electron. Dev.*, *ED-26* (July 1979), pp. 1068-78.
2. T. Toyabe et al., "A Numerical Model of Avalanche Breakdown in MOSFETs," *IEEE Trans. Electron. Dev.*, *ED-25* (July 1978), pp. 825-32.
3. H. Masuda, M. Nakai, and M. Kubo, "Characteristics and Limitations of Scaled Down MOSFETs Due to Two-Dimensional Field Effect," *IEEE Trans. Electron. Dev.*, *ED-26* (June 1979), pp. 980-6.
4. R. Reich and D. K. Ferry, "A Two-Dimensional Simulation of a Cooled Submicrometer Indium Arsenide Schottky-Gate FET," *IEEE Trans. Electron. Dev.*, *ED-27* (June 1980), pp. 1062-5.
5. B. Hoefflinger et al., "Model and Performance of Hot-Electron MOS Transistors for VLSI," *IEEE J. Solid State Circuits*, *SC-14* (April 1979), pp. 435-42.
6. T. J. Rodgers et al., "An Experimental and Theoretical Analysis of Double-Diffused MOS Transistors," *IEEE J. Solid State Circuits*, *SC-10* (October 1975), pp. 322-31.
7. F. N. Trofimenkoff, "Field-Dependent Mobility Analysis of the Field-Effect Transistor," *Proc. IEEE*, *53* (November 1965), pp. 1765-6.

8. B. Hoeneissen and C. A. Mead, "Current Voltage Characteristics of Small Size MOS Transistors," *IEEE Trans. Electron. Dev.*, *ED-19* (March 1972), pp. 382-3.
9. D. Roychoudhury and P. K. Basu, "A New Mobility Field Expression for the Calculation of MOSFET Characteristics," *Solid State Electron.*, *19* (July 1976), p. 656.
10. R. J. Jardonek and W. R. Bandy, "Velocity Saturation Effects in n-Channel Deep-Depletion SOS/MOSFETS," *IEEE Trans. Electron. Dev.*, *ED-25* (August 1978), pp. 894-8.
11. G. A. Armstrong and J. A. Magowan, "The Distribution of Mobile Carriers in the Pinch-Off Region of an Insulated-Gate Field-Effect Transistor and Its Influence on Device Breakdown," *Solid State Electron.*, *14* (August 1971), pp. 723-35.
12. G. Merckel, J. Borel, and N. Z. Cupcea, "An Accurate Large Signal MOS Transistor Model for Use in Computer-Aided Design," *IEEE Trans. Electron. Dev.*, *ED-19* (May 1972), pp. 681-90.
13. D. M. Caughey and R. F. Thomas, "Carrier Mobilities in Si Empirically Related to Doping and Field," *Proc. IEEE*, *55* (December 1967), pp. 2192-3.
14. E. M. Conwell, *High Field Transport in Semiconductors*, New York: Academic Press, 1967.
15. R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers," *Phys. Rev.*, *126* (June 1962), pp. 2002-14.
16. G. W. Taylor, "The Effects of Two-Dimensional Charge Sharing on the Above-Threshold Characteristics of Short-Channel IGFETs," *Solid State Electron.*, *22* (August 1978), pp. 701-18.
17. D. P. Kennedy and P. C. Murley, "Steady State Mathematical Theory for the Insulated Gate Field Effect Transistor," *IBM J. Res. Dev.*, *17*, No. 1 (January 1973), pp. 2-12.
18. S. C. Sun and J. D. Plummer, "Electron Mobility in Inversion and Accumulation Layers on Thermally Oxidized Silicon Surfaces," *IEEE Trans. Electron. Dev.*, *ED-27* (August 1980), pp. 1497-508.
19. J. P. Nougier, "Identity Between Spreading and Noise Diffusion Coefficients for Hot Carriers in Semiconductors," *Appl. Phys. Lett.* *32*, No. 10 (October 1978), pp. 671-3.
20. P. P. Guebels and F. Van de Wiele, *IEDM Tech. Digest* (December 1981), p. 211.
21. K. K. Thornber, "Relation of Drift Velocity to Low-Field Mobility and High-Field Saturation Velocity," *J. Appl. Phys.*, *51* (April 1980), pp. 2127-36.
22. H. K. J. Ihanntola and J. L. Moll, "Design Theory of a Surface Field Effect Transistor," *Solid State Electron.*, *7* (July 1964), pp. 423-30.
23. H. C. Pao and C. T. Sah, "Effect of Diffusion Current on Characteristics of Metal-Oxide (Insulator)-Semiconductor Transistor," *Solid State Electron.*, *9* (September 1966), pp. 927-37.
24. J. A. Magowan and G. A. Armstrong, "Modelling of Short-Channel M.O.S. Transistors," *Electron. Lett.*, *6*, No. 10 (May 1970), pp. 313-5.
25. G. W. Taylor, "Subthreshold Conduction in MOSFETs," *IEEE Trans. Electron. Dev.*, *ED-25* (March 1978), pp. 337-50.

APPENDIX A

A.1 Introduction

In existing analytical approaches to pinch-off operation, the field at the edge of the pinch-off region is held to the value \mathcal{E}_c , the critical field parameter in the velocity-field relationship.^{6,12,21} To avoid such an arbitrary condition, a description is presented here of the pinch-off region under conditions of hot-electron transport, which is based upon the boundary condition of a value of the electric field at the pinch-off point determined uniquely by the current and hence applied voltages. The field patterns and charge density distribution throughout the pinch-off zone are predicted.

The solution allows one to determine the effect of the gate voltage on the channel field at pinch-off through the channel current itself.

Hence, there is direct feedback between the channel current and the extent of channel length modulation. It is demonstrated that for low gate voltages the channel length modulation may be severe and the current then shows a large variation with drain voltage. However, for larger gate voltages, electrostatic feedback to the channel becomes a dominant effect, and the current shows little variation with drain voltage, i.e., the degree of channel length modulation becomes very small.

A.2 General considerations and assumptions

It is helpful initially to review the important approximations and salient results that are obtained from simplified long-channel theory. For the purposes of the discussion, Fig. 16 shows a cross section of the device, which indicates the major current flow patterns and electric-field lines.

A.2.1 Existing theories

A basic assumption of MOS theory is that the gradual channel approximation (GCA) is valid throughout the unsaturated region. In

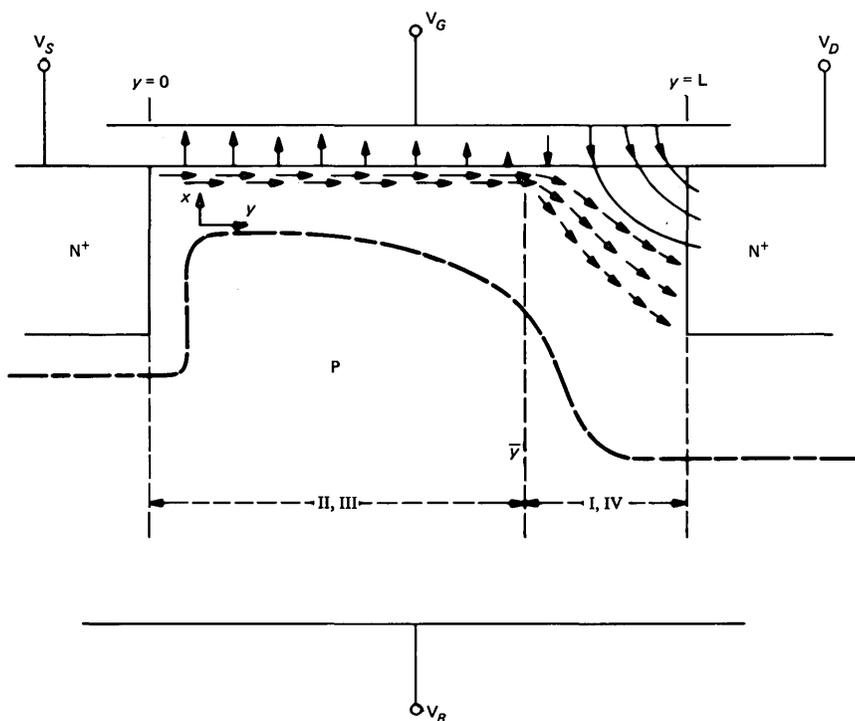


Fig. 16—MOS device cross section showing schematically the major lines of current flow (dashed arrows) and electric-field patterns (solid arrows).

a physical sense this approximation assumes that the transverse electric-field strength, $d\phi/dx$, is large compared to the longitudinal electric-field strength, $d\phi/dy$, in the channel region (ϕ is the electrostatic potential, x is the direction perpendicular to the gate, and y is the direction parallel to the channel). Hence, the charge $Q(y)$ in the channel may be determined by

$$Q(y) = C_o [V_G - V_{FB} - 2\phi_F - V(y)] - Q_B(y), \quad (54)$$

where C_o is the oxide capacitance, V_{FB} is the flatband voltage, $V(y)$ is the channel potential, and $Q_B(y)$ is the substrate depletion charge. The substrate charge is determined, using the depletion approximation, to be

$$Q_B(y) = \sqrt{2\epsilon q N_A [V(y) + 2\phi_F + V_{BS}]}, \quad (55)$$

where N_A is the substrate doping, q is the electronic charge, ϵ is the silicon dielectric permittivity, and

$$\phi_F = \frac{kT}{q} \ln \left(\frac{N_A}{n_i} \right). \quad (56)$$

The bulk Fermi potential is written in terms of n_i , the intrinsic silicon concentration. Another important assumption of the theory is that drift is the dominant conduction mechanism so that the continuity equation may be written

$$I = W\mu Q(y) \mathcal{E}(y), \quad (57)$$

where I is the source-to-drain current, W is the device width, μ is the channel mobility, and $\mathcal{E}(y)$ is the channel field. The channel current is obtained from (55), (56), and (57) to yield the familiar equation

$$I = \frac{\mu C_o W}{L} \left[\left(V_{GS} - V_{FB} - 2\phi_F - \frac{V_{DS}}{2} \right) V_{DS} - \frac{2\sqrt{2\epsilon q N_A}}{3C_o} \left\{ (V_{DS} + V_{BS} + 2\phi_F)^{3/2} + (V_{BS} - 2\phi_F)^{3/2} \right\} \right]. \quad (58)$$

The phenomena of current saturation and pinch-off in the channel are predicted to occur when the free charge in the channel goes to zero. From a solution of (54) for

$$Q(y) = 0 \quad (59)$$

one obtains

$$V_{SAT} = V_G - V_{FB} - 2\phi_F + \frac{\epsilon q N_A}{C_o^2} \left[1 - \sqrt{1 + \frac{2C_o^2 (V_G - V_{FB} + V_{BS})}{\epsilon q N_A}} \right]. \quad (60a)$$

The corresponding value of the saturated current is then found by substitution of (60a) into (58), and in the simple case, this procedure yields

$$I_{\text{SAT}} = \frac{\mu C_o W}{2L} (V_G - V_T)^2, \quad (60b)$$

where V_T is the simplified threshold value. In the general case, the more complex drain-voltage-dependent form of V_T ¹⁶ must be included in (60b), but the result is still evaluated in a straightforward manner.

A.2.2 Device characteristics in the saturation region

For voltages $V_{\text{DS}} > V_{\text{DSAT}}$, the drain depletion region extends towards the source. The position of pinch-off in the channel, occurring initially at the drain for $V(y) = V_{\text{DSAT}}$, will also move towards the source, since the condition (59) of zero charge may always be found somewhere in the channel for a function like (54), which is decreasing with voltage. Because of this feature, the potential at this position will remain constant for increasing drain voltage.

The use of (59) in the channel in saturation is, of course, an approximation that is used solely to determine the drain saturation voltage. Actually, the electron density is decreasing rapidly with distance in this region, and if we allow it to become arbitrarily small, then the electric field will have to become anomalously large if the drift component is to continue to provide continuity of current. We must therefore conclude that the drift component can no longer account for the total current flow and hence that the diffusion of carriers becomes an important conduction mechanism. This conclusion is supported by the numerical computations of other authors.^{11,17} We show here that diffusion constitutes a specific fraction \mathcal{P} of the current flow at the pinch-off point, and in Appendix B values for \mathcal{P} are derived in the range of one third to one half.

As a basis for this work, we assume that we may represent the MOS device by four different regions of operation, as shown in Fig. 17. The quadrants are divided on the vertical axis by the subthreshold and above-threshold regions of operation and on the horizontal axis by the position of pinch-off in the channel. The quadrants are distinguished by the approximations designated in the figure, which denote for each quadrant the dominant conduction mechanism and the principle direction of the electric-field lines. In quadrants II and III the GCA applies, which may be stated

$$\frac{d^2\phi}{dx^2} \gg \frac{d^2\phi}{dy^2}. \quad (61)$$

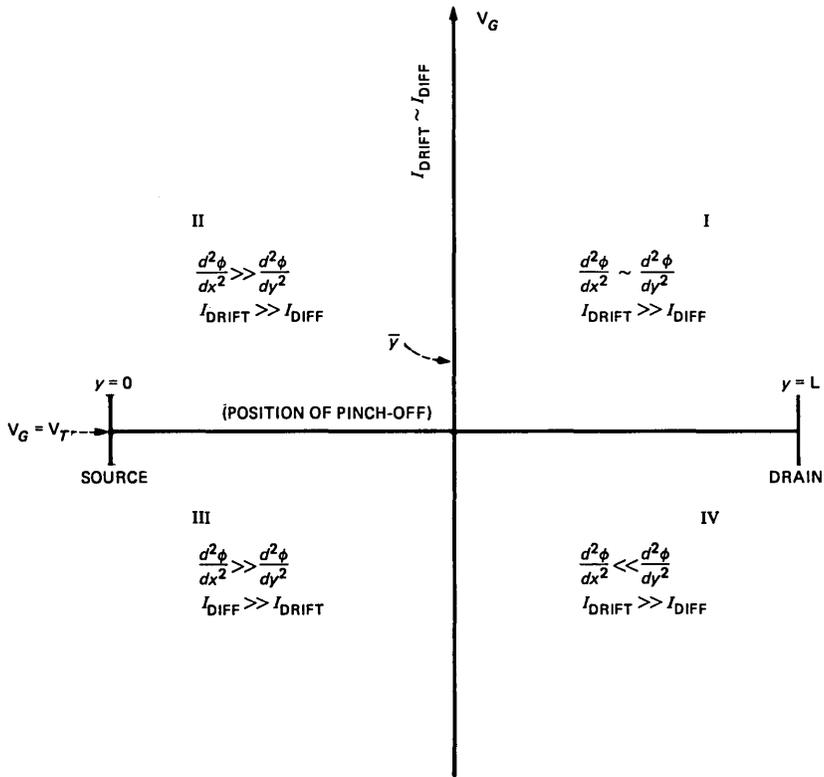


Fig. 17—Regions of operation for a MOSFET characterized by the approximations that may be used.

However, in quadrant I neither of these approximations is valid since the two terms are of the same order, which we write

$$d^2\phi/dx^2 \sim d^2\phi/dy^2. \quad (62)$$

The validity of (16) has been well established in MOS theory; the validity of (62) will be justified here by the solutions obtained. The other approximations indicated in Fig. 17 are that in quadrants II and IV we have

$$I_{DRIFT} \gg I_{DIFF} \quad (63a)$$

and in quadrant III we have

$$I_{DRIFT} \ll I_{DIFF}. \quad (63b)$$

In quadrant I we also have $I_{DRIFT} > I_{DIFF}$, but near the boundary we have

$$I_{\text{DRIFT}} \sim I_{\text{DIFF}}, \quad (63c)$$

which means that the components are comparable in size. The assumptions in quadrants II, III, and IV have been established by others; the validity of the assumptions in quadrant I and on its boundary will be considered here.

In the discussion of hot-electron transport we will use the mobility-field relationship²²

$$\mu(\mathcal{E}) = \frac{\mu_0}{\left[1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c}\right)^2\right]^{1/2}} \equiv \mu, \quad (64)$$

where \mathcal{E} is the electric field and \mathcal{E}_c is a critical field parameter that is related to the low-field mobility μ_0 by the relation

$$\mu_0 \mathcal{E}_c = v_s. \quad (65)$$

Also, the Einstein relationship will be assumed to hold; it states

$$D(\mathcal{E})/\mu(\mathcal{E}) = \frac{kT(\mathcal{E})}{q} \equiv \frac{kT}{q}, \quad (66)$$

where $D \equiv D(\mathcal{E})$ is the field-dependent diffusion coefficient and $T \equiv T(\mathcal{E})$ is the temperature of the electron in the hot-electron regime.

We will now discuss the conditions in regions I and II.

A.2.2.1 Region II. The current flow is approximately one-dimensional (parallel to the surface) because the dominant transverse field confines the carriers to a narrow potential well next to the surface. This region is described by (54), (55), and (57). Although drift is dominant in this region, it is also of interest to calculate the diffusion component. From (54) the mobile charge gradient is

$$\frac{dQ}{dy} = - (C_o + C_s) \frac{dV}{dy}, \quad (67)$$

where C_s , the space-charge capacitance, is

$$C_s \equiv C_s(y) = \sqrt{\frac{qN_A\epsilon}{2(V + 2\phi_F + V_{\text{BS}})}} \quad (68)$$

and $V \equiv V(y)$. The electric field is obtained from (57) as

$$\mathcal{E}(y) = \frac{I(1 - R)}{W\mu Q(y)}, \quad (69)$$

where R represents the fraction of the total current carried by diffusion. By using (69), (67), and (54), the diffusion component is

$$RI = I_{\text{DIFF}}$$

$$= \frac{D(\mathcal{L})I(C_o + C_s)(1 - R)}{\mu(\mathcal{L})[C_o(V_G - V_{\text{FB}} - 2\phi_F - V) - \sqrt{qN_A}2\epsilon(V + 2\phi_F + V_{\text{BS}})]}. \quad (70)$$

A.2.2.2 Region I. As the diagram in Fig. 16 shows, the problem in this region becomes two-dimensional. In region II, the electric field is directed from the channel towards the gate but is steadily decreasing as the channel potential increases. Somewhere beyond \bar{y} , the pinch-off position, the field in the oxide is equal to zero, and beyond this position the electric field is directed from the gate into the silicon; hence, many of the field lines terminate on the drain electrode, as shown, since the drain potential exceeds the gate potential in this region of operation. Because the transverse field in region I no longer creates a potential well, the mobile carriers may flow away from the surface as they approach the drain, resulting in a two-dimensional current flow.

The main features of this representation are depicted in Fig. 18, which shows an expanded view of region I. Consider the flow of carriers at some position y in region I. Because of the strong electric field indicated in Fig. 18, which is directed from the gate towards the substrate in region I, the electrons are forced away from the surface. Therefore, the density of electrons will be reduced at the surface. For sufficiently large values of x (i.e., deeper in the Si) the electron density must decrease again to its substrate value, and so we conclude that the electron density must exhibit a maximum as a function of x for a given value of y . We will, therefore, represent the electron density in a general way by a function of the form

$$n(\omega, r) = N(r)\exp\{-A(r)[\omega - \omega_0^*(r)]^2 - B(r)[\omega - \omega_1^*(r)]^4 - \dots\}, \quad (71)$$

where the higher-order terms in the distribution would be required for strong deviations from the normal case. The expansion for the electron density has been written for the generalized coordinate axes ω, r rather than for x, y to allow for the fact that the x, y system may not be the most convenient one in which to describe the distribution. The y or r dependence has been incorporated into this form through the parameters $N(r), A(r), B(r), \omega_0^*(r),$ and $\omega_1^*(r)$. In this way, the median value, the mean position, and the width of the distribution may change with y or r . The functions $\omega^*(r)$ describe the locus of the mean of the electron distribution between the pinch-off point and the drain as illustrated in Fig. 18. In the approach taken here, we have only three physical relationships available to us for the determination of the electron distribution, so we will restrict (71) to three unknown functions: i.e.,

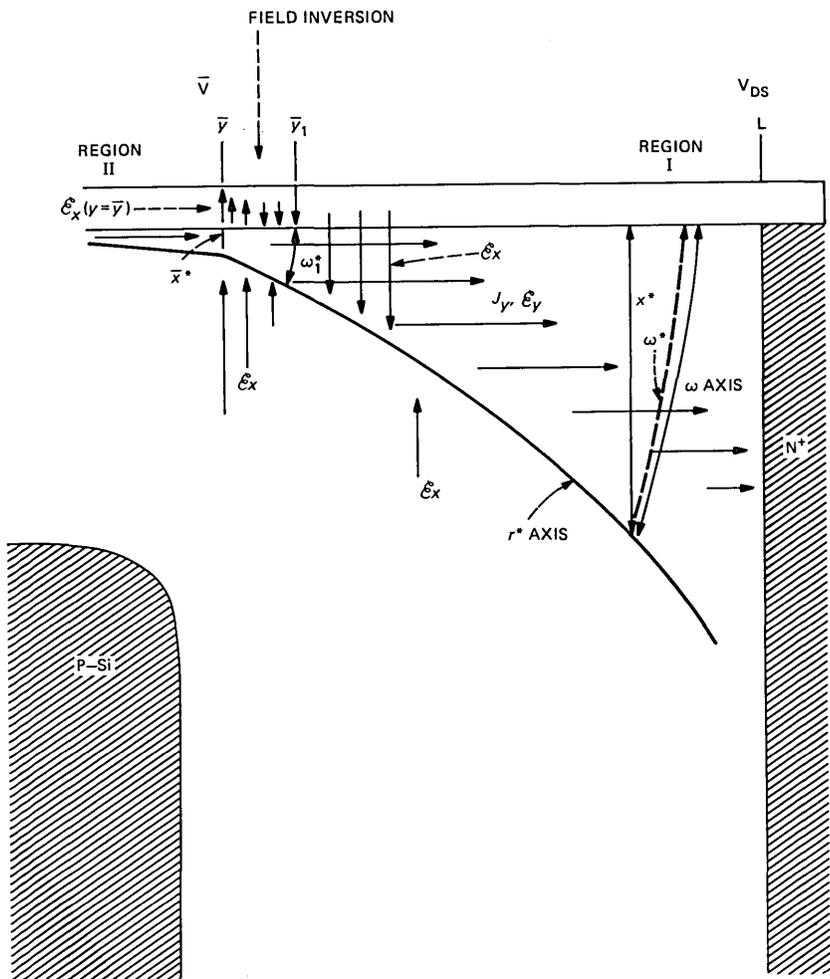


Fig. 18—Expanded view of the pinch-off zone with important physical and electrical parameters.

$$n \equiv n(\omega, r) = N(r) \exp \left\{ - \frac{[\omega - \omega^*(r)]^2}{2\sigma^2(r)} \right\}, \quad (72)$$

where $A(r)$ has been rewritten as $1/2\sigma^2(r)$ and in which the functions $N(r)$, $\sigma(r)$, $\omega^*(r)$ remain to be determined from three physical relationships, which will now be described.

The first relationship is the description of the current flow mechanism and quite generally it is

$$\frac{I}{qW} = \int_0^\infty D \frac{dn}{dr} d\omega + \int_0^\infty \mu \mathcal{E}_r n d\omega \quad (73)$$

for any y between \bar{y} and L where $\omega = 0$ is the silicon surface. The second relationship is the law of current continuity. Physically speaking, the current flow in the x direction must provide the current density gradient by means of which the current density in the y direction is able to change. This condition is stated formally as

$$\nabla \cdot \mathbf{J} = 0 \quad (74)$$

or, for the case being considered,

$$\frac{dJ_x}{dx} = -\frac{dJ_y}{dy},$$

where J_x and J_y are the current density components in the x and y directions, respectively. The third relationship is the two-dimensional Poisson equation

$$\nabla \cdot \mathcal{E} = \frac{q}{\epsilon} \rho, \quad (75)$$

which is written here as

$$\frac{\partial \mathcal{E}_x}{\partial x} + \frac{\partial \mathcal{E}_y}{\partial y} = -\frac{q}{\epsilon} (N_A + n), \quad (76)$$

subject to the boundary condition

$$\mathcal{E}_y = \bar{\mathcal{E}}, \quad y = \bar{y}, \quad (77)$$

where \bar{y} is the pinch-off point and $\bar{\mathcal{E}}$ is the value of longitudinal electric field in the channel at the pinch-off point. Both x and y components need to be retained in (76) since the concentration of field lines is comparable in both x and y directions. The value of \mathcal{E}_y along r^* will be discussed later.

In the section of the channel between the source and the field-inversion point, (64) is a one-dimensional relationship between the channel field and the mobility. Between the field inversion point and the drain, the mobility becomes a scalar field determined by the vector

$$\mathcal{E} = \mathbf{r} \mathcal{E}_r + \omega \mathcal{E}_\omega.$$

Since

$$\mathcal{E}^2 = \mathcal{E}_r^2 + \mathcal{E}_\omega^2,$$

(64) becomes

$$\mu = \frac{\mu_0}{\left[1 + \left(\frac{\mathcal{E}_\omega}{\mathcal{E}_c} \right)^2 + \left(\frac{\mathcal{E}_r}{\mathcal{E}_c} \right)^2 \right]^{1/2}}. \quad (78a)$$

Also, the one-dimensional hot-electron temperature

$$T_e = T_0 \left[1 + \left(\frac{\mathcal{E}}{\mathcal{E}_c} \right)^2 \right] \quad (78b)$$

may be expressed

$$\beta_e = \beta_0 + \beta_0 \left(\frac{\mathcal{E}_w}{\mathcal{E}_c} \right)^2 + \beta_0 \left(\frac{\mathcal{E}_r}{\mathcal{E}_c} \right)^2, \quad (78c)$$

which is a scalar field, where $\beta_e = kT_e/q$. Therefore, the mobility (78a) may be expressed in terms of the temperature T_e or, equivalently, the voltage β_e using (78a) and (78c) as

$$\frac{\mu}{\mu_0} = \sqrt{\frac{\beta_0}{\beta_e}}. \quad (78d)$$

A.3 Determination of physical parameters in pinch-off operation

A.3.1 Charge, voltage, and field at the pinch-off position

The boundary between regions II and I will be defined as the pinch-off position, and all variables at this point will be designated with a bar. As the position \bar{y} is approached from the source, the drift current will be a decreasing function of y and the diffusion current will be an increasing function of y . At \bar{y} , the drift and diffusion components of the current are

$$I_{\text{DRIFT}} = (1 - \bar{R})I \quad (79)$$

and

$$I_{\text{DIFF}} = \bar{R}I. \quad (80)$$

The method for the determination of \bar{R} is outlined in Appendix B [see (200)]. It is determined only by doping and oxide thickness and has typical values of one third to one half. From (70) and (66) we have

$$\begin{aligned} C_o(V_G - V_{\text{FB}} - 2\phi_F - \bar{V}) - \sqrt{qN_A 2\epsilon(\bar{V} + 2\phi_F + V_{\text{BS}})} \\ = \frac{1 - \bar{R}}{\bar{R}} \frac{kT}{q} [C_o + C_s(\bar{y})]. \end{aligned} \quad (81)$$

Equation (81) may also be written approximately as

$$V_{\text{SAT}} \equiv \bar{V} = V_G - V_T - \left(\frac{1 - \bar{R}}{\bar{R}} \right) \frac{kT}{q} \left(\frac{C_o + \bar{C}_s}{C_o} \right). \quad (82)$$

The relationship (81) is identical to that obtained in the original MOS theory²³ except for the extra term on the right-hand side. In other words, rather than using (59) to determine the pinch-off voltage, we are taking account of the charge in the channel at pinch-off, and from (81) and (54) the charge is determined to be

$$\bar{Q} = \left(\frac{1 - \bar{R}}{\bar{R}} \right) \frac{kT}{q} (C_o + \bar{C}_s). \quad (83)$$

For $T = T_0$ (the lattice temperature, as is the case in long-channel devices), \bar{Q} is small (provided that $\bar{R} \gtrsim 0.3$) and (82) yields the conventional result. The significance of (83)* is that we can now determine the longitudinal field in the channel at pinch-off, and from (83) and (69) it is

$$\bar{\mathcal{E}}_y = \frac{I\bar{R}}{W\mu \frac{kT}{q} (C_o + \bar{C}_s)}. \quad (84)$$

Equation (84) applies for any set of voltage variables in pinch-off operation. The dependence of \mathcal{E} on μ from (64) could be substituted here, and since the current at the onset of pinch-off operation (i.e., the boundary between triode and saturation regions) is known, then \mathcal{E} at the onset of pinch-off may also be expressed uniquely in terms of the applied voltages. It is therefore an ideal boundary condition for the pinch-off zone.

It is also of some interest to determine the sign and magnitude of \mathcal{E}_x at the pinch-off point. From the continuity of the divergence of \mathcal{E} we can relate the oxide and substrate fields as

$$\epsilon \mathcal{E}_{x_{si}} = \epsilon_{ox} \mathcal{E}_{ox}, \quad (85)$$

where $\mathcal{E}_{x_{si}}$ may be determined, as shown by Pao,²³ by a detailed solution of Poisson's equation in the x direction in which both mobile and fixed charge components are retained. An equivalent representation of \mathcal{E}_{ox} is

$$\mathcal{E}_{ox} = \frac{V_G - \phi_s - V(y) - V_{FB}}{t_{ox}}, \quad (86)$$

where ϕ_s is the surface potential at the source. As is well known, ϕ_s takes the value of $2\phi_F$ at threshold and increases only slightly for additional increases in gate voltage. Therefore, using (82) in (86) we find that

$$\bar{\mathcal{E}}_{ox} = \frac{\bar{Q} + \sqrt{2\epsilon q N_A (2\phi_F + V_{BS})} + V_{FB}/C_o}{\epsilon_{ox}} \quad (87a)$$

and

$$\bar{\mathcal{E}}_{x_{si}} = \frac{\bar{Q} + \sqrt{2\epsilon q N_A (2\phi_F + V_{BS})} + V_{FB}/C_o}{\epsilon}. \quad (87b)$$

* To include short-channel effects it is necessary to replace \bar{C}_s with $\bar{C}_s F$.

Hence, we conclude that both \mathcal{E}_{ox} and \mathcal{E}_{xi} start from large positive values at the source and decrease to the much smaller values given by (87a) and (87b) at the pinch-off position. These values are practically constant with increasing gate voltage except for the small increases in ϕ_S above $2\phi_F$. The fields are directed towards the gate electrode at \bar{y} . These results are illustrated by the plots of Pao,²³ which are reproduced in Fig. 19. We have extended the field values past the pinch-off point

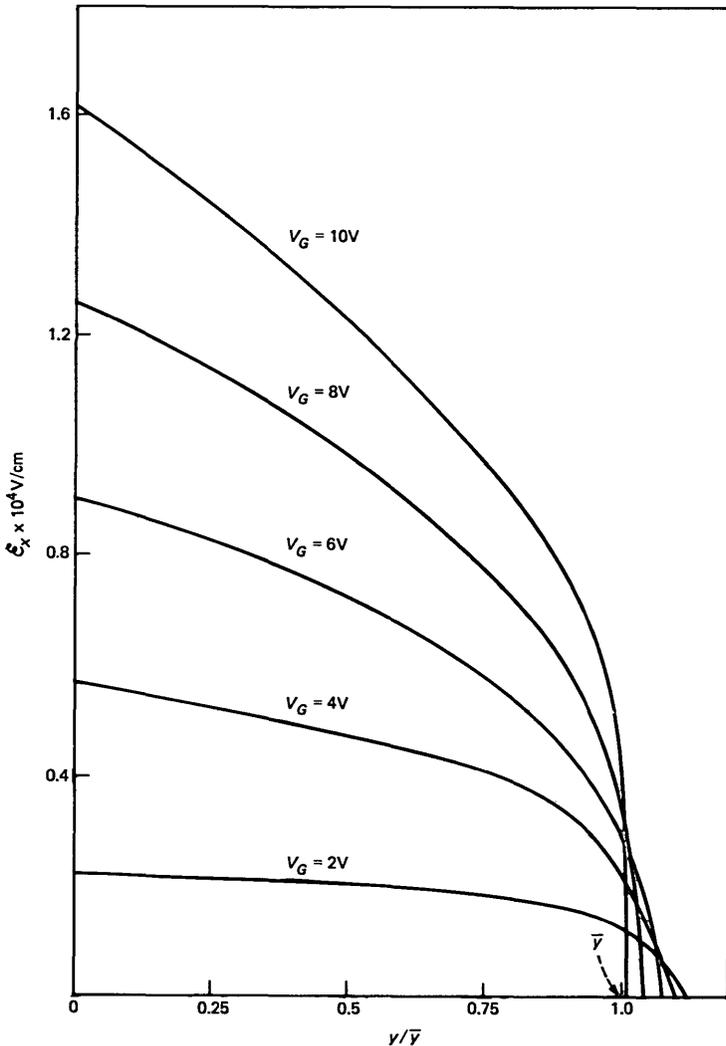


Fig. 19—Variation of transverse electric-field strength in the silicon at the surface as a function of distance along the surface from the source and through the pinch-off point parameters $t_{ox} = 2000\text{\AA}$, $N_D = 4.6 \times 10^{14} \text{ cm}^{-3}$, and $L = 5 \mu\text{m}$.

to show that a reversal in sign of \mathcal{E}_x occurs as shown by MaGowan's²⁴ numerical solution.²⁵

A.3.2 Formulation of the differential equations beyond the pinch-off point

Using the relationships described earlier, we may now find the variation of electron density and electric fields in region I. The two-dimensional nature of the problem, which we will now discuss, applies only after field inversion has occurred. The section between pinch-off and field inversion will be considered later.

Consider the vector field for the current flow in region I,

$$\mathbf{J} = J_x \mathbf{x} + J_y \mathbf{y}, \quad (88)$$

where \mathbf{x} and \mathbf{y} are unit vectors in the x and y directions. The vector components are the current densities in the x and y directions and may be written generally as

$$\frac{J_y}{q} = D \frac{dn}{dy} + \mu n \mathcal{E}_y \quad (89)$$

for the y direction and

$$\frac{J_x}{q} = D \frac{dn}{dx} + \mu n \mathcal{E}_x \quad (90)$$

for the x direction. At this point we would like to choose a set of coordinate axes to most suitably represent the current flow. We know that along the oxide-silicon interface the flow is parallel to the interface and that along the streamline the flow is in the direction of the streamline. This fact suggests that we should use polar coordinates \mathbf{r} and θ to represent the problem, where \mathbf{r} is the vector extending from the field inversion point at the surface to some point in the pinch-off zone and θ is the angle between the vector \mathbf{r} and the interface. The field inversion point is considered the origin or source point because it is the point in the channel where carriers first depart from the surface (from a charge sheet point of view). In the representation of the electron distribution by (72), then, we are taking ω to be the arc length measured from the interface for a fixed r and a variable θ . Since ω^* , σ , and N are functions of r only, the assumption becomes that we can represent the electrons by a Gaussian distribution that extends in a curvilinear fashion around a circular contour. In Fig. 20 the polar coordinates are shown schematically in an expanded view with all of the r axes emanating from the field-inversion point, y_i .

In polar coordinates (89) and (90) become

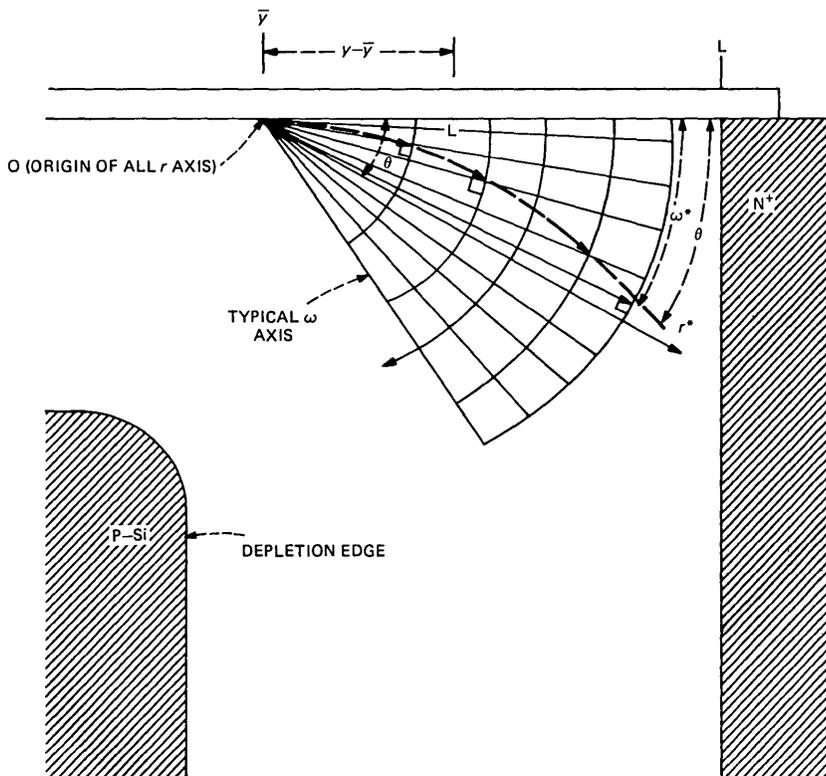


Fig. 20—Curvilinear coordinate axes in the pinch-off zone.

$$\frac{J_r}{q} = D \frac{\partial n}{\partial r} + \mu n \mathcal{E}_r \quad (91a)$$

$$\frac{J_\omega}{q} = D \frac{1}{r} \frac{\partial n}{\partial \theta} + \mu n \mathcal{E}_\omega. \quad (91b)$$

Then using n as represented by (72) in (91) and utilizing (66), we obtain for the current components

$$J_\omega = q\mu N \exp(-\eta) \left\{ \mathcal{E}_\omega - \beta_e \frac{(\omega - \omega^*)}{\sigma^2} \right\} \quad (92)$$

and

$$J_r = q\mu N \exp(-\eta)$$

$$\cdot \left\{ \mathcal{E}_r + \beta_e \left[\frac{N'}{N} + \frac{(\omega - \omega^*)}{\sigma^2} S_\omega - \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)' \right] \right\}, \quad (93)$$

where

$$\eta = \frac{(\omega - \omega^*)^2}{2\sigma^2}, \quad \mathcal{E}_\omega = -\frac{1}{r} \frac{\partial \psi}{\partial \theta},$$

$$\mathcal{E}_r = -\frac{\partial \psi}{\partial r}, \quad S_\omega = \frac{\partial \omega^*}{\partial r} - \frac{\partial \omega}{\partial r} \quad (94)$$

and the prime denotes differentiation with respect to r and ψ in the two-dimensional scalar potential. From the divergence condition (74) we find

$$\left(\frac{1}{r} \frac{\partial n}{\partial \theta} + \frac{n}{\mu} \frac{1}{r} \frac{\partial \mu}{\partial \theta} \right) \frac{1}{r} \frac{\partial \phi}{\partial \theta} + \left(\frac{\partial n}{\partial r} + \frac{n}{\mu} \frac{\partial \mu}{\partial r} \right) \frac{\partial \phi}{\partial r} = -n \nabla^2 \phi, \quad (95a)$$

where we can consider ϕ as some pseudo-potential field defined by the relations

$$-\frac{1}{r} \frac{\partial \phi}{\partial \theta} = \mathcal{E}_\omega - \beta_e \left(\frac{\omega - \omega^*}{\sigma^2} \right),$$

$$-\frac{\partial \phi}{\partial r} = \mathcal{E}_r + \beta_e \left[\frac{N'}{N} + \left(\frac{\omega - \omega^*}{\sigma^2} \right) S_\omega - \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)' \right]. \quad (95b)$$

Using (92) and (93) in (95a), we find the divergence relationship to be

$$\left\{ \mathcal{E}_\omega - \beta_e \frac{(\omega - \omega^*)}{\sigma^2} \right\} \left\{ -\frac{(\omega - \omega^*)}{\sigma^2} + \frac{1}{\mu} \frac{1}{r} \frac{\partial \mu}{\partial \theta} \right\}$$

$$+ \left\{ \mathcal{E}_r + \beta_e \left[\frac{N'}{N} + \frac{(\omega - \omega^*)}{\sigma^2} S_\omega - \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)' \right] \right\}$$

$$\cdot \left[\frac{N'}{N} + \left(\frac{\omega - \omega^*}{\sigma^2} \right) S_\omega - \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)' + \frac{1}{\mu} \frac{\partial \mu}{\partial r} \right] = \nabla^2 \phi. \quad (96)$$

At this point we need a representation for the electric fields \mathcal{E}_ω and \mathcal{E}_r . Consider \mathcal{E}_ω first. If drift and diffusion were equal and n were described by (72), then one could show that $\mathcal{E}_\omega = \beta_e [(\omega - \omega^*)/\sigma^2]$. We will assume that even though drift and diffusion are not equal, the functional form of \mathcal{E}_ω remains the same under conditions of current flow and is altered only in magnitude so that we may write

$$\mathcal{E}_\omega = K_1 \frac{\beta_e (\omega - \omega^*)}{\sigma^2}, \quad (97)$$

where K_1 is some number that is constant throughout the pinch-off zone. We will take the representation a step further by assuming that all electric fields may be represented by these functions so that we can

write

$$\mathcal{L}_r = -K_2 \frac{\beta_e \omega^*}{\sigma^2} \quad \text{and} \quad \mathcal{L}_r^* = -K_3 \frac{\beta_e \omega^*}{\sigma^2} \quad (98)$$

for the surface and along the streamline, respectively, where K_2 and K_3 are constants in the pinch-off zone, which must be determined. From (97) and (98) we obtain

$$\frac{1}{r} \frac{\partial \mathcal{L}_\omega}{\partial \theta} = \frac{K_1 \beta_e}{\sigma^2} + K_1 \frac{(\omega - \omega^*)}{\sigma^2} \frac{1}{r} \frac{\partial \beta_e}{\partial \theta} \quad (99)$$

and

$$\frac{\partial \mathcal{L}_r}{\partial r} = -\frac{K_2 \beta_e}{\sigma^2} S_\omega - K_2 \beta_e \omega^* \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) - K_2 \frac{\omega^*}{\sigma^2} \frac{\partial \beta_e}{\partial r}, \quad (100)$$

where

$$S_\omega^* = \frac{\partial \omega^*}{\partial r}.$$

Substituting (97) and (98) into (78c), we obtain for the surface (and for the region very close to the surface)

$$\beta_e = \frac{A}{2} + \sqrt{\left(\frac{A}{2} \right)^2 - \beta_0 A}, \quad (101)$$

where

$$A = \frac{(\sigma^2 \mathcal{L}_c)^2}{\beta_0 [K_1^2 (\omega - \omega^*)^2 + K_2^2 \omega^{*2}]}. \quad (102)$$

In the same way, in the vicinity of the streamline we have

$$\beta_e^* = \frac{A^*}{2} + \sqrt{\left(\frac{A^*}{2} \right)^2 - \beta_0 A^*}, \quad (103)$$

where

$$A^* = \frac{(\sigma^2 \mathcal{L}_c^*)^2}{\beta_0 [K_1^2 (\omega - \omega^*)^2 + K_3^2 \omega^{*2}]}. \quad (104)$$

Differentiating (101), we have

$$\frac{1}{r} \frac{\partial \beta_e}{\partial \theta} = \frac{1}{A} \frac{1}{r} \frac{\partial A}{\partial \theta} \left[\beta_e + \frac{\beta_0 A}{2 \sqrt{\left(\frac{A}{2} \right)^2 - \beta_0 A}} \right] \approx \frac{1}{A} \frac{1}{r} \frac{\partial A}{\partial \theta} \beta_e \quad (105)$$

and

$$\frac{\partial \beta_e}{\partial r} = \frac{1}{A} \frac{\partial A}{\partial r} \left[\beta_e + \frac{\beta_0 A}{2 \sqrt{\left(\frac{A}{2}\right)^2 - \beta_0 A}} \right] \approx \frac{1}{A} \frac{\partial A}{\partial r} \beta_e \quad (106)$$

and the approximation is good throughout most of the hot-electron regime because $A \gg \beta_0$. In (105) and (106) the derivatives of A are written

$$\frac{1}{A} \frac{1}{r} \frac{\partial A}{\partial \theta} = - \frac{2K_1^2(\omega - \omega^*)}{K_1^2(\omega - \omega^*)^2 + K_2^2\omega^{*2}} \quad (107a)$$

and

$$\frac{1}{A} \frac{\partial A}{\partial r} = \frac{-2[-K_1^2(\omega - \omega^*)S_\omega + K_2^2\omega^*S_\omega^*]}{K_1^2(\omega - \omega^*)^2 + K_2^2\omega^{*2}} - 2\sigma^2 \frac{d}{dr} \left(\frac{1}{\sigma^2} \right), \quad (107b)$$

where we have used

$$\sigma' = - \frac{\sigma^3}{2} \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) \quad (107c)$$

to transform the term in σ' .

Using (78c) for the mobility we now find

$$\frac{1}{\mu} \frac{1}{r} \frac{\partial \mu}{\partial \theta} = - \frac{1}{2} \frac{1}{\beta_e} \frac{1}{r} \frac{\partial \beta_e}{\partial \theta} \quad (108)$$

$$\frac{1}{\mu} \frac{\partial \mu}{\partial r} = - \frac{1}{2} \frac{1}{\beta_e} \frac{\partial \beta_e}{\partial r}, \quad (109)$$

so that the derivatives of the mobility may be evaluated using (105) and (106).

Poisson's equation may now be reduced to a simpler form using these results. Substituting (107c) and (107b) into (105) and (106), we find

$$\frac{1}{\beta_e} \frac{1}{r} \frac{\partial \beta_e}{\partial \theta} = \frac{2K_1^2}{\omega^*(K_1^2 + K_2^2)}, \quad \frac{1}{\beta_e^*} \frac{1}{r} \frac{\partial \beta_e^*}{\partial \theta} = 0 \quad (110)$$

and

$$\frac{1}{\beta_e} \frac{\partial \beta_e}{\partial r} = \frac{1}{\beta_e^*} \frac{\partial \beta_e^*}{\partial r} = - \frac{2S_\omega^*}{\omega^*} - 2\sigma^2 \frac{d}{dr} \left(\frac{1}{\sigma^2} \right). \quad (111)$$

The Poisson equation (76) in cylindrical coordinates is

$$\frac{\partial \mathcal{E}_r}{\partial r} + \frac{1}{r} \mathcal{E}_r + \frac{1}{r} \frac{\partial \mathcal{E}_\theta}{\partial \theta} = - \frac{q}{\epsilon} (N_A + n). \quad (112)$$

Using (110) and (111) in (99) and (100) to find $\partial \mathcal{L}_\omega / \partial \omega$ and $\partial \mathcal{L}_r / \partial r$, we can reduce (112) to an equation in $(1/\sigma^2)$ for the surface ($\omega = 0$), which is

$$K_2 \omega^* \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) + \frac{K_2 S_\omega^*}{\sigma^2} - \frac{K_2 \omega^*}{r \sigma^2} + \left(\frac{K_2^2 - K_1^2}{K_1^2 + K_2^2} \right) \frac{K_1}{\sigma^2} = - \frac{1}{\lambda_e^2}. \quad (113)$$

Using symmetrical arguments for the streamline locus we have

$$K_3 \omega^* \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) + K_3 \frac{S_\omega^*}{\sigma^2} - \frac{K_3 \omega^*}{r \sigma^2} + \frac{K_1}{\sigma^2} = \frac{-1}{\lambda_e^{*2}}. \quad (114)$$

The parameters λ_e, λ_e^* are modified Debye lengths

$$\lambda_e = \sqrt{\frac{\epsilon \beta_e}{q(N_A + N e^{-\eta})}} \quad (115)$$

and

$$\lambda_e^* = \sqrt{\frac{\epsilon \beta_e}{q(N_A + N)}}$$

due to the presence of mobile charge. The traditional Debye length is given by

$$\lambda_0 = \sqrt{\frac{\epsilon \beta_e}{q N_A}}.$$

A useful result that we shall need later is found by subtracting (113) from (114) to give

$$K_3 \left(1 - \frac{2K_1^2}{K_1^2 + K_2^2} \right) - K_2 = \frac{K_2}{K_1} \left(\frac{\sigma}{\lambda_e^*} \right)^2 - \frac{K_3}{K_1} \left(\frac{\sigma}{\lambda_e} \right)^2. \quad (116)$$

We now employ the condition

$$\oint \mathcal{L} ds = 0 \quad (117)$$

over any path in the pinch-off zone since ψ is a conservative field. Using (97), (98), and (117) we have

$$\int_0^{\omega^*} K_1 \beta_e \frac{(\omega - \omega^*)}{\sigma^2} d\omega = \int_0^r (K_3 - K_2) \beta_e \frac{\omega^*}{\sigma^2} dr'. \quad (118)$$

To simplify this result, we will assume that we may ignore the dependence of β_e upon ω on the left-hand side. Then by integrating over ω , differentiating with respect to r , and using (110) and (111), we find

$$\omega^* \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) = - \frac{2(K_3 - K_2)}{K_1 \sigma^2}. \quad (119)$$

Substituting (118) into (112) we find

$$\frac{d\omega^*}{dr} - \frac{\omega^*}{r} = \frac{2(K_3 - K_2)}{K_1} - \frac{K_1}{K_2} \left(1 - \frac{2K_1^2}{K_1^2 + K_2^2} \right) - \frac{1}{K_2} \left(\frac{\sigma}{\lambda_e} \right)^2. \quad (120)$$

Substituting (118) into (113) we have

$$\frac{d\omega^*}{dr} - \frac{\omega^*}{r} = \frac{2(K_3 - K_2)}{K_1} - \frac{K_1}{K_3} - \frac{1}{K_3} \left(\frac{\sigma}{\lambda_e^*} \right)^2. \quad (121)$$

Subtracting (119) from (120) would yield (116) as before. Returning to (110) we use (118) to give

$$\frac{1}{\beta_e^*} \frac{\partial \beta_e^*}{\partial r} = \frac{1}{\beta_e} \frac{\partial \beta_e}{\partial r} = \frac{2}{\omega^*} \left[2 \frac{(K_3 - K_2)}{K_1} - S_\omega^* \right]. \quad (122)$$

The results (109) and (121) are then used in (108) and (109) to obtain functions of K_1 , K_2 , and K_3 , which are

$$\frac{1}{\mu} \frac{1}{r} \frac{\partial \mu}{\partial \theta} = \frac{-K_1^2}{(K_1^2 + K_2^2)} \cdot \frac{1}{\omega^*} \quad (\omega = 0, \text{ silicon interface}), \quad (123a)$$

$$\frac{1}{\mu^*} \frac{1}{r} \frac{\partial \mu^*}{\partial \theta} = 0 \quad (\omega = \omega^*, \text{ streamline}), \quad (123b)$$

and

$$\frac{1}{\mu^*} \frac{\partial \mu^*}{\partial r} = \frac{1}{\mu} \frac{\partial \mu}{\partial r} = - \frac{1}{\omega^*} \left[2 \left(\frac{K_3 - K_2}{K_1} \right) - S_\omega^* \right]. \quad (123c)$$

We may now return to the evaluation of (96). The right-hand side of (96) may be evaluated using (95b) to be

$$\begin{aligned} \nabla^2 \phi = & - \frac{1}{r} \frac{\partial \mathcal{L}\omega}{\partial \theta} + \frac{\beta_e}{\sigma^2} + \left(\frac{\omega - \omega^*}{\sigma^2} \right) \frac{1}{r} \frac{\partial \beta_e}{\partial \theta} - \frac{\partial \mathcal{L}\mathcal{L}}{\partial r} \\ & - \left[\frac{N'}{N} + \left(\frac{\omega - \omega^*}{\sigma^2} \right) S_\omega - \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)' \right] \frac{\partial \beta_e}{\partial r} - \frac{\beta_e}{\partial r} \left(\frac{N'}{N} \right) \\ & + \frac{\beta_e}{\sigma^2} S_\omega^2 - \beta_e (\omega - \omega^*) S_\omega \left(\frac{1}{\sigma^2} \right)' - \beta_e (\omega - \omega^*) S_\omega \left(\frac{1}{\sigma^2} \right)' \\ & + \beta_e \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)'' - \beta_e \frac{(\omega - \omega^*)}{\sigma^2} \frac{\partial^2 \omega^*}{\partial r^2} - \frac{\mathcal{L}\mathcal{L}}{r} \\ & - \frac{\beta_e}{r} \left[\frac{N'}{N} + \frac{(\omega - \omega^*)}{\sigma^2} S_\omega - \frac{(\omega - \omega^*)^2}{2} \left(\frac{1}{\sigma^2} \right)' \right]. \quad (124) \end{aligned}$$

We need to consider (96) along the two separate loci. Along the main streamline ω^* , we have $\omega - \omega^* = 0$, and we will neglect the diffusion component N'/N compared to the drift component. Using (124) and (122) in (96) with $\mathcal{E}_r = \mathcal{E}_r^*$, we then find

$$\mathcal{E}_r^* \left[\frac{N'}{N} + \frac{1}{\mu^*} \frac{\partial \mu^*}{\partial r} \right] = \frac{\beta_e}{\lambda_e^{*2}} + \frac{\beta_e}{\sigma^2} (1 + S_\omega^2) - \frac{1}{r} \frac{2\beta_e}{\omega^*} \left[2 \left(\frac{K_3 - K_2}{K_1} \right) - S_\omega \right], \quad (125)$$

where we have considered the second derivatives of β_e and ω^* to be unimportant. The second locus is along the interface ($\omega = 0$), and in that case (96), (97), (98), (110), and (123) are used to give

$$\begin{aligned} & \left(-K_1 \frac{\omega^*}{\sigma^2} + \frac{\omega^*}{\sigma^2} \right) \times \left(\frac{\omega^*}{\sigma^2} - \frac{K_1^2}{K_1^2 + K_2^2} \frac{1}{\omega^*} \right) \\ & + \left\{ -K_2 \frac{\omega^*}{\sigma^2} - \frac{\omega^* S_\omega}{\sigma^2} - \frac{\omega^2}{2} \left(\frac{1}{\sigma^2} \right)' \right\} \\ & \cdot \left\{ -\frac{\omega}{\sigma^2} S_\omega - \frac{\omega^2}{2} \left(\frac{1}{\sigma^2} \right)' - \frac{1}{\omega} \left[2 \left(\frac{K_3 - K_2}{K_1} \right) - S_\omega \right] \right\} = \nabla^2 \phi, \quad (126) \end{aligned}$$

which may be rewritten, using (119), as

$$\begin{aligned} (-K_1 + 1) \left(1 - \frac{K_1^2}{K_1^2 + K_2^2} \left(\frac{\sigma}{\omega^*} \right)^2 \right) &= \left\{ K_2 + S_\omega - \left(\frac{K_3 - K_2}{K_1} \right) \right\} \\ &\cdot \left\{ -S_\omega + \left(\frac{K_3 - K_2}{K_1} \right) - \left(\frac{\sigma}{\omega^*} \right)^2 \left[2 \left(\frac{K_3 - K_2}{K_1} \right) - S_\omega^* \right] \right\} \\ &+ \left(\frac{\sigma^2}{\omega^*} \right)^2 \nabla^2 \phi. \quad (127) \end{aligned}$$

We will now assume that the electron distribution is well localized in the pinch-off zone, so that $\sigma \ll \omega^*$.

Then the term $\nabla^2 \phi$ may be dropped and (127) may be reduced to

$$K_2 = \left(\frac{K_1}{K_3 - K_2} \right) \frac{\left[1 + \left(\frac{K_3 - K_2}{K_1} - S_\omega \right)^2 - K_1 \right]}{\left(1 - K_1 \frac{S_\omega}{K_3 - K_2} \right)}. \quad (128)$$

Another relationship between the electron distribution parameters is found from the total current (73). Using (78a) and (72) in (73),

$$\frac{I}{qW} = N\mu_0 \left\langle \frac{\mathcal{E}_r}{\sqrt{1 + \left(\frac{\mathcal{E}_r}{\mathcal{E}_0}\right)^2 + \left(\frac{\mathcal{E}_\omega}{\mathcal{E}_c}\right)^2}} \right\rangle \int_{-\infty}^{\infty} e^{-\eta} d\omega$$

$$+ D_0 \left\langle \sqrt{1 + \left(\frac{\mathcal{E}_r}{\mathcal{E}_c}\right)^2 + \left(\frac{\mathcal{E}_\omega}{\mathcal{E}_c}\right)^2} \right\rangle \frac{d}{dr} \int_{-\infty}^{\infty} Ne^{-\eta} d\omega, \quad (129)$$

where the notation $\langle \rangle$ represents an average over the ω axis. The assumption is that the respective arguments are weak functions of ω compared to the exponential term. Then the equation may be written in simplified form as

$$\frac{\mathcal{E}_r^*}{\beta_e} \int Ne^{-\eta} d\omega + \frac{d}{dr} \int Ne^{-\eta} d\omega = \frac{I}{qWD_0 \sqrt{1 + \left(\frac{\mathcal{E}_r^*}{\mathcal{E}_c}\right)^2}}, \quad (130)$$

where the average value of \mathcal{E}_ω is taken to be zero. These two terms are the contributions of drift and diffusion, respectively, to the total current for a position r in the pinch-off zone. Equation (119) or (120) and (130) must be solved in the pinch-off zone by using boundary conditions that are appropriate to describe the electron distribution (72). Because this distribution can only make sense after a reversal of the x field in the silicon has taken place, then we cannot use the boundary conditions (83) and (84), which were established for the pinch-off point where the \mathcal{E}_x field is directed towards the gate. Instead, we must establish new boundary conditions somewhere beyond the field reversal point, which itself is beyond the pinch-off point. It is also noted that (125) and (128) are equations in K_2 , K_3 , and K_1 that must be solved together with another relation provided by these revised boundary conditions.

A.3.3 Revised boundary conditions

The pinch-off point to which the boundary conditions (82) through (87) apply occurs in the channel at a location closer to the source than to the field inversion point. In a long-channel device these points are fairly close; in a short-channel device the separation grows since the charge at the pinch-off point grows. The potential between the two points is $V_G - V_T - V_{SAT}$. The boundary conditions at a point y_1 in the channel just after field inversion has occurred are estimated in the following way.

From (69) we obtain

$$\left. \frac{d\mathcal{E}}{dy} \right|_{y=\bar{y}} = - \frac{\bar{R}}{(1 - \bar{R})} \frac{\bar{\mathcal{E}}^2}{\beta_e} - \frac{I(1 - \bar{R})}{Z\mu^2\bar{Q}} \left. \frac{d\mu}{dy} \right|_{y=\bar{y}}. \quad (131)$$

In calculating (131) we are considering the derivatives of R to be negligible. The second term may be evaluated using (109) with $r = y$ and the one-dimensional form (64) of μ to obtain

$$\frac{d\mu}{dy} = -\mu \frac{\beta_0}{\beta_e} \frac{\bar{\mathcal{E}}}{\mathcal{E}_c^2} \frac{d\bar{\mathcal{E}}}{dy} \Big|_{y=\bar{y}}. \quad (132)$$

Substituting (132) into (131), using (79) for β_e , and solving for $(d\bar{\mathcal{E}})/(dy)$, we find

$$\frac{d\bar{\mathcal{E}}}{dy} \Big|_{y=\bar{y}} = -\frac{\bar{R}}{(1-\bar{R})} \frac{\bar{\mathcal{E}}^2}{\beta_0}. \quad (133)$$

Using (133), (99), and (110) in (76), we obtain

$$-\frac{\bar{R}}{(1-\bar{R})} \frac{\bar{\mathcal{E}}^2}{\beta_0} + \frac{\beta_e K_1}{\sigma_1^2} \left(\frac{1 - (K_1/K_2)^2}{1 + (K_1/K_2)^2} \right) = -\frac{q}{\epsilon} (N_A + n). \quad (134)$$

In (134) we have mixed terms for two different positions. The term in $\bar{\mathcal{E}}^2$ actually applies only at the pinch-off position, whereas the term in σ_1^2 applies to a point y_1 just beyond the field-reversal point. We will assume that we can use the term in $\bar{\mathcal{E}}^2$ at y_1 , if we change $\bar{\mathcal{E}}$ to $h\bar{\mathcal{E}}$, where h is a parameter >1 , which determines how much \mathcal{E}_y has increased in the interval \bar{y} to y_1 . For almost all levels of current above threshold we can ignore the charge term on the right-hand side so that we obtain the result

$$\sigma_1 = \sqrt{K'_1 \frac{(1-\bar{R})}{\bar{R}} \frac{\beta_0}{h} \frac{1}{\bar{\mathcal{E}}}} \sqrt{1 + \left(\frac{h\bar{\mathcal{E}}}{\mathcal{E}_c} \right)^2}, \quad (135)$$

where

$$K'_1 = K_1 b$$

and

$$b = \frac{1 - (K_1/K_2)^2}{1 + (K_1/K_2)^2}. \quad (136)$$

The result predicts that as $\bar{\mathcal{E}}$ increases, σ_1 at first decreases as $\sqrt{K'_1}/\bar{\mathcal{E}}$ for small values of $\bar{\mathcal{E}}$ and then decreases as $\sqrt{K'_1}$ for larger values of $\bar{\mathcal{E}}$. Using (98) to express the continuity of the longitudinal field, we can use (135) to find the boundary condition

$$\omega_1^* = \frac{K'_1}{K_2} \frac{\beta_0}{h\bar{\mathcal{E}}} \frac{(1-\bar{R})}{\bar{R}}. \quad (137)$$

A further boundary condition is imposed by Gauss' law at the oxide-silicon interface between \bar{y} and L as

$$\epsilon \mathcal{E}_\omega|_{\omega=0} = \epsilon_o \mathcal{E}_{ox}|_N, \quad (138)$$

where $\mathcal{E}_{ox}|_N$ is the normal field in the oxide. The reader may show fairly easily that this condition may be written

$$\frac{d\mathcal{E}_\omega}{dy} = \frac{C_o}{\epsilon} \mathcal{E}_y. \quad (139)$$

Now the left-hand side using (97), (111), and (118) may be expressed as K_1^2/K_2 so that we find, after considerable algebra, the relationship

$$\left(\frac{K_1}{K_2}\right) = \frac{C_o \beta_o}{\epsilon h} \mathcal{E} \left(\frac{1 - \bar{R}}{\bar{R}}\right) b$$

or

$$\frac{K_1}{K_2} = \alpha b \equiv \alpha',$$

where

$$\alpha = \frac{C_o \beta_o (1 - \bar{R})}{\epsilon h \mathcal{E} \bar{R}}. \quad (140)$$

If we knew Q_1 , we could obtain the boundary condition N_1 from the normalization relation

$$N_1 = \frac{qQ_1}{\sqrt{2\pi\sigma_1}}. \quad (141)$$

We will determine Q_1 later since it is not required at this point to proceed with a solution of these equations. Also, the boundary condition on \mathcal{E}_ω is, from (140),

$$\mathcal{E}_{\omega 1} = \frac{K_1}{K_2} h \mathcal{E} = \frac{C_o \beta_o}{\epsilon} \frac{(1 - \bar{R})}{\bar{R}} b. \quad (142a)$$

To establish the potential V_1 for the new boundary condition, we need two contributions, $\Delta_1 V$ and $\Delta_2 V$, determined as follows. From the pinch-off point to the field-reversal point we have $\Delta_1 V = \bar{Q}/C_o + V_T - V_{FB}$, and from the field reversal point to y_1 we have $\Delta_2 V$. In this section, as a general approximation we may say

$$\mathcal{E}_\omega \approx \frac{\epsilon_{ox}}{\epsilon t_{ox}} \left(\int_{\bar{y}}^y \mathcal{E}_y dy - V_G + \bar{V} \right)$$

or, alternatively, if we integrate from y_i , the field reversal point, then

$$\mathcal{E}_\omega = \frac{\epsilon_{ox}}{\epsilon t_{ox}} \int_{y_i}^y \mathcal{E}_y dy. \quad (142b)$$

For $y = y_1$, the new boundary condition, we have $\int_{y_i}^{y_1} \mathcal{L}_y dy = \Delta_2 V$, which we can then use in (142b), along with (142a), to find

$$\Delta_2 V = \beta_0 \left(\frac{1 - \bar{R}}{\bar{R}} \right) b.$$

Using \bar{Q}/C_o from (18c), the total voltage drop between \bar{y} and y_1 is therefore

$$\begin{aligned} \Delta V &\equiv V_1 - V_{\text{SAT}} = \Delta_1 V + \Delta_2 V \\ &= V_T - V_{\text{FB}} + \beta_0 \frac{(1 - \bar{R})}{\bar{R}} b + \frac{\bar{Q}_0}{C_o} \sqrt{1 + \left(\frac{I}{W v_s \bar{Q}_0} \right)^2}. \end{aligned} \quad (143)$$

We then have for r_1

$$r_1 = \frac{2}{(h+1)} \frac{\Delta V}{\bar{\mathcal{L}}}, \quad (144)$$

where an average value of \mathcal{L} between \bar{y} and y_1 has been used. Using (143) with $b \approx 1$, we have

$$\begin{aligned} r_1 &= \frac{2}{(h+1)\bar{\mathcal{L}}} \left[V_T - V_{\text{FB}} + \beta_0 \left(\frac{1 - \bar{R}}{\bar{R}} \right) \right. \\ &\quad \left. \cdot \left(1 + \frac{C_o + C_s}{C_o} \sqrt{1 + \left(\frac{I}{W V_s \bar{Q}_0} \right)^2} \right) \right]. \end{aligned} \quad (145)$$

A.3.4 Approximate solutions to the equations

Using the integrating factor $-(1/r)$ and the boundary values ω_1^* , r_1 , we may solve (120) to obtain

$$\omega^* = Cr \ln \left(\frac{r}{r_1} \right) + \frac{\omega_1^*}{r_1} r, \quad (146a)$$

$$S_\omega^* = \frac{\partial \omega^*}{\partial r} = C \ln \left(\frac{r}{r_1} \right) + C + \frac{\omega_1^*}{r_1}, \quad \frac{\omega^*}{r} = C \ln \left(\frac{r}{r_1} \right) + \frac{\omega_1^*}{r_1}, \quad (146b)$$

where

$$C = 2 \left(\frac{K_3 - K_2}{K_1} \right) - \frac{K_1}{K_2} \left(1 - \frac{2K_1^2}{K_1^2 + K_2^2} \right) - \frac{1}{K_2} \left(\frac{\sigma}{\lambda_e} \right)^2 \quad (146c)$$

and is considered to be constant. Used in (119), this result can be written, by using (140) and ignoring the term in σ/λ_e , as

$$\frac{1}{v} \frac{dv}{d\omega^*} = - \frac{(C + \alpha)}{S_\omega^*} \frac{1}{\omega^*},$$

where $v = (1/\sigma^2)$, which then yields

$$\sigma^2 = \sigma_1^2 \left(\frac{\omega^*}{\omega_1^*} \right)^a, \quad (147a)$$

where

$$a = \frac{C + \alpha}{S_\omega^*}. \quad (147b)$$

We now use (115); assuming that $K_1 \ll K_2$ and that $K_2 \approx K_3$ we can write (115) as

$$K_3 - K_2 \approx \frac{K_2}{K_1} \frac{\sigma^2}{\epsilon \beta_e} qN. \quad (148)$$

This condition on K_1 amounts to saying that the field that is pushing carriers away from the surface is much smaller than the field that is driving the current; this is always found to be the case in the on region.

To find N , we return to (130) and assume that at the position y_1 the dominant conduction component is drift. That is, at the saturation point the diffusion current accounts for about one half to one third of the current flow but at a distance, r_1 , further down the channel, drift has again become dominant. If drift has become dominant, then we can say from (130)

$$Q = \frac{I}{W\mu \mathcal{E}_r^*}, \quad (149a)$$

and as velocity saturation becomes predominant, this becomes

$$Q = \frac{I}{Wv_s}. \quad (149b)$$

From (149b) we can see that the gradient of the charge along the streamline actually goes to zero in the limit of velocity saturation. Therefore, the assumption of total drift is reasonable. With this assumption (149b) gives us a boundary condition Q_1 at y_1 , and therefore, N_1 , which would be

$$N_1 = \frac{I}{\sqrt{2\pi}} W\mu_1^* \mathcal{E}_{r1}^* \sigma_1 = > \frac{I}{W\sqrt{2\pi}\sigma_1 v_s},$$

from the normalization of the charge, which approaches the velocity-saturated condition shown. From (149b) we can write the equation for N for any position as

$$N = \frac{I}{W\mu \mathcal{E}_r^* \sqrt{2\pi}\sigma}. \quad (150)$$

Using (150) and (103) in (148) and then the result in (146c), we find

$$C = 2 \frac{K_2}{K_1^2} \frac{q}{\epsilon} \frac{\beta_0 K_3 \bar{Q}_0 \bar{\mathcal{E}}^2}{\mathcal{E}_c^4 \sqrt{2\pi}} \frac{\omega_1^{3a/2}}{\sigma_1^3} \cdot \omega^{2-\frac{3a}{2}}. \quad (151)$$

Since C is independent of position, we have from (151)

$$a \simeq 4/3. \quad (152)$$

Substituting from (135), (137), and (140) we can reduce (151) to

$$C \simeq C + \alpha = \frac{\gamma}{h^2 K_1^{3/2}}, \quad \gamma = \sqrt{\frac{2}{\pi}} \left(\frac{1 - \bar{R}}{\bar{R}} \right)^{3/2} \frac{\beta_0 (C_o + C_s)}{\epsilon \mathcal{E}_c}. \quad (153)$$

In writing (148) we assumed that $K_3 - K_2 \ll K_3, K_2$, and we have used this fact again here. It is noted from (147) that (152) implies the result

$$S_\omega = \frac{3}{4} (C + \alpha). \quad (154)$$

We now use (150) in (125). From (150) or (149) we have

$$\frac{N'}{N} = - \left[\frac{1}{\mathcal{E}_r^*} \frac{d\mathcal{E}_r^*}{dr} + \frac{1}{\mu^*} \frac{d\mu^*}{dr} + \frac{1}{\sigma} \frac{d\sigma}{dr} \right], \quad (155)$$

and using this result and (107c) in (125), we obtain

$$\mathcal{E}_r^* \left[- \frac{1}{\mathcal{E}_r^*} \frac{d\mathcal{E}_r^*}{dr} + \frac{\sigma^2}{2} \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) \right] = \beta_e \left[\frac{1}{\lambda_e^{*2}} + \frac{1 + S_\omega^2}{\sigma^2} \right],$$

where we have dropped all terms in $(\sigma/\omega^*)^2$ (such as the term in β/ω^*) and $\sigma^2/r\omega^*$ because we are assuming that the Gaussian is localized well enough that $(\sigma/\omega^*)^2 \ll 1$. Using (100) for the streamline and (114), we obtain

$$K_3 \frac{S_\omega}{\sigma^2} + \frac{3}{2} K_3 \omega^* \frac{d}{dr} \left(\frac{1}{\sigma^2} \right) = - \left(\frac{1}{\lambda_e^{*2}} + \frac{1 + S_\omega^2}{\sigma^2} \right),$$

and then substituting from (119) and (146c)

$$K_3 = \frac{1 + S_\omega^2}{\frac{3(C + \alpha)}{2} - S_\omega^*}, \quad (156)$$

where the term $(1/K_2)(\sigma/\lambda_e)^2$ has been considered small.

We are going to assume that in the range of high fields in which we are interested,

$$C \gg 1, \quad K_1$$

and $S_\omega \approx 0$, which means that $(\partial\omega^*/\partial r) \approx (\partial\omega/\partial r)$ [cf. (94)], so that the streamline locus is effectively a straight line. Then (128) and (156) may be subtracted, and (146c) may be used to give

$$S_\omega^* = \frac{3}{2} (C + \alpha) \frac{1}{(1 + K_1) \left(\frac{C + \alpha}{2}\right)} \quad (157)$$

$$K_2 = \frac{2}{C + \alpha} \left[1 + \frac{(C + \alpha)^2}{4} \right] \quad (158)$$

and

$$K_3 = \frac{2}{C + \alpha} \left[1 + \frac{(C + \alpha)^2}{4} \right] + \left(\frac{C + \alpha}{2}\right) K_1. \quad (159)$$

Using (138) and (153) we then find

$$K_1 = \left(\frac{\gamma\alpha}{2h^2}\right)^{2/5}, \quad K_2 = \left(\frac{\gamma}{2h^2}\right)^{2/5} \left(\frac{1}{\alpha}\right)^{3/5} \quad (160, 161)$$

$$K_3 = \left(\frac{\gamma}{2h^2}\right)^{2/5} \left(\frac{1}{\alpha}\right)^{3/5} + \frac{1}{2^{2/5}} \left(\frac{\gamma}{2h^2}\right)^{4/5} \left(\frac{1}{\alpha}\right)^{1/5} \quad (162)$$

$$C + \alpha = \left(\frac{\gamma}{h^2}\right)^{2/5} 2^{1/5} \left(\frac{1}{\alpha}\right)^{3/5}. \quad (163)$$

These are the constants that characterize the electric fields. They are determined by the parameter α , or, equivalently, \mathcal{E} , through (140). These results are plotted in Fig. 21 over the total range of \mathcal{E} for a practical device. All of the assumptions we have made are validated by the plots. We now use these results to determine the length of the pinch-off zone $L - \bar{y}$. From (117) we have

$$\frac{\omega^{*2}}{\sigma^2} = \frac{2(K_3 - K_2) (V_{DS} - V_1)}{K_1 K_3 \beta_e}. \quad (164)$$

Using (103) for β_e along the streamline and substituting (147) for σ , we obtain a result for ω^* , which is

$$\omega^* = \frac{\omega_1^*}{\sigma_1^{3/2}} (C + \alpha)^{3/4} \left[\frac{(V_{DS} - V_1)\beta_0}{\mathcal{E}_c^2} \right]^{3/4}. \quad (165)$$

Now $\omega^* = r^*\theta$ and $L - y_i = r^*\cos\theta$, so that we may write the length of the pinch-off zone as

$$L - y_i = \frac{\cos\theta}{\theta} \frac{\omega_1^*}{\sigma_1^{3/2}} (C + \alpha)^{3/4} \left[\frac{(V_{DS} - V_1)\beta_0}{\mathcal{E}_c^2} \right]^{3/4}, \quad (166)$$

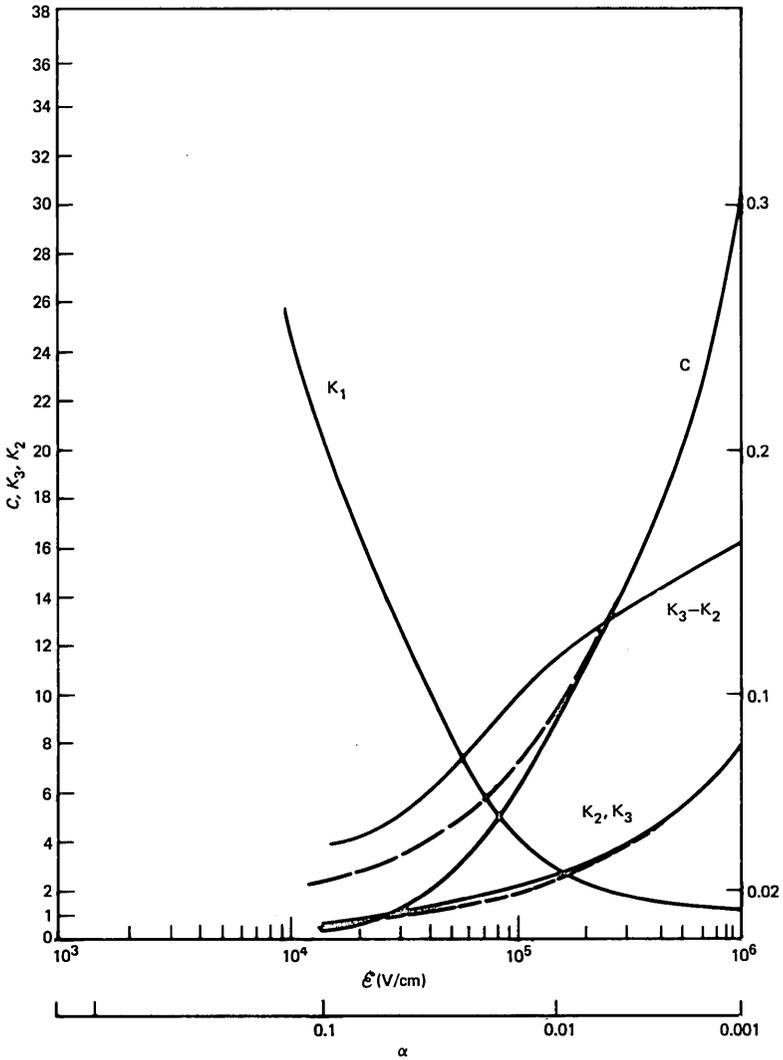


Fig. 21—Variation of the field constants with pinch-off field \mathcal{E} or its normalized form, α . The device parameters are $t_{ox} = 250\text{\AA}$ and $N_A = 5 \times 10^{16} \text{ cm}^{-3}$.

where θ is the angle formed by the interface and the vector from the point y_1 to the streamline axis. Using (135), (137), (140), and (163) for σ_1 , ω_1 , and $(C + \alpha)$, respectively (under the condition $b \approx 1$), we have

$$L - y_i = \frac{\cos \theta}{\theta} \cdot \alpha^{5/4} \cdot \frac{\epsilon}{C_o} \cdot 8^{3/4} \cdot \left(\frac{\bar{R}}{1 - \bar{R}} \right)^{3/4} \beta_0^{1/4} (V_{DS} - V_1)^{3/4}. \quad (167)$$

We now require an estimate of θ . In solving for the constants K_1 ,

K_2 , and K_3 , we have assumed that $S_\omega \approx 0$ and that S_ω^* was constant, which implies that θ can be considered constant since $\omega = \gamma\theta$, and so

$$\theta = C \left\langle \ln \frac{\gamma}{\gamma_1} \right\rangle, \quad (168)$$

where the \ln term must be written as some average. We cannot actually estimate θ because the solution does not take into account the effect of the finite depth of the drain junction. We will, therefore, use a value of $\theta = 45$ degrees to expedite calculation of (167). This is a reasonable assumption based on the examination of two-dimensional numerical solutions of this region. (Note that the function $\cos \theta/\theta$ is not a strong function anyway, varying between 0.5 and 0.85 for θ between 30 and 45 degrees).

The result (167) applies for voltages $V_{DS} > V_1$ but not to the situation when $V_1 > V_{DS} > V_{SAT}$. To extend the result over the interval $\Delta V = V_1 - V_{SAT}$ and thus to extend $L - y_1$ to $L - \bar{y}$, we will write (167) approximately as

$$\Delta L_H = L - \bar{y} = \frac{h^{1/5}}{\mathcal{E}^{3/2}} \cdot A_1 [(V_{DS} - V_{SAT} + A_2 \Delta V)^{3/4} - (A_2 \Delta V)^{3/4}], \quad (169)$$

where

$$A_1 = \frac{\cos \theta}{\theta} \left(\frac{1}{\gamma} \right)^{4/5} 8^{1/2} \left(\frac{C_o}{\epsilon} \right)^{1/2} \left(\frac{1 - \bar{R}}{\bar{R}} \right)^{3/4} \beta_0^{3/4} \quad (170)$$

and h is determined by the condition that $L - \bar{y} = r_1$ for $V_{DS} = V_{SAT} + \Delta V$ and A_2 is an additional constant that we must find. Using (145) in (169) and for the saturation condition, noting that the current in velocity saturation may be written $I = W\mu Q \cdot \mathcal{E}^2 / (1 - \bar{R}) \mathcal{E}_c$, we have

$$\frac{2}{(h + 1)} \left(\frac{\bar{Q}_0}{C_o} \right)^{1/4} \frac{\mathcal{E}^{1/2}}{\mathcal{E}_c^{1/2}} \cdot \frac{1}{\mathcal{E}} = \frac{A_1 h^{1/5}}{\mathcal{E}^{3/2}} [(1 + A_2)^{3/4} - A_2^{3/4}]. \quad (171)$$

To determine A_2 and h , we require another relation, which is given by the condition of the continuity of the derivative of the current at the boundary between the triode and the saturation regions. This condition is found to be [cf. (48)]

$$\bar{Q}/C_o = - \mathcal{E}_c \frac{d\bar{y}}{dV_{DS}} \left(\frac{I}{Wv_s C_o} \right). \quad (172)$$

Since $\bar{Q} \approx I/Wv_s$ for higher channel fields, (172) becomes

$$1 = - \frac{\mathcal{E}_c d\bar{y}}{dV_{DS}}. \quad (173)$$

From (169) we have

$$\frac{d\bar{y}}{dV_{DS}} = \frac{h^{1/5}}{\mathcal{E}^{3/2}} A_1 \frac{3}{4} \frac{1 + dV_T/dV_{DS}}{(V_{DS} - V_{SAT} + A_2 \Delta V)^{1/4}} - \frac{3h^{1/5} A_1}{2 \mathcal{E}^{5/2}} [(V_{DS} - V_{SAT} + A_2 \Delta V)^{3/4} - (A_2 \Delta V)^{3/4}]. \quad (174)$$

For $V_{DS} = V_{SAT}$, (174) reduces to

$$\frac{d\bar{y}}{dV_{DS}} = - \frac{3}{4} \frac{h^{1/5} A_1}{\mathcal{E}^{3/2}} \frac{1}{(A_2 \Delta V)^{1/4}}, \quad (175)$$

where we have considered $(dV_T/dV_{DS}) \ll 1$. Using (175) in (173) gives the condition

$$\frac{3}{4} \frac{h^{1/5} A_1}{A_2^{1/4} \Delta V^{1/4} \mathcal{E}^{1/2}} = 1. \quad (176)$$

In writing (174) we have assumed that the dependence of $h^{1/5}$ on \mathcal{E} and hence V_{DS} can be ignored. Noting that for $\bar{y} = y$ we have $L - \bar{y} = L - y_1 = r_1$ and $V_{DS} - V_{SAT} = \Delta V$; then using (144) in (169) we have

$$\frac{2\Delta V}{(h+1)\mathcal{E}} = \frac{h^{1/5} A_1 \Delta V^{3/4}}{\mathcal{E}^{3/2}}. \quad (177)$$

Using this result in (176) yields

$$A_2 = \left(\frac{3}{2(h+1)} \frac{\mathcal{E}_c}{\mathcal{E}} \right)^4. \quad (178)$$

Since $h > 1$ and $\mathcal{E} \geq \mathcal{E}_c$ in the range of interest here, A_2 will be a small number (i.e., < 0.1 , typically). We are, therefore, justified in treating the square brackets as 1 in (171) and so we find for h

$$(h+1)h^{1/5} = \mathcal{E}^2 \left(\frac{\bar{Q}_0}{C_0} \right)^{1/4} \frac{1}{\mathcal{E}_c^{1/2}} \frac{1}{A_1}. \quad (179)$$

If we considered a typical value of $\mathcal{E} \approx 10^5$ V/cm (i.e., well into the hot-electron regime) and use $A_1 \approx 175$ [cf. (170)], then we find $h \approx 2.8$. For higher values of \mathcal{E} , h would increase, although we would not expect values much higher than $\mathcal{E} = 10^5$ V/cm in practice. Therefore, we will treat h generally as a constant with a value 2.5 through 3.0, especially since it appears in $L - \bar{y}$, the result of major interest, only as $h^{1/5}$.

A.4 Discussion of results

The result (169) is valid for high values of pinch-off field \mathcal{E} so that the hot-electron approximations [cf. (105) and (106)] are valid. For low values of \mathcal{E} , $L - y$ becomes anomalously large according to (169) and is obviously incorrect. However, it has been shown that for gate voltages near and below V_T , a good representation of ΔL is

$$\Delta L_L = L - \bar{y} = \sqrt{2 \frac{\epsilon}{qN_A} (V_{DS} - V_{SAT})}. \quad (180a)$$

For gate voltages in the 'on' region (180) becomes notoriously gross. If we now combine the results (180) and (169) in the form

$$\frac{1}{\Delta L} = \frac{1}{\Delta L_H} + \frac{1}{\Delta L_L}, \quad (180b)$$

we obtain a representation of $\bar{y} = L - \Delta L$, which is good for all regions of operation.

We demonstrate this agreement by the curve shown in Fig. 22. The curves show data and theory for a 0.5- μm device. The lower solid theory line shows the agreement obtained using (179) to determine \bar{y} , which is then used in the velocity-saturated model of the companion paper. The agreement is quite good. The upper solid line shows the result that is obtained if the ΔL_L [cf. (180a)] is used alone to predict \bar{y} and there is considerable error. However, there is not as much error as one might expect on the basis of $(L/L - \Delta L_L)I_{SAT}$, which is normally considered to be the case where I_{SAT} is the current at the onset of saturation. This is clear from the comparison of ΔL , ΔL_L , ΔL_H , and $\Delta L'$ shown in Fig. 23. The variation of ΔL in Fig. 23 is much more than is manifested in Fig. 22 because the velocity-saturated current [cf. (20b)]

$$I = Wv_s C_o \left[\sqrt{\left(\frac{\mathcal{E}_c \bar{y}}{a^*}\right)^2 + \frac{(V_{GS} - V_T) - (Q_0/C_o)^2}{a^*}} - \frac{\mathcal{E}_c \bar{y}}{a^*} \right]$$

loses its dependence upon \bar{y} . In the limit of total velocity saturation there is no dependence at all. Therefore, we are completely justified in making the approximation such as (168). In fact, the coefficient A_1 could be considerably in error without having much effect. The same may be said about the radius of curvature of the junction. The fact that we treated a vertical instead of a cylindrical junction and did not consider explicitly the junction depth is really of very little consequence. The most important feature of the result is the inverse dependence of ΔL_H upon $\mathcal{E}^{3/2}$. It is this dependence that allows ΔL to track the device geometries and applied voltages in a continuous fashion.

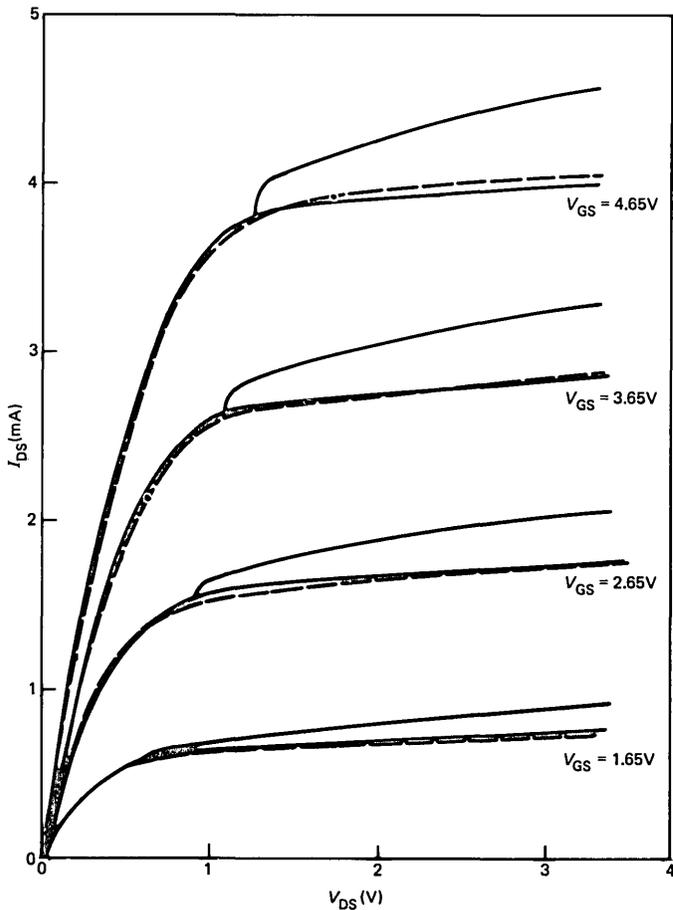


Fig. 22—Comparison of drain voltage data (dashed curve) of a velocity saturated device with theory using the composite result (128) (lower solid curve) and the simple depletion result (127) (upper solid curve). The effect of series resistance ($\approx 50\Omega$) has been removed from the data and the device parameters are $L = 0.5 \mu\text{m}$, $t_{\text{ox}} = 250\text{\AA}$, $W = 10 \mu\text{m}$, $N_A = 5 \times 10^{16} \text{cm}^{-3}$, $r_j = 0.25 \mu\text{m}$, $\mu_0 = 650 \text{cm}^2/\text{V-s}$, $\theta = 0.03$, and $v_s = 10^7 \text{cm/s}$.

APPENDIX B

B.1 Introduction

In a previous work, we had examined the problem of ensuring the continuity of current in moving from the subthreshold to the above-threshold regions by introducing a voltage parameter δV_G into the expression for the saturated current so that it became

$$I_{AT} = \frac{W\mu C_o}{2\bar{y}} (V_G - V_T + \delta V_G)^2, \quad (181)$$

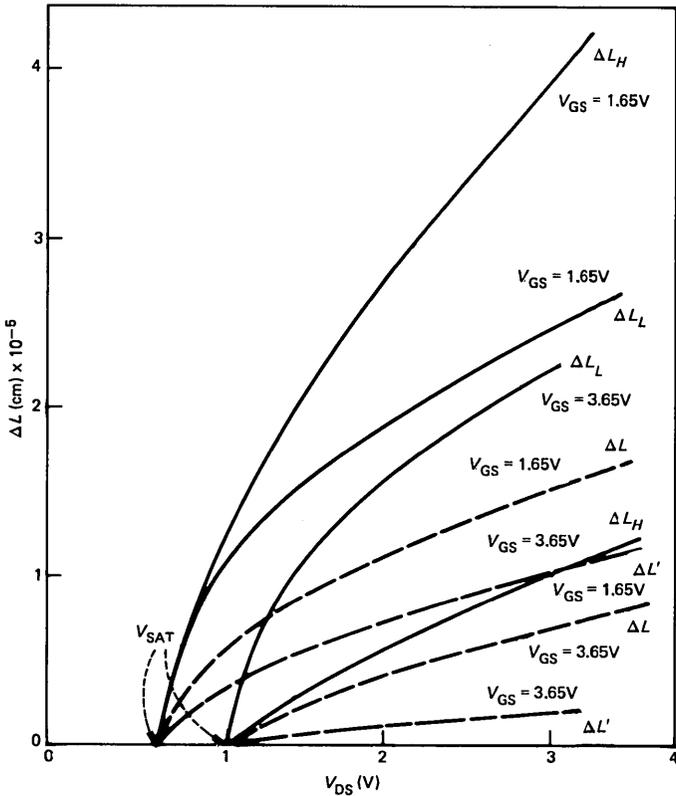


Fig. 23—Variations of ΔL_H , ΔL_L (solid curve) and ΔL (dashed curve) for the device in Fig. 11. Also indicated is the effective $\Delta L'$ if the current in saturation is $I = I_{SAT}L/L - \Delta L$, where I_{SAT} is the pinch-off value. $\Delta L'$ is less than ΔL because of the effects of velocity saturation.

where I_{AT} refers to the above-threshold current.¹⁶ The parameter δV_G was introduced to account for the absence of diffusion in the above-threshold formulation. It was determined on the basis of a unique value of current at the threshold condition from consideration of the above-threshold and subthreshold currents. However, to be consistent in this approach and to have a useful result, we must also require continuity of the derivative of the current between the subthreshold and above-threshold regions of operation.

In this appendix we will re-examine the conditions of the continuity of current and its derivative at two important transitions in the device, namely, (1) the transition between the subthreshold and the above-threshold saturation regions, and (2) the transition between the triode region and the saturation region, both of which are above threshold. These conditions are needed to determine \bar{R} , the fraction of the total current carried by diffusion at the pinch-off point above threshold.

It is noted that the former approach¹⁶ of substituting the parameter δV_G into the above-threshold result for all gate voltages is not complete. When determined self-consistently, the parameter δV_G gives only the voltage for which the currents below and above threshold are equal but says nothing about their derivatives.

The transition from the nonsaturated region below threshold to the triode region above threshold will be considered only after the results have been obtained since it only exists over a drain voltage of about $2kT/q$ and hence is relatively unimportant. Also, the transition from linear to triode regions below threshold does not need to be considered since its continuity has already been established.²⁵

B.2 Transition from subthreshold to above-threshold saturation conduction

Since we are concerned at this transition with currents flowing just above the threshold voltage (i.e., almost at the threshold point), then we are justified in using the above-threshold current expressions that do not include velocity saturation because the channel fields for such voltages are less than \mathcal{E}_c . From (83) we have

$$\frac{\bar{Q}}{C_o} = \left(\frac{1 - \bar{R}}{\bar{R}} \right) \frac{kT}{q} \frac{(C_o + \bar{C}_S F)}{C_o}, \quad (182a)$$

where

$$\bar{C}_S = \sqrt{\frac{\epsilon q N_A}{2(V + V_{BS} + 2\phi_F)}} \quad (182b)$$

is the semiconductor depletion capacitance at the pinch-off point in the channel. At the threshold condition we will have $\bar{C}_S = C_S$, i.e., the same value at the source and the pinch-off point. For a constant T , \bar{Q} is dependent on V_{GS} only through \bar{C}_S , on V_{DS} only through $F(V_{DS})$, and on V_{BS} through both of these. (It may also have some dependence upon V_{GS} and V_{DS} through the parameter \bar{R} .) We will assume at this point that the dependence of \bar{Q} upon bias parameters is sufficiently weak that we may ignore its derivatives in the calculation of the derivatives of the current above threshold. We shall re-examine the validity of the assumption later. The result (82) suggests that we should use it in the triode region equation

$$I_{DS} = \frac{\mu C_o W}{L} \left[(V_{GS} - V_T(V_{DS})) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (183)$$

to obtain a modified form of the saturation current, which is

$$I_{SAT} = \frac{\mu C_o W}{2\bar{y}} \left[(V_{GS} - V_T(V_{DS}))^2 - \left(\frac{\bar{Q}}{C_o} \right)^2 \right]. \quad (184)$$

The threshold voltage has been written as a function of drain voltage to show that the $3/2$ power terms of the conventional triode expression may be represented.¹⁶ We conclude from (184), therefore, that true square law behavior of the device may only be observed for $V_{GS} - V_T(V_{DS}) \gg \bar{Q}/C_o$.

In considering the saturation current at the transition between the two regions, we are going to assume that an offset gate voltage of η above the threshold voltage must be applied to the device in order for the above-threshold theory to predict a finite current flowing at the transition point. This situation is illustrated in Fig. 24, which shows data for a reasonably short-channel device ($L = 2.5 \mu\text{m}$, $t = 500\text{\AA}$, and $N_A \approx 10^{16} \text{cm}^{-3}$) in the region near the threshold voltage. The voltage η is shown as the voltage increment from V_T to the onset of the straight section. An equivalent way of stating this condition is that the above-threshold current merges with the subthreshold current not

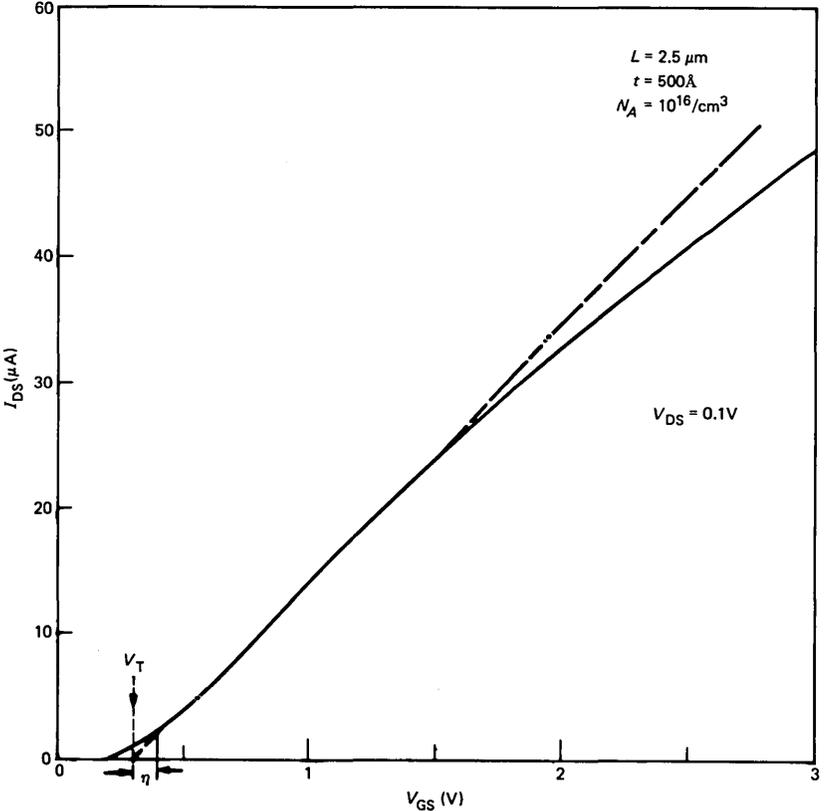


Fig. 24(a)—Variation of drain current with gate voltage showing the determination of η and its relation to V_T for linear region characteristic.

for $V_{GS} = V_T$ but for a slightly larger value of $V_{GS} = V_T + \eta$, so that (184) becomes

$$I_{SAT}|_{V_T+\eta} = \frac{\mu WC_o}{2\bar{y}} \left[\eta^2 - \left(\frac{Q}{C_o} \right)^2 \right]. \quad (185)$$

In writing the saturation current we have used \bar{y} in place of L in (184) and (185), where $\Delta L = L - \bar{y}$ is the channel-length modulation in saturation operation. The relative size of ΔL to L can be significant in a short-channel device.

At the gate voltage V_T , the subthreshold theory predicts the flow of a finite diffusion current, which would correspond to an effective

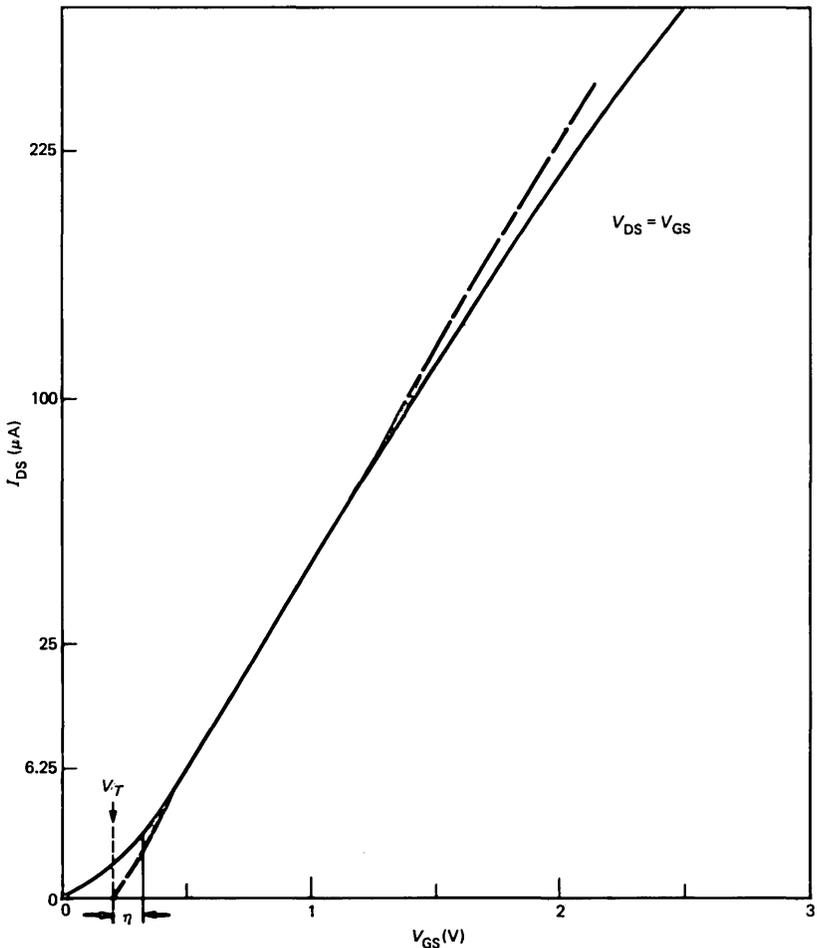


Fig. 24(b)—Variation of drain current with gate voltage showing the determination of η and its relation to V_T for saturation region characteristic.

surface potential $\hat{\phi}_S = 2\phi_F$. Since the transition between the regions takes place for $V_G = V_T + \eta$, in a similar way we should introduce a corresponding offset surface potential into the subthreshold formulation. It may be defined as the additional surface potential above the threshold value ($\hat{\phi}_S = 2\phi_F$) that is achieved when the additional voltage η is applied. The subthreshold current may be written²⁵.

$$I_{ST} = \frac{qWD}{\bar{y}} \sqrt{\frac{\epsilon}{2qN_A(V_{BS} + \hat{\phi}_S)}} \left(\frac{kT}{q}\right) n_i e^{\frac{-q\phi_F}{kT}} e^{\frac{q\hat{\phi}_S}{kT}}, \quad (186)$$

where $\hat{\phi}_S$ is the effective surface potential in subthreshold operation. Therefore, when a merging of the subthreshold and above-threshold currents is considered, the transition will occur not for $\hat{\phi}_S = 2\phi_F$ but rather for a slightly larger surface potential of $2\phi_F + \xi$. The voltage variation of the current is predominantly in the exponential term through the effective surface potential $\hat{\phi}_S$, which is obtained from the equation

$$V_{GS} = V_{FB} + \hat{\phi}_S + \frac{1}{C_o} \sqrt{2\epsilon q N_A (V_{BS} + \hat{\phi}_S)} \cdot F. \quad (187)$$

Therefore, we will neglect the voltage dependence of $\hat{\phi}_S$ in the pre-exponential term. The factor F has been defined previously for short-channel devices.¹⁶ At the threshold condition it takes the specific form

$F =$

$$\frac{L + r_j - \frac{1}{2} \sqrt{KV_{DS}} - \frac{1}{2} [KV_{DS} + 2r_j \sqrt{K(V_{bi} + V_{BS} + V_{DS})} + r_j^2]^{1/2} - \frac{1}{2} [2r_j \sqrt{K(V_{bi} + V_{BS})} + r_j^2]^{1/2}}{L - \sqrt{KV_{DS}}}, \quad (188)$$

where

$$K = 2\epsilon/qN_A,$$

$$r_j = \text{the junction depth,}$$

and

$$N_A = \text{the doping concentration}$$

and it is assuming that $V_{bi} \approx 2\phi_F$.

The transition from subthreshold to above-threshold conduction as a function of drain voltage is obvious in a short-channel device because the variation with drain-source voltage of the effective $\hat{\phi}_S$ (i.e., the average ϕ_S in the channel determined from two-dimensional charge-sharing techniques) below threshold and the V_T above threshold are

quite pronounced. Actually, this transition always exists even in a long-channel device since we may never have absolutely no channel length modulation except in the limit of a device of infinite length. Hence, there will always be some curve of I versus V_{DS} , which passes from the subthreshold to the above-threshold region at some V_{DS} , so that we may consider this condition in a general way. By imposing the condition of continuity of current, we obtain from (185) and (186) the relation

$$\frac{1}{2} \left[\eta^2 - \left(\frac{\bar{Q}}{C_o} \right)^2 \right] = \left(\frac{kT}{q} \right)^2 \frac{C_s}{C_o} \frac{q\xi}{e^{kT}}, \quad (189)$$

where C_s , as mentioned in the definition (182b), is the capacitance of the semiconductor depletion region. This capacitance will be approximately the same at any position between the source and the edge of the drain-depletion region (i.e., the saturation point in the channel) for gate voltages up to the threshold condition, since we are assuming a negligible field in the channel for the subthreshold formulation. The term ξ represents the additional surface potential that is required above threshold to achieve a matching of the solutions. We expect this increase in surface potential introduced in (186) to correspond to the concomitant increase in gate voltage η . The solutions are joined, therefore, not at $V_{GS} = V_T$ but at a slightly higher voltage, which is determined from (187) to be

$$V_T + \eta = V_{FB} + 2\phi_F + \xi + \frac{1}{C_o} \sqrt{2\epsilon q N_A (V_{BS} + 2\phi_F + \xi)} F + \frac{kT}{q} \frac{C_s}{C_o}, \quad (190)$$

where the term $[(kT)/q]/(C_s/C_o)$ is an attempt to represent the change in mobile charge in the channel between the two conditions. We have simply used $Q_n = N_A \bar{X}_d$, the charge in the channel under diffusion-limited conditions, where \bar{X}_d is the effective depth of channel charge.²⁵ We must now impose the condition of continuity of the derivatives at the transition. From (184) we have (neglecting the derivatives of \bar{Q})

$$\frac{dI}{dV_{DS}} = -\frac{I}{\bar{y}} \frac{d\bar{y}}{dV_{DS}} - \mu C_o \frac{W}{\bar{y}} \eta \frac{dV_T}{dV_{DS}}, \quad (191)$$

and from (186) we have

$$\frac{dI}{dV_{DS}} = -\frac{I}{\bar{y}} \frac{d\bar{y}}{dV_{DS}} + \frac{q}{kT} I \frac{d\hat{\phi}_s}{dV_{DS}}. \quad (192)$$

We therefore have at the transition from (191) and (192) that

$$\eta \frac{dV_T}{dV_{DS}} = - \left(\frac{kT}{q} \right) \frac{C_S}{C_o} e^{\frac{q\xi}{kT}} \frac{d\hat{\phi}_S}{dV_{DS}}. \quad (193)$$

From (187) we obtain for the derivative of the threshold voltage

$$\frac{dV_T}{dV_{DS}} = \frac{1}{C_o} \sqrt{2\epsilon q N_A (V_{BS} + 2\phi_F + \xi)} \frac{dF}{dV_{DS}} \quad (194)$$

or, for the surface potential at the transition

$$\left. \frac{d\hat{\phi}_S}{dV_{DS}} \right|_{\phi_S=2\phi_F+\xi} = - \frac{1}{(C_o + C_S F)} \sqrt{2\epsilon q N_A (V_{BS} + 2\phi_F + \xi)} \frac{dF}{dV_{DS}}. \quad (195)$$

Using (194) and (195) in (193) we obtain

$$\eta = \frac{kT}{q} \frac{C_S}{C_o + C_S F} e^{\frac{q\xi}{kT}}. \quad (196)$$

Using (190) and (187) at $V_G = V_T$, we find

$$\eta = \xi + \frac{1}{C_o} \sqrt{2\epsilon q N_A} \cdot (\sqrt{V_{BS} + 2\phi_F + \xi} - \sqrt{V_{BS} + 2\phi_F}) F + \frac{kT}{q} \frac{C_s}{C_o}, \quad (197)$$

and by expanding the square root term, since $\xi \ll V_{BS} + 2\phi_F$, we have

$$\eta \approx \frac{(C_o + C_S F)}{C_o} \xi + \frac{kT}{q} \frac{C_s}{C_o}. \quad (198)$$

Combining (196) and (198) we obtain

$$\xi = \frac{kT}{q} \frac{C_S C_o}{(C_o + C_S F)^2} e^{\frac{q\xi}{kT}} - \frac{kT}{q} \frac{C_s}{(C_o + C_S F)}$$

or

$$\xi' = \frac{C_S C_o}{(C_o + C_S F)^2} e^{\xi'} - \frac{C_s}{C_o + C_S F}, \quad (199)$$

where $\xi' = q\xi/kT$. We may, therefore, determine ξ once values of oxide thickness, doping density, and substrate bias have been specified by using this transcendental relationship. An interesting feature of this parameter is its lack of dependence upon T .

By this method we have ensured continuity of the current and of its

derivative in passing from the subthreshold to the above-threshold region. It is realized that the absolute value of the predicted current will be somewhat low around $V_G = V_T$ because the subthreshold region does not incorporate drift and because the above-threshold result does not incorporate diffusion. However, this is not expected to be of any serious consequence. To calculate the current for voltages between $V_T + \eta$ and V_T , the surface potential is found from (198) and then (186) is used.

The results of calculating ξ and η are shown in Figs. 25 and 26 as a function of oxide thickness and substrate doping over a wide range of these parameters. These parameters are independent of gate bias but have a slight dependence on V_{DS} through the parameter F . In Figs. 25 and 26 a value of $F = 0.8$ was used, and to illustrate the influence of the drain bias, the curves are also shown in dashed lines for $F = 0.6$. Generally, the results show that in the range of useful device operation (say $N_A \sim 1\text{--}2 \times 10^{16} \text{ cm}^{-3}$ for $t_{ox} \approx 500\text{\AA}$ or $N_A \approx 3\text{--}5 \times 10^{16} \text{ cm}^{-3}$ for $t_{ox} = 250\text{\AA}$), the parameters η and ξ are slowly decreasing functions of both doping and oxide thickness and may be readily determined from these simple calculations. For rather thick oxides and high doping levels, η and ξ tend to rise again, which shows that the subthreshold region is penetrating farther into the above-threshold region. One can

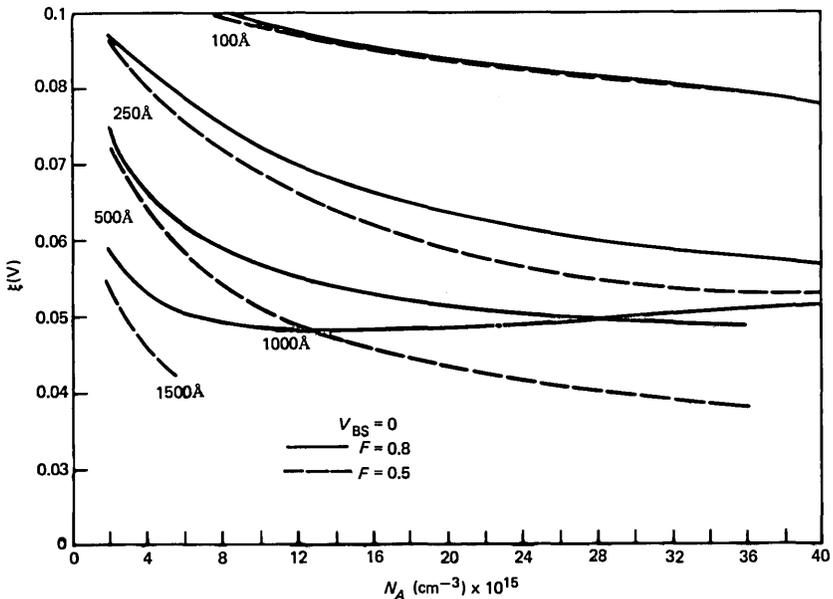


Fig. 25—Variation of ξ , the surface potential increment above threshold, as a function of substrate doping for several oxide thicknesses. The dashed curves show the effect of drain voltage through the short-channel factor F .

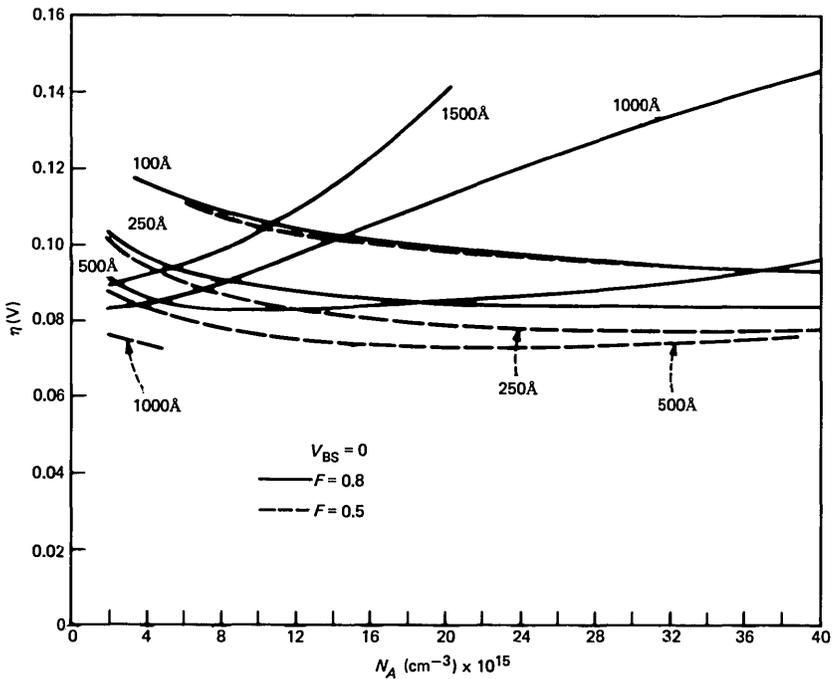


Fig. 26—Variation of η , the gate voltage above V_T for a merging of regions, as a function of substrate doping and oxide thickness. The variation with the factor F is shown by dashed lines.

think of η and ξ as offset voltage parameters that describe the protrusion of the subthreshold region into the above-threshold region.

We may now solve for \bar{Q}/C_o using (196) and (189) and obtain

$$\frac{\bar{Q}}{C_o} = \frac{kT}{q} \frac{(C_o + C_s F)}{C_o} \xi' \sqrt{1 - 2/\xi'}. \quad (200)$$

We may then use this result, together with (182a), to determine

$$\bar{R} = \frac{1}{1 + \xi' \sqrt{1 - 2/\xi'}}. \quad (201)$$

From (200) and (199) we conclude that \bar{Q}/C_o is approximately independent of bias parameters. The approximation is equivalent to ignoring the dependence of C_s upon \bar{V} . However, this is the same approximation we made to arrive at (191), so it is a consistent one. Within the same approximation, we see from (201) that \bar{R} is also weakly dependent upon bias parameters, another fact we have used to arrive at (191). We show later in the discussion how good the approximation is.

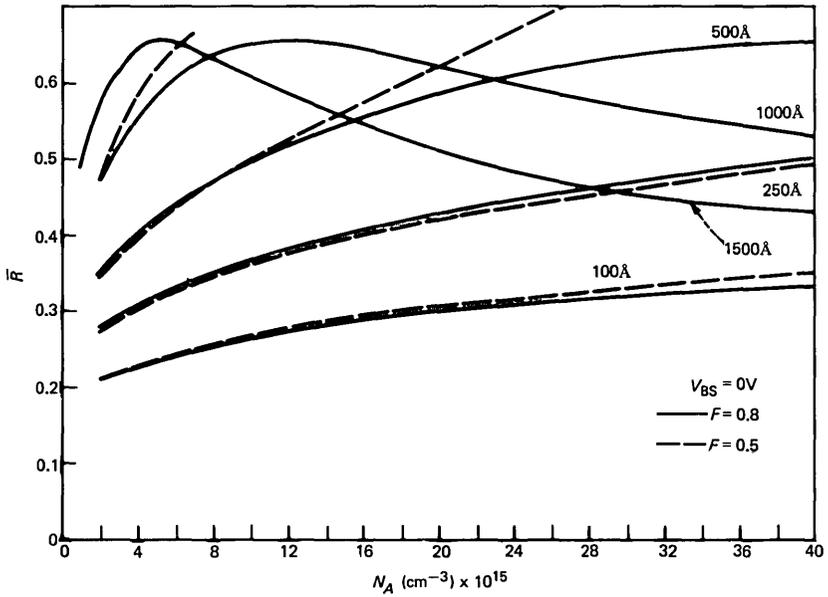


Fig. 27—Variation of \bar{R} , the fraction of diffusion at the pinch-off point as a function of N_A and t_{ox} for two values of F (and therefore V_{DS}).

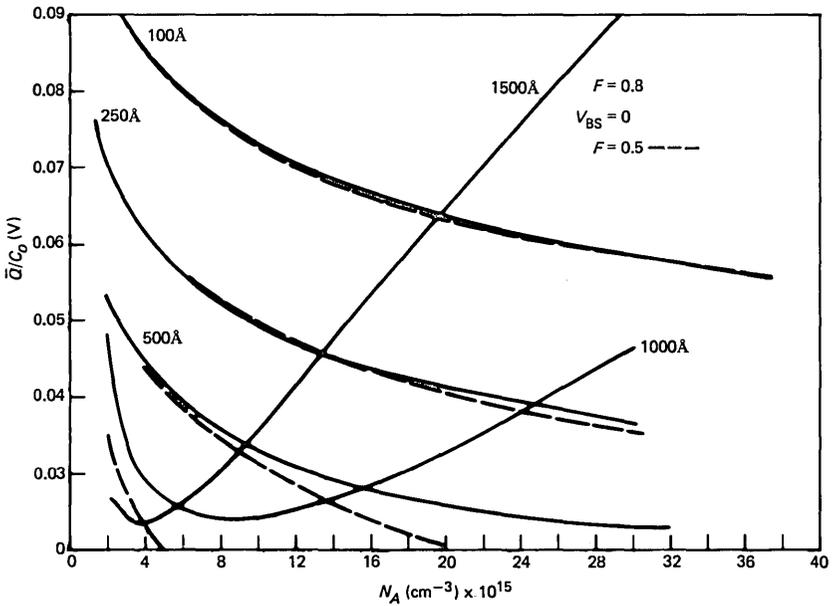


Fig. 28—Variation of \bar{Q} , the channel charge at the pinch-off point as a function of N_A and t_{ox} for two values of F .

The values of \bar{R} and \bar{Q} are plotted in Figs. 27 and 28 for substrate biases of $V_{BS} = 0, -3V$ as a function of substrate doping and oxide thickness. As for the other parameters, the effect of V_{DS} through the parameter F is not that great. Both \bar{R} and \bar{Q} are moderate functions of doping for a given oxide thickness. The value of \bar{R} ranges from 0.2 to 0.6, but typically for $t_{ox} = 500\text{\AA}$, $N_A = 10^{16}$, or $t_{ox} = 250\text{\AA}$, $N_A = 5 \times 10^{16}$, we find $\bar{R} \approx 0.4$, i.e., about one half of the current is drift and one half of the current is diffusion at the pinch-off point. As in the case of η and ξ for higher dopings and thick oxides, \bar{Q} increases considerably and \bar{R} drops off. Therefore, as the subthreshold transition point pushes higher above the threshold voltage, the pinch-off charge increases, and the fraction of the total current carried by drift tends to increase.

It should be noted that the parameter \bar{R} has been derived from conditions near threshold in which region velocity saturation of carriers can be ignored. The form of \bar{Q} [cf. (182b)] applies with or without velocity saturation; our interpretation of \bar{Q} is that the effects of velocity saturation enter only through T , which becomes a hot-electron temperature under velocity-saturated conditions. From (182b) this is equivalent to \bar{R} being independent of velocity saturation, and this is supported by (201) since ξ' is temperature-independent [cf. (199)]. It is, therefore, reasonable for \bar{R} to be used for all gate voltages even though it was derived from conditions at the transition.

AUTHOR

Geoffrey W. Taylor, Ph.D. (Electrical Engineering), 1972, University of Toronto; AT&T Bell Laboratories, 1979—. After two years of postdoctoral studies at the University of Toronto Mr. Taylor joined the Solid State Electronics Centre at Honeywell, where he worked on the n-channel charge-coupled devices (CCDs), dynamic RAMs, and MOS device development. In 1976 he joined the Central Research Lab at Texas Instruments, where he worked for one year on the development of large-area CCDs for visible imaging and for two years on the development of metal semiconductor field-effect transistors (MESFETs) MOSFETs, CCDs, and dynamic memory structures for VLSI applications. At AT&T Bell Laboratories he has been involved in NMOS circuit design and exploratory MOS device work. He is presently investigating new device concepts in Si and GaAs.

LETTERS

Comments on "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design," by M. S. Phadke, R. N. Kacker, D. V. Speeney, and M. J. Grieco*

I have recently read with extreme interest the article "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design," by M. S. Phadke et al. (hereafter denoted as Phadke).

The authors acknowledge the contributions of Professor Genichi Taguchi in this report. I believe my understanding of the Taguchi method is sufficient to raise the issues below. It is believed that the issues are sufficient so as to challenge the usefulness of the Taguchi Method if Phadke is used as a primer. If this should happen it is believed that a new and useful tool might be lost to the engineer.

The introduction to Phadke gives a concise presentation of the Taguchi Method. The problem arises when Phadke has to establish a signal-to-noise ratio (s/n) for analysis and selection of the optimal process by maximizing s/n . One must first understand the meaning of the 5 (experiments 5, 15, and 18) or 10 data points in Table III of Phadke. It is abundantly clear from the table headings (and the text) that these multiple measurements at each experimental condition fit into two categories:

1. They are the result of multiple measurements at different spots on a given wafer, and
2. They are the result of measurements on a second wafer processed at the same time as the first.

Phadke processes these 5 or 10 measurements to establish the mean pre-etch line width and the respective standard deviation listed in Table IV. It is at this point Phadke breaches the teachings of Taguchi. The standard deviation listed in Table IV is *not* the variance (square of the standard deviation) of which Taguchi speaks. The standard deviation values in Table IV are in fact due to at least four possible factors:

1. Experimental measurement error (treated only cursorily on page 1281).
2. Variations from wafer to wafer (undoubtedly small for adjacent wafers cut from a large boule). No information is given about the source of the wafers, except that they were scarce.
3. Variations at the five locations due to the effects of the mask variations, or other asymmetries in the wafer processing (e.g., the means of all 33 measurements of the pre-etch line width data—Table

* B.S.T.J., 62, No. 5 (May-June 1983), pp. 1273-309.

III—show that the “bottom” line width is distinctly different from the other four, i.e., its average value is 2.81 μm versus 2.62 to 2.69 for the other four; in fact, one is about 90-percent confident that the “bottom” data are different from the “left” data).

4. Variations in location due to the experimental process (i.e., experiment 2 versus experiment 9, etc.).

It is extremely clear that the signal-to-noise ($\eta = \log \bar{x}/s$) data in Table IV (page 1286) are not even remotely related to s/n that Taguchi teaches since the standard deviation is composed primarily of the effects of repeat measurements for one experiment. That is, the s/n in Table IV is simply another response, just as line width (\bar{x}) is a response. On the other hand, Taguchi’s noise (variance) is due to repeat experiments at the same or nearly the same experimental conditions (see the Wheatstone Bridge example in Ref. 3 of Phadke).

It is in the reduction to practicality where Phadke seemingly slips. The average s/n (η) in Table V is simply the appropriate average of the $s/n = \log \bar{x}/s$ in Table IV. For example, the average s/n for factor A level 1 is $(1.4803 + 1.3512 + \dots + 1.2709)/9 = 1.28568$ with a standard deviation of 0.13749. If one were interested in calculating the s/n ratio for Table V using the s/n of Table IV as a response variable, then the values would be those shown below for two of the 24 values that can be calculated (more will be said about the log transformation later).

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

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

Overall average s/n = 1.4011.

THIS VALUE SHOULD BE

$$\frac{1.2857}{0.1375} = 9.3510$$

BEFORE ANY TRANSFORMATION

$$\frac{1.51657}{0.23747} =$$

6.3864

In the case of the correct data in Table V (as per Jugle) the noise is due to fluctuations in the other eight process variables. *This* is the noise of which Taguchi speaks. However, the object (first paragraph of Section I in Phadke) of this experiment was to determine a process for which the line width was under control and its variance was minimal (i.e., the line width s/n ratio was a maximum). If this is the case, the data in Table V would look like Table Va below (note that only three values have been calculated and no transformation has been applied to the \bar{x}/s values).

One can now follow the Taguchi Method and ANOVA on the data shown in Table VII and Table Va to establish a process condition under which the appropriate process control variable can shift the mean pre-etch line width with little, or no, change in the variance and select the other process variable settings so as to minimize the variance (maximize the s/n). If one uses the data in Table V (as per Jugle) and the data in Table IV, one can determine a process variable, which controls the mean s/n across the wafer, and a set of process variables, which minimizes variance in this s/n value. The two conditions may, or may not, be the same, so the appropriate trade-off would have to be made.

There are two other comments which must be made. First, the log transformation of \bar{x}/s is of concern and it should be used only discriminately. Taguchi (as well as Phadke et al.) are products of the communications industry where \bar{x} is generally very large (10^3 through 10^{15}) and s is comparatively small (10^0 through 10^3). In this case the log transformation not only makes sense, but it is mandatory. However, in the Phadke experiment, and in many other experiments, $0.02\bar{x} \leq s \leq 0.20\bar{x}$. In this case the compression of larger values by the log transformation may deemphasize larger effects and might have contributed to the second comment below. Second, on page 1303 Phadke talks about "implementation and the benefits of optimum levels." It

Table Va—Pre-etch line width s/n

Factor	Level 1	Level 2	Level 3
A	5.448	6.804	—
BD	8.587	•	•
•	•	•	•
•	•	•	•
•	•	•	•

$$A_1 \quad \bar{x} = \frac{2.500 + 2.684 + \dots + 2.829}{9} = 2.487$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^9 (x_i - \bar{x})^2 = 0.4565$$

$$s/n = 2.487/.4565 = 5.448$$

is somewhat disturbing that after optimization by their analysis they had to change the control variable (Taguchi's signal factor) called exposure setting from a nominal position (90) to beyond the experimental range (140 versus 108, assuming a fixed aperture of 2). This would suggest that there exists either a major uncontrolled variable in the experiment or a strong unrecognized interaction controlling the results, and that the positive outcome (in terms of implementation and benefits) is largely fortuitous.

Finally, in the calculation for Table Va the mean value for the pre-etch width calculated by this author does not agree with the value published in Phadke in Table VII (i.e., 2.487 versus 2.39). The data in Table IV are self-consistent (x, s, η); hence one would expect that the x values are correct. This may be a transcription error. However, there are enough small errors in the calculated data I have checked (i.e., only about 5 percent of the calculations) to suggest some further review is necessary.

I would like to thank T. Barker and L. Smith (both of Xerox Corporation) for helpful discussions.

Don B. Jugle
Project Manager
Xerox Corporation

Reply to Letter by D. B. Jugle

Designing a process or a product that is robust against all noises is the key objective of off-line quality control. However, not only do we usually not know all noise factors, but even if we knew them it is inefficient and unnecessary to include all of them in an off-line quality control experiment. It is generally adequate to consider a few important noises in the experiment, with an anticipation that a design robust against the chosen noises will be robust against all noises. The confirmation experiment, then, verifies the robustness of the optimum design.

Signal-to-noise ratio (s/n) is a measure of process variation, which in turn is a measure of robustness. Depending on the problem, a different method is used to compute s/n . In the article by M. S. Phadke et al.* the s/n was calculated from the variation of the line width between wafers and within wafers. This s/n takes into consideration

* M. S. Phadke, R. N. Kacker, D. V. Speeney, and M. J. Grieco, "Off-Line Quality Control in Integrated Circuit Fabrication Using Experimental Design," B.S.T.J., 62, No. 5 (May-June 1983), pp. 1273-309.

the line width variation resulting from material variation between wafers, the nonflatness of a wafer, nonuniformity of projection printing across a wafer and from one exposure to another, etc. But, this s/n does not include the effect of normal variation in the process factors around their nominal values. Another way to compute s/n is to consider the data from all experiments at a particular level of a factor. For example, to calculate the s/n for level 1 of factor A we would use all observations on pre-etch line width corresponding to experiments 1 through 9. This s/n would then include the effect of variation in the other eight process parameters. However, the variation in the process parameters among experiments 1 through 9 is too wide compared to their normal variation. So, this s/n is also not perfect. The s/n used in our paper was selected on the basis of engineering judgment.

In his letter to the editor, Mr. Jugle has suggested in Table V (as per Jugle) that the s/n for various factor levels should be computed from the s/n for individual experiments. This should never be done because we are not interested in the variation of s/n . Also, the s/n suggested in Table Va (as per Jugle) is not appropriate because it ignores the linewidth variation within wafers and between wafers.

The main reason for taking the log of \bar{x}/s is that the factorial effects have better additivity in the log domain. For example, without the log transformation it is possible for the prediction of \bar{x}/s to be negative. This unrealistic prediction is avoided by the log transformation.

After publication of the article,* the authors realized that there were some typographical errors in the data. However, these errors have little impact on the final results. The presence of these errors is regretted.

The confirmation experiment reported in the article referenced showed without doubt that there was a four-fold reduction in the process variance and a three-fold reduction in the cases of unopened window. Also note that the off-line quality control method has been applied successfully in improving numerous processes in AT&T and elsewhere.

M. S. Phadke
R. N. Kacker
D. V. Speeney
M. J. Grieco
AT&T Bell Laboratories

G. Taguchi
Consultant

PAPERS BY AT&T BELL LABORATORIES AUTHORS

COMPUTING/MATHEMATICS

- Bentley J. L., **A Case-Study in Applied Algorithm Design.** *Computer* 17(2):75-88, 1984.
- Cargill T. A., **The Blit Debugger.** *J Syst Soft* 3(4):277-284, 1983.
- Christen C., Hwang F. K., **Detection of a Defective Coin With Partial Weight Information.** *Am Math Mo* 91(3):173-179, 1984.
- Chung F. R. K., Graham R. L., **Edge-Colored Complete Graphs With Precisely Colored Subgraphs.** *Combinatori* 3(3-4):315-324, 1983.
- Fishburn P. C., **Probabilities of Dominant Candidates Based on First-Place Votes.** *Discr App M* 7(2):131-140, 1984.
- Gale W. A., Koenker R., **Pricing Interactive Computer Services.** *Computer J* 27(1):8-17, 1984.
- Graham R. L., Yao F. F., **Finding the Convex Hull of a Simple Polygon.** *J Algorithm* 4(4):324-331, 1983.
- Hanson S. J., Kraut R. E., Farber J. M., **Interface Design and Multivariate Analysis of UNIX Command Use.** *ACM T Offic* 2(1):42-57, 1984.
- Hartigan J. A., Kleiner B., **A Mosaic Of Television Ratings.** *Am Statistn* 38(1):32-35, 1984.
- Johnson D. S., **The Np-Completeness Column—An Ongoing Guide.** *J Algorithm* 4(4):397-411, 1983.
- Logan B. F., **Integrals of High-Pass Functions.** *SIAM J Math* 15(2):389-405, 1984.
- Rose D. J., **Convergent Regular Splittings for Singular M-Matrices.** *SIAM J Alg* 5(1):133-144, 1984.
- Saksena V. R., Cruz J. B., **Robust Nash Strategies for a Class of Nonlinear Singularly Perturbed Problems.** *Int J Contr* 39(2):293-310, 1984.
- Togai M., **Japan's Next Generation of Robots.** *Computer* 17(3):19-25, 1984.
- Tullis T. S., **The Formatting of Alphanumeric Displays—A Review and Analysis.** *Human Fact* 25(6):657-682, 1983.

ENGINEERING

- Abidi A. A., Meyer R. G., **Noise in Relaxation Oscillators.** *IEEE J Soli* 18(6):794-802, 1983.
- Arroyo C. J., **A Helical-Wrap Peel Strength Tester.** *Adhes Age* 27(2):12-15, 1984.
- Aumiller G. D., **Infrared Dye-Laser in the 685- to 880-nm Range (Letter).** *Appl Optics* 23(5): 651, 1984.
- Banu M., Tsvividis Y., **Fully Integrated Active RC-Filters in MOS Technology.** *IEEE J Soli* 18(6):644-651, 1983.
- Burr D. J., Ackland B. D., Weste N., **Array Configurations for Dynamic Time Warping.** *IEEE Acoust* 32(1):119-128, 1984.
- Chraplyvy A. R., **Optical Power Limits in Multichannel Wavelength-Division-Multiplexed Systems Due to Stimulated Raman Scattering.** *Electr Lett* 20(2):58-59, 1984.
- Dautremont-Smith W. C., Lopata J., **Optical Monitoring for Rate and Uniformity Control of Low-Power Plasma-Enhanced CVD.** *J Vac Sci B Technol* 1(4):943-946, 1983.
- Ebeling K. J., Coldren L. A., **Wavelength Self-Stabilization of Coupled-Cavity Semiconductor Lasers.** *Electr Lett* 20(2):69-70, 1984.
- Elvidge J. et al., **Microprocessor Implementation of an FFT for Ionospheric VLF Observations (Letter).** *IEEE Geosci* 22(2):171-174, 1984.
- Fichtner W., Nagel L. W., Penumalli B. R., Petersen W. P., Darcy J. L., **The Impact of Supercomputers on IC Technology Development and Design.** *P IEEE* 72(1):96-112, 1984.

- Gehrels N. et al., **The Development of a Segmented N-Type Germanium Detector, and Its Application to Astronomical Gamma-Ray Spectroscopy.** *IEEE Nucl S* 31(1):307-311, 1984.
- Green M. L., Coleman E., Bader F. E., Sproles E. S., **The Physical Metallurgy and Electrical Contact Resistance of Mechanically Alloyed Cu-Ru Composites.** *Mater Sci E* 62(2):231-239, 1984.
- Haber F., Barness Y., Yeh C. C., **An Adaptive Interference Canceling Array Utilizing Hybrid Techniques.** *IEEE Aer El* 19(6):795-804, 1983.
- Hall P. M., Dudderar T. D., Argyle J. F., **Thermal Deformations Observed in Leadless Ceramic Chip Carriers Surface Mounted to Printed Wiring Boards.** *IEEE Compon* 6(4):544-552, 1983.
- Hayes J. R., Capasso F., Gossard A. C., Malik R. J., Wiegmann W., **Bipolar Transistor With Graded Band-Gap Base.** *Electr Lett* 19(11):410-411, 1983.
- Henein G., **Origin of Blistering Observed on Forming Gas-Annealed Ti-Pt Ohmic Contacts.** *Thin Sol Fi* 109(2):115-125, 1983.
- Kastalsky A., Luryi S., Gossard A. C., Hendel R., **A Field-Effect Transistor With a Negative Differential Resistance.** *IEEE Elec D* 5(2):57-60, 1984.
- Koren U., Arai S., Tien P. K., **Three-Channel Buried-Crescent InGaAsP Laser With 1.51- μ m Wavelength on Semi-Insulating InP.** *Electr Lett* 20(4):177-178, 1984.
- Kramer B. et al., **Data Acquisition, Reduction, and Transmission of Atmospheric Electricity Data on High-Altitude Scientific Balloons (Letter).** *IEEE Geosci* 22(2):169-171, 1984.
- Levine B. F., Bethea C. G., **Ten-MHz Single Photon Counting at 1.3 μ m.** *Appl Phys L* 44(6):581-582, 1984.
- Mucha J. A., **Correction of Nonlinear Derivative Diode-Laser Data in Standard Addition Analyses.** *Appl Spectr* 38(1):68-73, 1984.
- Novembre A. E., Bowden M. J., **Effect of Varying the Composition of Copolymers of Glycidyl Methacrylate and 3-Chlorostyrene (GMC) on Electron Lithographic Performance.** *Polym Eng S* 23(17):975-979, 1983.
- Olsson N. A., Dutta N. K., Liou K. Y., **Dynamic Linewidth of Amplitude-Modulated Single-Longitudinal-Mode Semiconductor-Lasers Operating At 1.5 μ m Wavelength.** *Electr Lett* 20(3):121-122, 1984.
- Olsson N. A., Dutta N. K., Tsang W. T., Logan R. A., **Threshold Current Characteristics of GaAs Lasers Under Short Pulse Excitation.** *Electr Lett* 20(2): 63-64, 1984.
- Ralls K. S., Skocpol W. J., Jackel L. D., Howard R. E., Fetter L. A., Epworth R. W., Tennant D. M., **Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1-f?) Noise.** *Phys Rev L* 52(3):228-231, 1984.
- Russ G. J., **A Comparative Study of Electroplated Palladium as a Contact Finish.** *IEEE Compon* 6(4):389-395, 1983.
- Sharma S. P., Dasgupta S., **Reaction of Contact Materials With Vapors Emanating From Connector Products.** *IEEE Compon* 6(4):553-559, 1983.
- Smith P. W., Tomlinson W. J., **Nonlinear Optical Interfaces—Switching Behavior.** *IEEE J Q El* 20(1):30-36, 1984.
- Song B. S., Gray P. R., **A Precision Curvature-Compensated CMOS Bandgap Reference.** *IEEE J Soli* 18(6):634-643, 1983.
- Sundahl R. C., **Electrooptic Materials for Optical Communications.** *Ferroelectr* 50(1-4):457-466, 1983.
- Taylor G. N., Wolf T. M., Stillwagon L. E., **The Role of Inorganic Materials in Dry-Processed Resist Technology.** *Sol St Tech* 27(2):145-155, 1984.
- Thurston R. N., Boyd G. D., **Brightness and Contrast of Nematic Liquid-Crystal Bistable Configuration Displays.** *Displays* 5(1):15-20, 1984.
- Vasile M. J., **The Velocity Dependence of Secondary Ion Yields.** *Nucl Instru* 218(1-3):319-323, 1983.
- Wagner A., **Applications of Focused Ion Beams.** *Nucl Instru* 218(1-3):355-362, 1983.
- Winters J. H., **Differential Detection With Intersymbol Interference and Frequency Uncertainty.** *IEEE Commun* 32(1):25-33, 1984.

Xydeas C. S., Kostic B., Steele R., **Embedding Data Into Pictures by Modulo Masking.** IEEE Commun 32(1):56-69, 1984.

MANAGEMENT/ECONOMICS

Graham D. R., Kaplan D. P., Sibley D. S., **Efficiency and Competition in the Airline Industry.** Bell J Econ 14(1):118-138, 1983.

Porter R. H., **A Study of Cartel Stability—The Joint Executive Committee, 1880-1886.** Bell J Econ 14(2):301-314, 1983.

PHYSICAL SCIENCES

Abrahams S. C., Bair H. E., DiSalvo F. J., Marsh P., Deuring L. A., **Magnetic and Structural Phase-Transition at 363.7-K in Tetramethylammonium Hexacyano-trimethylenecyclopropanide.** Phys Rev B 29(3):1258-1262, 1984.

Abrahams S. C., Marsh P., Liminga R., Lundgren J. O., **Pyroelectric Sr(NO₂)₂·H₂O: Room Temperature Crystal Structure.** J Chem Phys 79(12):6237-6241, 1983.

Aeppli G. et al., **Spin Correlations and Reentrant Spin-Glass Behavior in Amorphous Fe-Mn Alloys. 2. Dynamics.** Phys Rev B 29(5):2589-2605, 1984.

Aeppli G., Guggenheim H., Uemura Y. J., **Spin Dynamics Near the Magnetic Percolation Threshold.** Phys Rev L 52(11):942-945, 1984.

Aur S., Kofalt D., Waseda Y., Egami T., Chen H. S., Teo B. K., Wang R., **Local Atomic-Structure in Amorphous Mo₅₀Ni₅₀ by Resonance X-Ray Diffraction Using Synchrotron Radiation.** J Non-Cryst 61-2(Jan):331-336, 1984.

Benedek G., Toennies J. P., Doak R. B., **Surface-Phonon Spectroscopy of LiF(001) by Inelastic Scattering of He Atoms: Theory and Interpretation of Time-of-Flight Spectra.** Phys Rev B Condensed Matter 28(12):7277-7287, 1983.

Berreman D. W., **Domain-Wall Tension Allows Bistability in Imperfect Lamellar Cholesteric Twist Cells.** J Appl Phys 55(4):806-809, 1984.

Bertz S. H., **The Role of Symmetry in Synthetic Analysis—The Concept of Reflexivity.** J Chem S Ch (4):218-219, 1984.

Besomi P., Degani J., Wilson R. B., **Heat-Treatment Effects on Indium Gallium-Arsenide Phosphide Double Heterostructure Material.** J Appl Phys 55(4):1135-1138, 1984.

Broek H. W., Tom C., **Spatial Coherence of Multipath Underwater Acoustic Transmission—An Experiment At Two Frequencies Concurrently.** J Acoust So 75(2):395-405, 1984.

Bruinsma R., Aeppli G., **Metamagnets and Frustration.** Phys Rev B 29(5):2644-2651, 1984.

Capizzi M., Modesti S., Frova A., Staehli J. L., Guzzi M., Logan R. A., **Electron-Hole Plasma in Direct-Gap Ga_{1-x}Al_xAs and K-Selection Rule.** Phys Rev B 29(4):2028-2035, 1984.

Chen H. S., Inoue A., **Sub-Tg Enthalpy Relaxation In PdNiSi Alloy Glasses.** J Non-Cryst 61-2(Jan):805-810, 1984.

Coldren L. A., Ebeling K. J., Rentschler J. A., Burrus C. A., Wilt D. P., **Continuous Operation of Monolithic Dynamic Single-Mode Coupled-Cavity Lasers.** Appl Phys L 44(4):368-370, 1984.

Dean P. J., Thomas D. G., Frosch C. J., **New Isoelectronic Trap Luminescence in Gallium-Phosphide.** J Phys C 17(4):747-762, 1984.

Denbroeder F. J., Vandenberg J. M., Draper C. W., **Microstructures of Cu-Zr Phases Formed by Laser Surface Treatment.** Thin Sol Fi 111(1):43-51, 1984.

Di Mauro L. F., Heaven M., Miller T. A., **Laser-Induced Fluorescence Spectroscopy of Ionic Clusters Between Organic Cations and Inert Gases.** Chem P Lett 104(6):526-532, 1984.

Dinnsen D. A. et al., **Phonological Neutralization, Phonetic Implementation and Individual Differences.** J Phonetics 12(1):49-60, 1984.

Donnelly V. M., Flamm D. L., Dautremont-Smith W. C., Werder D. J., **Anisotropic**

- Etching of SiO₂ in Low-Frequency CF₄/O₂ and NF₃/Ar Plasmas.** *J Appl Phys* 55(1):242-252, 1984.
- Dutta N. K., Wilt D. P., Besomi P., Dautremont-Smith W. C., Wright P. D., Nelson R. J., **Improved Linearity and Kink Criteria for 1.3- μ m InGaAsP-InP Channelled Substrate Buried Heterostructure Lasers.** *Appl Phys L* 44(5):483-485, 1984.
- Eergen J. R., Julesz B., **Rapid Discrimination of Visual Patterns.** *IEEE Syst M* 13(5):857-863, 1983.
- Eibschutz M., Lines M. E., Chen H. S., **Linewidth Asymmetries in the Mossbauer Zeeman Spectrum of Amorphous Iron-Metalloid Alloys.** *J Non-Cryst* 61-2 (Jan):1219-1224, 1984.
- Elman B. S. et al., **Stoichiometric Determination of SbCl₅-Graphite Intercalation Compounds Using Rutherford Backscattering Spectrometry.** *J Appl Phys* 55(4):894-898, 1984.
- Fleury P. A., Lyons K. B., **Dynamic Central Peaks in Solids—Progress and Prospects.** *Ferroelectr* 52(1-3):3-14, 1983.
- Gottscho R. A., Miller T. A., **Optical Techniques in Plasma Diagnostics.** *Pur A Chem* 56(2):189-208, 1984.
- Haavasoja T., Narayanamurti V., Chin M. A., Bhatt R. N., **Observation of a High-Frequency Cutoff for Phonon Propagation in Liquid ⁴He.** *Phys Rev L* 51(26):2400-2409, 1983.
- Harbison J. P., Williams A. J., Lang D. V., **Effect of Silane Dilution on Intrinsic Stress in Glow-Discharge Hydrogenated Amorphous-Silicon Films.** *J Appl Phys* 55(4):946-951, 1984.
- Holmes R. J., Kim Y. S., Brandle C. D., Smyth D. M., **Evaluation of Crystals of LiNbO₃ Doped With MgO or TiC₂ for Electrooptic Devices.** *Ferroelectr* 51(1-2):41-45, 1983.
- Hopkins J. B., Farrow L. A., Fisanick G. J., **Raman Microprobe Determination of Local Crystal Orientation in Laser Annealed Silicon.** *Appl Phys L* 44(5):535-537, 1984.
- Huse D. A., **Multicritical Scaling in Baxter's Hard Square Lattice Gas.** *J Phys A* 16(18):4357-4368, 1983.
- Jackel J. L., Rice C. E., Veselka J. J., **Proton Exchange in LiNbO₃.** *Ferroelectr* 50(1-4):491-496, 1983.
- Johnson A. M., Glass A. M., Olson D. H., Simpson W. M., Harbison J. P., **High Quantum Efficiency Amorphous-Silicon Photodetectors With Picosecond Response Times.** *Appl Phys L* 44(4):450-452, 1984.
- Julesz B., Harmon L. D., **Noise and Recognizability of Coarse Quantized Images (Letter).** *Nature* 308(5955):211-212, 1984.
- Kahn D., Rabiner L. R., Rosenberg A. E., **On Duration and Smoothing Rules in a Demissyllable-Based Isolated-Word Recognition System.** *J Acoust So* 75(2):590-598, 1984.
- Keith H. D., Loomis T. C., **Estimation of Crystallization Temperatures in Quenched Polyethylene.** *J Pol Sc PP* 22(2):295-306, 1984.
- Kihlstrom K. E., Mael D., Geballe T. H., **Tunneling $\alpha^2F(\omega)$ and Heat-Capacity Measurements in High-T_c Nb₃Ge.** *Phys Rev B Condensed Matter* 29(1):150-158, 1984.
- Levine B. F., Bethea C. G., **Single Photon Detection at 1.3 μ m Using a Gated Avalanche Photodiode.** *Appl Phys L* 44(5):553-555, 1984.
- Liu P. L. et al., **Measurements of Intensity Fluctuations of an InGaAsP External Cavity Laser.** *Appl Phys L* 44(5):481-483, 1984.
- Lovinger A. J. et al., **Curie Transitions in Copolymers of Vinylidene Fluoride.** *Ferroelectr* 50(1-4):553-562, 1983.
- Macrander A. T., Chu S. N. G., Strege K. E., Bloemeke A. F., Johnston W. D., **Correlation Between Background Carrier Concentration and X-Ray Line-width for InGaAs/InP Grown by Vapor-Phase Epitaxy.** *Appl Phys L* 44(6):615-617, 1984.
- Maloney T. J., Aspnes D. E., Arwin H., Sigmon T. W., **Spectroscopic Ellipsometric and He Backscattering Analyses of Crystalline Si-SiO₂ Mixtures Grown by Molecular-Beam Epitaxy.** *Appl Phys L* 44(5):517-519, 1984.

- Mankiewich P. M., Craighead H. G., Harrison T. R., Dayem A. H., **High-Resolution Electron-Beam Lithography on CaF_2** . *Appl Phys L* 44(4):468-469, 1984.
- Martinez O. E., Heritage J. P., Miller B. I., Dutta N. K., Nelson R. J., **Threshold Temperature Dependence of Subnanosecond Optically-Excited 1.3- μm In-GaAsP Lasers**. *Appl Phys L* 44(6):578-580, 1984.
- Miller R. C., Dupuis R. D., Petroff P. M., **High-Quality Single GaAs Quantum Wells Grown by Metalorganic Chemical Vapor Deposition**. *Appl Phys L* 44(5):508-510, 1984.
- Mitchell J. W., Kessler J. E., **Purification of Optical Waveguide Glass Forming Reagents—Phosphorus Oxychloride**. *J Elchem Sc* 131(2):361-365, 1984.
- Moller A., Nordheim A., Kozlowski S. A., Patel D. J., Rich A., **Bromination Stabilizes Poly(dG-dC) in the Z-DNA Form Under Low-Salt Conditions**. *Biochem* 23(1):54-62, 1984.
- Ng K. K., Polito W. J., Ligenza J. R., **Growth-Kinetics of Thin Silicon Dioxide in a Controlled Ambient Oxidation System**. *Appl Phys L* 44(6):626-628, 1984.
- Olsson N. A., Tsang W. T., Logan R. A., Kaminow I. P., Ko J. S., **Spectral Bistability in Coupled Cavity Semiconductor Lasers**. *Appl Phys L* 44(4):375-377, 1984.
- Pelleg J., **Diffusion of Phosphorus in TaSi_2 Thin-Films. 1. Lattice and Grain-Boundary Diffusion in TaSi_2/Si (Polycrystalline)**. *Thin Sol Fi* 110(2):115-127, 1983.
- Pelleg J., **Diffusion of Phosphorus in TaSi_2 Thin Films. 2. Lattice and Short-Circuit Diffusion in $\text{TaSi}_2/\text{SiO}_2$** . *Thin Sol Fi* 110(2):129-138, 1983.
- Phillips J. C., **Chemical Bonding and Heats of Formation in Chalcogenide Network Compounds $\text{As}_2(\text{S},\text{Se})_3$ and $\text{Ge}(\text{S},\text{Se})_2$** . *Phys Rev B Condensed Matter* 28(12):7038-7039, 1983.
- Pini R. et al., **Continuously Tunable Multiple-Order Stimulated 4-Photon Mixing in a Multimode Silica Fiber**. *Optics Lett* 9(3):79-81, 1984.
- Raghavan R. S., **Growing Up Recoil-Free in India**. *Phys Today* 37(2):38-44, 1984.
- Rossetti R., Beck S. M., Brus L. E., **Direct Observation of Charge-Transfer Reactions Across Semiconductor—Aqueous-Solution Interfaces Using Transient Raman Spectroscopy**. *J Am Chem S* 106(4):980-984, 1984.
- Schweizer K. S., Stillinger F. H., **High-Pressure Phase-Transitions and Hydrogen-Bond Symmetry in Ice Polymorphs**. *J Chem Phys* 80(3):1230-1240, 1984.
- Silberg E., Chang T. Y., Ballman A. A., Caridi E. A., **Doping and Electrical Properties of Mn in $\text{In}_{1-x}\text{y Ga}_x\text{Al}_y$ as Grown by Molecular Beam Epitaxy**. *J Appl Phys* 54(12):6974-6981, 1983.
- Stavola M., **Infrared Spectrum of Interstitial Oxygen in Silicon**. *Appl Phys L* 44(5):514-516, 1984.
- Talpey T. E., Worley R. D., **Infrasonic Ambient Noise Measurements in Deep Atlantic Water (Letter)**. *J Acoust So* 75(2):621-622, 1984.
- Targ S. S. et al., **Use of a Diode Laser to Observe Room-Temperature, Low-Power Optical Bistability in a GaAs-AlGaAs Etalon**. *Appl Phys L* 44(4):360-361, 1984.
- Tell B. et al., **Beryllium Implantation Doping of InGaAs**. *Appl Phys L* 44(4):438-440, 1984.
- Temkin H., Dupuis R. D., Logan R. A., Van Der Ziel J. P., **Schottky-Barrier Restricted Arrays of Phase-Coupled AlGaAs Quantum-Well Lasers**. *Appl Phys L* 44(5):473-475, 1984.
- Teo B. K., Chen H. S., Wang R., Antonio M. R., **EXAFS of Glassy Metallic Alloys: Amorphous and Crystalline MoNi**. *J Non-Cryst Solids* 58(2-3):249-274, 1983.
- Van Der Ziel J. P., Temkin H., Dupuis E. D., Mikulyak R. M., **Mode-Locked Picosecond Pulse Generation From High-Power Phase-Locked GaAs-Laser Arrays**. *Appl Phys L* 44(4):357-359, 1984.
- Warren W. W., Elhanany U., Brenner G. F., **NMR Studies of Expanded Liquid Cesium**. *J Non-Cryst* 61-2(Jan):23-28, 1983.
- Weston H. T., Daniels W. B., **Temperature and Volume Dependence of the Thermal-Conductivity of Solid Neon**. *Phys Rev B* 29(5):2709-2716, 1984.
- Wood T. H., **Actual Modal Power Distributions in Multimode Optical Fibers and Their Effect on Modal Noise**. *Optics Lett* 9(3):102-104, 1984.

SOCIAL AND LIFE SCIENCES

Poltrock S. E., Schwartz D. R., Comparative Judgments of Multidigit Numbers. J Exp Psy L 10(1):32-45, 1984.

ERRATUM

A. G. Fraser and S. P. Morgan, "Queueing and Framing Disciplines for a Mixture of Data Traffic Types," AT&T Bell Lab. Tech. J., 63, No. 6, Part 2 (July-August 1984), pp. 1061-87.

The illustrations in Figures 6, 7, and 8 in this paper were misplaced. In all cases the captions are correct, but the illustrations are in the wrong positions. The illustration captioned Figure 6 (page 1074) should be Figure 7, Figure 7 (page 1075) should be Figure 8, and Figure 8 (page 1077) should be Figure 6.

CONTENTS, OCTOBER 1984

Part 1

- Contrasting Performance of Faster Binary Signaling With QAM
G. J. Foschini
- Adaptive Transversal Equalization of Multipath Propagation for
16-QAM 90-Mb/s Digital Radio
G. L. Fenderson, J. W. Parker, P. D. Quigley, S. R. Shepard,
and C. A. Siller, Jr.
- Enhancement of ADPCM Speech by Adaptive Postfiltering
V. Ramamoorthy and N. S. Jayant
- On Using the Itakura-Saito Measures for Speech Coder Performance
Evaluation
B.-H. Juang
- A Packet/Circuit Switch
Z. L. Budrikis and A. N. Netravali
- An Approximate Analysis of Sojourn Times in the M/G/1 Queue
With Round Robin Service Discipline
P. J. Fleming
- Analysis of a TDMA Network With Voice and Data Traffic
M. L. Honig

Part 2

THE *UNIX* SYSTEM

- Preface
R. L. Martin
- Foreword
A. V. Aho
- The Evolution of the *UNIX* Time-sharing System
D. M. Ritchie
- Program Design in the *UNIX* System Environment
R. Pike and B. W. Kernighan
- The Blit: A Multiplexed Graphics Terminal
R. Pike
- Debugging C Programs With the Blit
T. Cargill

- UNIX* Operating System Security
F. T. Grampp and R. H. Morris
- File Security and the *UNIX* System Crypt Command
J. A. Reeds and P. J. Weinberger
- The Evolution of C—Past and Future
L. Rosler
- Data Abstraction in C
B. Stroustrup
- Multiprocessor *UNIX* Systems
M. J. Bach and S. J. Buroff
- A *UNIX* System Implementation for System/370
W. A. Felton, G. L. Miller, and J. M. Milner
- UNIX* Operating System Porting Experiences
D. E. Bodenstab, T. F. Houghton, K. A. Kelleman, G. Ronkin,
and E. P. Schan
- Evolution of *UNIX* System Performance
J. Feder
- Cheap Dynamic Instruction Counting
P. J. Weinberger
- Theory and Practice in the Construction of a Working Sort Routine
J. P. Linderman
- The Fair Share Scheduler
G. Henry
- The Virtual Protocol Machine
M. J. Fitton, C. J. Harkness, R. A. Kelleman, P. F. Long, and
C. Mee
- Large Software Development Environments Using the *UNIX*
System
T. E. Fritz, J. E. Hefner, and T. M. Raleigh
- A Stream Input-Output System
D. M. Ritchie

AT&T BELL LABORATORIES TECHNICAL JOURNAL is abstracted or indexed by *Abstract Journal in Earthquake Engineering*, *Applied Mechanics Review*, *Applied Science & Technology Index*, *Chemical Abstracts*, *Computer Abstracts*, *Current Contents/Engineering, Technology & Applied Sciences*, *Current Index to Statistics*, *Current Papers in Electrical & Electronic Engineering*, *Current Papers on Computers & Control*, *Electronics & Communications Abstracts Journal*, *The Engineering Index*, *International Aerospace Abstracts*, *Journal of Current Laser Abstracts*, *Language and Language Behavior Abstracts*, *Mathematical Reviews*, *Science Abstracts (Series A, Physics Abstracts; Series B, Electrical and Electronic Abstracts; and Series C, Computer & Control Abstracts)*, *Science Citation Index*, *Sociological Abstracts*, *Social Welfare, Social Planning and Social Development*, and *Solid State Abstracts Journal*. Reproductions of the Journal by years are available in microform from University Microfilms, 300 N. Zeeb Road, Ann Arbor, Michigan 48106.



AT&T
Bell Laboratories

