

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING
ASPECTS OF ELECTRICAL COMMUNICATION

Volume 58

September 1979

Number 7

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

The EL2 Electret Transmitter: Analytical Modeling, Optimization, and Design

By J. C. BAUMHAUER, Jr. and A. M. BRZEZINSKI

(Manuscript received February 21, 1979)

We describe here the development of an EL2 Electret Transmitter that provides desirable attributes for use in hands-free-answer telephony and future electronic and special-purpose sets. The EL2 has lower sensitivity to spurious electromagnetic and mechanical signals than do existing magnetic transmitters. It offers lower dc power consumption, smaller size, and lower intrinsic noise and distortion than the carbon microphone. Formulation of an electro-mechano-acoustic electret model allows parameter optimization, in which side conditions on electrostatic stability and a prescribed transmit frequency response are adhered to. We show that higher sensitivities are possible with larger air film thickness, until the decreasing source capacitance becomes a limiting factor. Multiple diaphragm supports allow decreased film stress with no change in the stable electret charge or sensitivity. We describe theoretically a thermal stabilization procedure that minimizes long-range stress relaxation effects by accelerating viscoelastic changes. Based on film data, we project nominal sensitivity variations within ± 1 dB over 20 years of service. In the design, metallized poly(tetrafluoroethylene) electret film is tensioned, supported, and clamped above a selectively metallized stationary electrode forming three cells acoustically and electrically in parallel. A preamplifier completes the subassembly, which is housed in a rectangular aluminum enclosure shielding the transducer. Typical EL2 parameters are -32 dBV/N/m² sensitivity at 1 kHz, 1 k Ω output impedance, 3.2 kHz response resonance frequency, and 2 to 16 V required dc supply.

I. INTRODUCTION

Electromagnetic transducers based on the modulation of a biasing magnetic field in an air film appeared as microphones (often referred to as transmitters in telephony), receivers, and loudspeakers very early in the development of telephony.^{1,2} Today, this mechanism is employed in general-purpose U- and L-type telephone receivers in the Bell System. A smaller electromagnetic transducer, the AF1, was used as the electroacoustic transmitter in speakerphone modules until mid-1978. However, because of its operating principle, in recent years it was recognized to be inherently susceptible to an increasing incidence of spurious electromagnetic signals at customer locations. At the same time, interest was growing in a new telephone transmitter for electronic and special-purpose residential sets as well as for hands-free-answer features in business sets. The former residential applications generally require lower dc power consumption and smaller size than are typical of the variable-resistance granular carbon transmitter³ such as the T-type used in general-purpose sets. Moreover, the carbon transmitter, being dependent upon periodic mechanical agitation, is not suitable as the stationary transmitter in hands-free-answer applications. These events encouraged Bell Laboratories development efforts⁴ and the subsequent design of the EL2 Electret Transmitter. Compared to the AF1, the electret promised a greatly reduced sensitivity to electromagnetic as well as mechanical vibration interference. Its signal power efficiency is surpassed only by the carbon transmitter. However, compared to the latter, the electret could be used in applications where lower bias power consumption, smaller size, lower intrinsic noise and distortion, and the potential for longer service life were required.

A variable-capacitance transducer is based on the modulation of a biasing electrostatic field in an air film. In an electret, that field is provided by an "electret charge" distribution⁵ typically implanted in, and near the surface of, a thin, solid, dielectric film abutting the air film. The externally biased condenser microphone was first demonstrated by A. E. Dolbear⁶ in 1878, although E. C. Wente⁷ was the first to develop a practical instrument in 1917 at the AT&T Research Laboratories. In 1928, S. Nishikawa and D. Nukiyama used a thick, wax plate electret element in building an early electret transducer.⁸ In the United States, R. T. Rutherford was granted an electret microphone patent in 1935.⁸ While the Japanese used wax electrets during World War II, their extremely low capacitance and unstable electret charge retention remained a problem.⁸ In 1962, Sessler and West developed the electret-biased, polymer-film transducer^{9,10} at Bell Laboratories. Made with charged thin Teflon† films, these microphones

† Registered trademark of E. I. DuPont de Nemours.

overcame the earlier limitations of self-biased units. Through the 1960s, large investments in the development and design of polymer-film electret microphones were made, largely outside telephony. Bell Canada introduced their electret transmitter design into an operator headset product in 1970¹¹ and later into a speakerphone. Recent efforts at Bell Laboratories culminated in initial Western Electric production shipments of the Bell System EL2 electret transmitter¹² in May 1977. It has now replaced the AF1 in the 4A Speakerphone and is being used as the stationary transmitter in a number of business sets for hands-free-answer services.

This paper covers the transducer device aspects of the EL2 development and design, while a companion article¹³ will treat new technological aspects of the project. Here, analytical modeling and parameter optimization are first presented (Section II). Those results, coupled with physical, electrical, material, and telephone constraints are then shown applied in the physical design (Section III) to achieve performance and reliability objectives (Section IV).

II. ANALYTICAL MODELING AND OPTIMIZATION†

2.1 Transducer physics and model

Capacitance microphones are based on the modulation of a biasing dc electric field in a gas dielectric; the field in turn modulates the induced surface charge on an adjacent electrode. In the electret microphone (see Fig. 1), the biasing field in the air film is the result of a fixed "electret" surface charge¹⁴ which may be considered trapped on an imaginary electrode at the surface of the polymer film adjacent the air film. Free charge is thus induced in the "stationary" electrode. Assuming a homogeneous isotropic linear solid dielectric and a one-dimensional electric field normal to the stationary electrode, an application of (i) the charge equation of electrostatics, (ii) Gauss's equation on the line integral of electric field, and (iii) jump conditions on the continuity of electric displacement¹⁵ may be shown¹⁶ to yield the electric field in the air film and the induced free surface charge on the stationary electrode as, respectively,

$$\begin{aligned} E_o^1 &= (V_E + \sigma d/\epsilon)/(h^1 + d\epsilon_o/\epsilon), \\ \sigma_B^1 &= \epsilon_o E_o^1. \end{aligned} \quad (1)$$

Here, ϵ and ϵ_o are the dielectric permittivity of the polymer and air films, respectively, $-\sigma$ is the negative electret surface charge per unit area, d and h^1 are the polymer and air film thickness, respectively, and an externally applied dc voltage V_E across the stationary and moving

† See appendix for partial list of symbols.

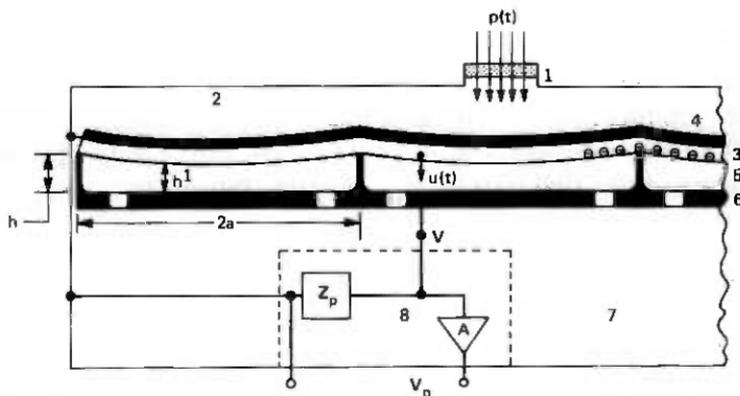


Fig. 1—Multicell electret transmitter: 1. Sound port and impedance material; 2. Front acoustic chamber; 3. "Electret-charged" polymer film diaphragm; 4. Diaphragm metalized electrode; 5. Air film; 6. Stationary electrode with holes; 7. Rear acoustic chamber; 8. Preamplifier.

diaphragm electrodes has been assumed presently, in addition to the electret self-bias. Meter-kilogram-second units have been employed. The electrostatic traction, T , exerted on the polymer film diaphragm is given by the normal surface component of the Maxwell electrostatic stress tensor¹⁷ in the air film, i.e.,

$$T = \epsilon_0 (E_o^1)^2 / 2. \quad (2)$$

Since T , together with the polymer film membrane tension, will determine the biased diaphragm equilibrium displacement, it is clear from eqs. (2) and (1a) that electrostatically an electret bias charge σ of magnitude such that

$$\sigma d / \epsilon = V_E \quad (3)$$

is equivalent to an external voltage V_E present in a nonelectret capacitance (condenser) microphone. This equivalence, previously shown by Warren et al.¹⁸ by employing energy considerations, can likewise be shown to hold for dynamic operation as first discovered by Sessler.¹⁴ Henceforth, only a passive electret self-bias will be assumed present; that is, we take $V_E \equiv 0$ in eq. (1a). Of course, all results can easily be applied to capacitance microphones using the equivalence (3). It is noted that quantities representing the biasing electrostatic equilibrium (or intermediate) state are being designated with the superscript "1."

In Fig. 1, a multicell electret diaphragm supported by ribs, spaced at distance $2a$ and each of length $2b$, is shown in its equilibrium position. Identical individual cells operating acoustically and electrically in parallel are formed. A membrane diaphragm's effective "lumped parameter" mass, stiffness, and area (all per unit cell) may be written,

respectively,

$$M_e = \kappa \rho d a^2, \quad K_e = \gamma S, \quad A_e = \beta a^2. \quad (4)$$

Here, ρ is the film mass per unit volume and S is the applied membrane force per unit edge, hereafter referred to as tension. For a rectangular diaphragm used in a capacitive transduction device, the constants may be shown to be appropriately based upon the average membrane displacement and are given by¹⁶

$$\kappa = \pi^4 \xi / 16 \lambda, \quad \gamma = \pi^2 \kappa / 4, \quad \beta = 4 \xi / \lambda, \quad (5)$$

where ξ is the number of cells formed and $\lambda = \xi 2a / 2b$ is the prescribed overall diaphragm aspect ratio; the prescribed overall diaphragm area is then ξA_c . Relations (5) were derived assuming one-dimensional tension in the horizontal direction in Fig. 1, and with zero displacement at all boundaries surrounding each unit cell. Referring to the figure, the electrostatic displacement, w , of the effective rigid piston model is defined by

$$h^1 = h - w. \quad (6)$$

The input sound port and acoustic material act as a damped mass, M_F , of incompressible air, having effective area A_F and damping coefficient R_F , which add a degree of freedom to the system. When an acoustic signal $p(t)$ is impressed upon this mass, coupling to the diaphragm is effected through a front acoustic chamber stiffness K_F . Diaphragm motion is likewise influenced by the rear acoustic chamber stiffness K_R , thin air film damping¹⁹ R_R , its own mechanical mass and stiffness, and air mass loading. Due to the inherently low capacitance of the combined cells, a preamplifier of high input impedance,

$$\mathbf{Z}_p = 1 / (1 / R_p + j \omega C_p), \quad (7)$$

and near-unity gain A is housed in the microphone rear acoustic chamber. Hereafter, a bold-face quantity denotes that it is to be considered complex.

2.2 Equations of motion and stability

One may show that the electrostatic potential energy stored in the dielectric/electrode system in Fig. 1 may be written in terms of σ , σ_B^1 , d , and h^1 . Energy and external work expressions for all mechanical, electrical, and acoustic elements may likewise be written. With these expressions, one can treat the coupled electrostatic-dynamic system by letting

$$\sigma_B^1 \rightarrow \sigma_B^1 + \bar{\sigma}_B(t), \quad h^1 \rightarrow h^1 - u(t), \quad (8)$$

$\bar{\sigma}_B$ being the "small" time-varying charge and u the small dynamic

diaphragm displacement. An application of Lagrange's equation yields the equations of motion¹⁶

$$\begin{bmatrix} PA_e \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} Z_K & 0 & Z_F \\ Z_M & -\chi & Z_K \\ -\chi & Z_e & 0 \end{bmatrix} \begin{bmatrix} \dot{U} \\ I \\ \dot{Y} \end{bmatrix}, \quad (9)$$

where, assuming small dynamic fields superposed on a "large" electrostatic bias, we have used $u(t) \ll h^1$, $\dot{\sigma}_B(t) \ll \sigma_B^1$. Above,

$$\begin{aligned} \chi &= (\sigma d/\epsilon)/j\omega(h^1 + d\mu), & \mu &= \epsilon_o/\epsilon, & j^2 &= -1, & Z_K &= -K_F/j\omega \\ Z_F &= (R_F + j\omega M_F)/\tau^2 - Z_K, & \tau &= A_F/A_e, & Z_e &= Z_p + 1/j\omega C^1, \\ C^1 &= \epsilon_o A_e \nu / (h^1 + d\mu), & Z_M &= j\omega M_e + R_R + (K_e + K_R + K_F)/j\omega, \\ K_{F,R} &= \rho_a c^2 A_e^2 \xi / \nu_{F,R}, & M_e &\rightarrow M_e + (1/3)\rho_a (\nu_F + \nu_R)/\xi. \end{aligned} \quad (10)$$

In eqs. (9), P , \dot{U} , $I \equiv \dot{\sigma}_B A_e \nu$, and $\dot{Y} \xi A_e$ are the steady-state complex amplitudes of the input sound pressure, piston diaphragm velocity, cell current, and sound port volume velocity, respectively, and χ is the electromechanical coupling coefficient. The factor ν is defined as that fraction of A_e metallized on the stationary electrode, ω is the circular frequency, ν_F and ν_R are the front and rear acoustic chamber volumes (ν_R includes the volume of the stationary electrode holes), and ρ_a , c are the density of, and sound wave velocity in, air. Relation (10k) on M_e allows for the air mass loading on the diaphragm by the "gas spring" acoustic chambers. In addition, Lagrange's equations, due to equilibrium of the bias state, yield the electrostatic stability criterion

$$\eta - L/(1 - \eta)^2 = 0, \quad (11)$$

where, for a stable and physically realizable solution,

$$\begin{aligned} L &\equiv (\sigma d/\epsilon)^2 \epsilon_o \alpha^2 \beta \nu / 2\gamma S (h + d\mu)^3 \leq 4/27, \\ \eta &\equiv w/(h + d\mu) \leq 1/3, \end{aligned} \quad (12)$$

a well-known result.²⁰

The open-circuit "cell sensitivity" and associated cell source impedance are defined, respectively, as

$$V = \text{Limit}_{Z_p \rightarrow \infty} (Z_p I)/P, \quad Z_s = (VP)/I|_{Z_p=0} \quad (13)$$

In the absence of small motional impedance terms evident near resonance, Z_s^{-1} can be shown to reduce to $j\omega C^1$, where C^1 is the cell capacitance in the electrostatic equilibrium state; see eq (10h). Of course, (13b) will be employed in calculations. The electret "transmit-

ter sensitivity" is given by

$$V_p = V[Z_p/(Z_p + Z_s/\xi)]A, \quad (14)$$

where ξ accounts for the fact that the cells are in parallel, and A was defined following eq. (7). Before proceeding, it is noted from Griffin's work¹⁹ that the thin air film damping R_R is proportional to $1/(h^1)^3$, and to the cube of the spacing between rows of stationary electrode holes in each cell as shown in Figs. 1 and 5. This degree of freedom over the magnitude of R_R will prove most helpful.

2.3 Optimization and EL2 design

In telephony, it is desirable that the transmit frequency response rises and peaks at the upper end of the telephone bandwidth, that is, between 3 and 4 kHz. This specification on frequency and level, which places a side condition on microphone design, may be approximated by allowing M_F and R_F , i.e., the sound port impedance, to vanish presently in the expression for V . Then, finding the maxima of $|V|$ with frequency in terms of the well-known dynamic magnification factor,²¹ $\mathcal{M}(\omega)$, of a mass-spring-damper system can be shown to yield

$$\omega_d^2/\omega_n^2 = (1 - \mathcal{M}_M^{-2})^{1/2} = 1 - R_R^2/2M_c^2\omega_n^2, \quad (15)$$

where \mathcal{M}_M is the maximum of \mathcal{M} occurring at the damped resonance frequency ω_d , and the natural frequency is given by

$$\omega_n^2 = (K_c + K_R)/M_c. \quad (16)$$

Since \mathcal{M}_M and ω_d will be prescribed, (15a) yields a side condition on ω_n . It is also seen from eq. (15b) that the desired \mathcal{M}_M may be achieved by realizing the proper R_R as previously indicated. While the analytical optimization will be described more thoroughly in Ref. 16, most results are provided here. First, with stability criterion (12a) and side-condition (16), it may be shown that maximum $|V|$ is achieved for $\omega_n^2 = K_c/M_c$, that is, $v_R = \infty$. A design at $K_c = K_R$ will be 3 dB below maximum $|V|$, but will require lower film tension. This result is unlike that reported by Fraim et al.²² where the frequency side condition was treated but not as an integral part of the optimization. Second, it may be shown that $|V| \propto (h + d\mu)^{1/2}$ owing to a higher electrostatically stable electret charge level with increased h [see eqs. (11) and (12)] as seen in Fig. 2. However, since increasing h lowers the source capacitance, eq. (14) can yield a transmitter sensitivity which reaches a maximum with h , as seen. It may be shown from side condition (16) that

$$S \propto 1/\xi^2. \quad (17)$$

Then, in the stiffness-controlled region it follows that $|V|$ is independ-

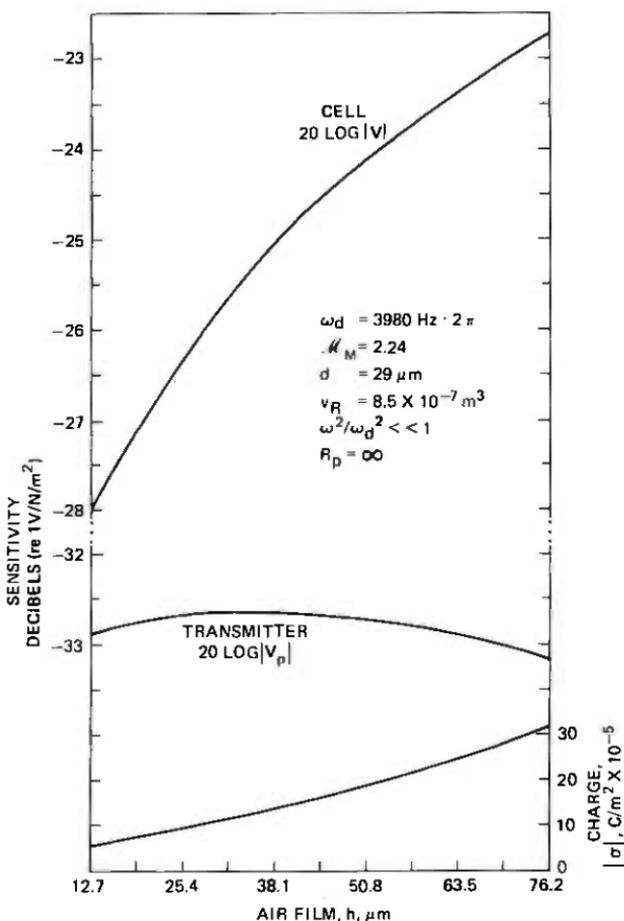


Fig. 2—One-day cell sensitivity, transmitter sensitivity, and corresponding stable electret charge vs air film thickness for a rectangular design. For all h , $S = 60.1 \text{ N/m}$. No front chamber is present.

ent of ξ and thus cell size. Similarly, $|Z_s|$ and $|V_p|$ may be shown independent of ξ allowing relation (17) to be freely employed in keeping film stress levels below strength limitations by increasing ξ .

The design approach is now to use condition (16) to determine S for a given geometry. Then constraint (12a) determines the stable charge level possible. Such data were obtained for the EL2 for $\omega^2/\omega_d^2 \ll 1$ while iterating parameters such as d , h and v_R . In applying (12a), the upper bound on L , i.e., $4/27$, was reduced to $\zeta(4/27)$ to allow margin for (i) worst-case tolerances expected, (ii) anticipated aging effects, and (iii) a safety factor related to extreme service conditions, modeling assumptions, etc. Item (ii) is discussed later. The following values,

descriptive of the EL2, were used to generate Fig. 2:

$$\begin{aligned} \omega_{cl} &= 2\pi(3980 \text{ Hz}), \dagger & \mathcal{M}_M &= 2.24, \dagger & \xi A_c &= 7.3 \times 10^{-5} \text{ m}^2, \\ \lambda &= 1.74, & \xi &= 3, & \nu &= 0.73, & d &= 29 \text{ } \mu\text{m}, \\ v_R &= 8.5 \times 10^{-7} \text{ m}^3, \\ \zeta &= 0.114, & \rho &= (2.23 \times 10^3 + 38.6 \times 10^{-4}/d) \text{ kg/m}^3, \\ \epsilon &= 2.08\epsilon_0, & A &= 0.9, & C_p &= C_{in} + C_{stray} = (5 + 6) \text{ pF}. \end{aligned} \quad (18)$$

The second term in eq. (18j) on ρ is added to the EL2 polymer film density to account for the metallized electrode mass (film and metalization are described in Section III), and ϵ_0 , ρ_0 , and c are well known. For all h in Fig. 2, condition (16) yielded tension $S = 60.1 \text{ N/m}$. Curves such as those in Fig. 2 together with additional physical, material, and electrical constraints allowed optimum final parameters such as h , d , and v_R to be chosen. With the parameters in eq. (18) and choosing $h = 36.1 \text{ } \mu\text{m}$, Fig. 2 gives $\sigma = 13 \times 10^{-5} \text{ C/m}^2$ ($\sigma d/\epsilon = 204 \text{ volts}$) for the EL2.

2.4 Thermal stabilization and EL2 performance

Given the above EL2 design parameters, we find that $h^1 = 35.2 \text{ } \mu\text{m}$ and $R_R = 0.032 \text{ Ns/m}$ which, together with cell geometry, yield¹⁹ the required spacing between rows of holes in the stationary electrode. Then the transmitter performance including its sound port acoustic impedance²³ influence is seen in Fig. 3, where we have used for the EL2

$$\begin{aligned} R_p &= 10^8 \text{ ohms}, & v_F &= 0.7 \times 10^{-7} \text{ m}^3, & A_F &= 1.72 \times 10^{-6}/\xi \text{ m}^2, \\ R_F &= 0.4 \rho_0 \phi / \xi + 2.53 \times 10^{-6} \rho_0 (\omega \phi)^{1/2} / \xi \text{ Ns/m}, \\ M_F &= 2.21 \times 10^{-9} \rho_0 / \xi \text{ kg}, \end{aligned} \quad (19)$$

ϕ being the kinematic viscosity of air. The calculated response is seen to match that typically measured quite favorably across the telephone bandwidth. The response, influenced by input acoustics, peaks at about 3080 Hz and $|V_p|$ at 1 kHz and ξC^1 are $-31.8 \text{ dBre } 1 \text{ V/N/m}^2$ and 9.6 pF , respectively. Above the frequency of peak response (associated with the first diaphragm resonance), the results begin to differ due to the single degree-of-freedom model used for the continuous diaphragm.

To this point, we have ignored tension and electret charge variation with time; accordingly, all previous results are referred to 1 day past manufacture, as represented in Figs. 2 and 3. However, the viscoelastic polymer film undergoes stress relaxation and/or creep under load. Similarly, extensive investigation has been conducted on electret charge retention at Bell Laboratories and will be reviewed in a com-

† Prescribed in the absence of sound port impedance.

panion article.¹³ Here, we simply give a curve fit for the assumed EL2 electret charge retention at room conditions normalized to the 1-day level, $\sigma[1]$:

$$\sigma(t)/\sigma[1] = \sum_i \Gamma_i e^{-t/\delta_i}, \quad (600 \text{ s} \leq t \leq 20 \text{ yrs}),$$

$$\Gamma_i \rightarrow \begin{bmatrix} 0.954 \\ 0.039 \\ 0.001 \\ 0.004 \\ 0.003 \\ 0.005 \end{bmatrix}, \quad \delta_i \rightarrow \begin{bmatrix} 9.66 \times 10^{11} \\ 1.08 \times 10^9 \\ 2.37 \times 10^7 \\ 3.81 \times 10^6 \\ 2.50 \times 10^5 \\ 1.03 \times 10^4 \end{bmatrix} \text{ s.} \quad (20)$$

Data indicate the initial electret charge to be about 1 percent above $\sigma[1]$, which will be employed here. Since $\sigma/\sigma[1]$ is only down to 0.97 at 20 years, it is not surprising that the diaphragm tension relaxation will be of far greater impact. Recently, Wang and Matsuoka²⁴ studied the viscoelastic behavior of metallized annealed EL2 film. Through time-temperature superposition, they obtained a reduced relaxation modulus vs reduced time over a 500-year period at room temperature (using 18,000-s high temperature isotherms-ambient to 75°C). Using a generalized Maxwell model,²⁵ their result for the room temperature relaxation modulus is here expressed numerically by

$$\mathcal{G}^o(t) = \sum_i \mathcal{G}_i e^{-t/\tau_i}, \quad (600 \text{ s} \leq t \leq 500 \text{ yrs}),$$

$$\mathcal{G}_i \rightarrow \begin{bmatrix} 30.11 \\ 3.61 \\ 4.04 \\ 4.42 \\ 4.97 \\ 7.03 \end{bmatrix} \times 10^7 \text{ N/m}^2, \quad \tau_i \rightarrow \begin{bmatrix} 1.14 \times 10^{11} \\ 1.59 \times 10^9 \\ 1.88 \times 10^8 \\ 6.36 \times 10^6 \\ 2.38 \times 10^5 \\ 6.48 \times 10^3 \end{bmatrix} \text{ s,} \quad (21)$$

where superscript "o" denotes room temperature. To their Arrhenius plot of the temperature shift factor a_T , the Arrhenius equation,

$$\ln a_T = 25,530 [1/(T + 273) - 1/(T^o + 273)], \quad (22)$$

has here been fit, where T is temperature in °C, and $T^o = 22.7$ °C.

In EL2 manufacture, the diaphragm undergoes a fixed stress $\bar{\sigma}^M$ for a short duration, $t = 0$ to \bar{t} , during which time the material creeps to strain $\bar{\epsilon}^M$. At time \bar{t} , the film diaphragm is clamped permanently at strain $\bar{\epsilon}^M$. Stress relaxation in accordance with the decreasing modulus then ensues. While $\bar{\sigma}^M$ could be chosen such the proper membrane tension, S , and hence frequency response previously prescribed at 1 day were obtained, unsatisfactory increases in sensitivity and decreases in resonance frequency would occur during, say, a 20-year service life.

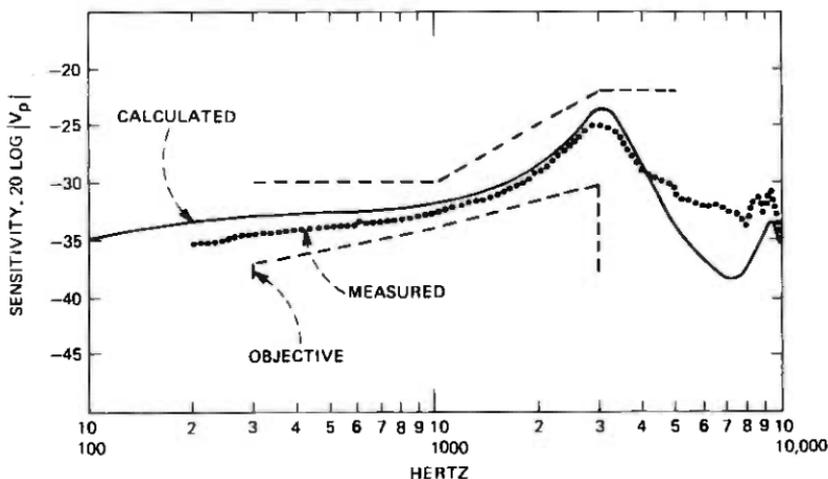


Fig. 3—Typical EL2 frequency response one day after assembly (following thermal stabilization), at room temperature.

For this reason,²⁶ a "thermal stabilization" procedure has been developed whereby the electret devices are placed at high temperature T^* for duration $\Delta t = t_2 - t_1$ where $\bar{t} \ll t_1 < t_2 \leq 86,400$ s (1 day). While the mechanics of the resulting thermoviscoelastic problem are only sketched below, the reader may refer to Ref. 16 for more details. For a thermorheologically simple viscoelastic material undergoing "small" strain, the following linear constitutive equation holds²⁷

$$\sigma^M(t) \equiv \hat{\sigma}^M(\theta) = \int_0^\theta \mathcal{G}^o(\theta - \theta') \frac{\partial \epsilon^M(\theta')}{\partial \theta'} d\theta',$$

$$\theta(t) \equiv \int_0^t 1/a_T dt, \quad (23)$$

where t is the real, and θ the reduced, time. For $t \gg \bar{t}$, it may be shown that the creep between $t = 0$ and \bar{t} may be viewed as a step in strain of magnitude $\bar{\epsilon}^M$ at $t = 0$. From the convolution integral (23a), it may be shown using the Dirac-delta function and Laplace transforms that

$$\sigma^M(t) = \hat{\sigma}^M(\theta) = \bar{\epsilon}^M \mathcal{G}^o(\theta), \quad t \gg \bar{t}. \quad (24)$$

Now, for $\bar{t} \ll t \leq t_1$, $\theta = t$ and eq. (24) yields a well-known result.²⁸ During stabilization, eq. (23b) gives $\theta = t_1 + (t - t_1)/a_T(T^*)$, which is the basis for time-temperature superposition.²⁹ For the service life, $t \leq t_2$,

$$\theta = t_1 + (t - t_2) + \Delta t/a_T(T^*). \quad (25)$$

At 1 day, the stress $\sigma^M = S/d$ is prescribed as can be seen from eqs.

(15), (16), and (4b)—see the argument following eq. (16). Thus, $\bar{\epsilon}^M$ may be determined from eqs. (24) and (25). In a manner similar to that used in obtaining eq. (24), it can be shown that the actual step change in $\bar{\sigma}^M$ at $t = 0$ yields

$$\epsilon^M(t) = \bar{\sigma}^M J^o(\bar{t}), \quad 0 < t \leq \bar{t}. \quad (26)$$

With the creep compliance $J^o(\bar{t})$ known, we can then find from eq. (26) the required $\bar{\sigma}^M$ that must be applied to the film for duration \bar{t} to achieve the desired 1-day tension following stabilization. Microphone performance for all $t \gg \bar{t}$ (i.e., during the service life) is thus independent of the choice of a $(\bar{t}, \bar{\sigma}^M)$ pair satisfying eq. (26). However, since the creep compliance increases more rapidly with time initially, manufacturing tolerances on \bar{t} , to maintain a given (acceptable) variance on σ^M for all $t \gg \bar{t}$, are less restrictive for larger \bar{t} . Desiring time-independent transmitter performance through the service life, the following parameters were chosen for the EL2: $T^* = 60^\circ\text{C}$, $\Delta t = (2/3 \text{ day})$, and $t_2 = (1 \text{ day})$, yielding $\bar{\epsilon}^M = 0.65$ percent. Then, for $\bar{t} = 60 \text{ s}$, we use²⁴ $J^o(\bar{t}) = 10.6 \times 10^{-6} \text{ m}^2/\text{N}$ and obtain the EL2 initial stress $\bar{\sigma}^M = 4.2 \times 10^6 \text{ N/m}^2$ or tension, 122 N/m (more than twice the 1-day level), to be applied. Note that $\bar{\sigma}^M$ could be lowered by increasing \bar{t} . It is interesting that at 1 day, $\theta \approx (29 \text{ years})$, that is, the equivalent of 29 years of room temperature relaxation has been effected. Now, with time variations in both eqs. (20) and (24) present, quantities like h^1 , R_R , and K_e change accordingly, and the results in Fig. 4 are obtained; EL2 performance during the service life has been "stabilized." The 1-kHz transmitter

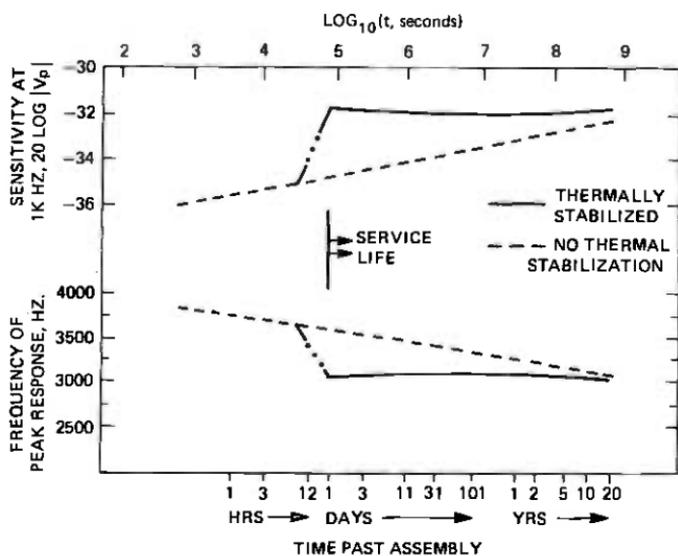


Fig. 4—Predicted EL2 sensitivity at 1 kHz and frequency of peak response level vs time at room temperature, stabilized at 60°C for 16 hours.

sensitivity is seen to drop very slightly during the first year due to minor but dominant charge decay assumed, after which time slight tension relaxation causes $|V_p|$ to rise and the resonance frequency to decline slightly.

III. PHYSICAL DESIGN

3.1 Constraints and objectives

The physical design of the EL2 electret transmitter was guided by the following factors: potential applications, reliability, and manufacturability. The primary application for the new electret microphone has been the replacement of the AF1 in the 680-type transmitter module used with the 4A Speakerphone. This application dictated physical constraints based on the existing module configuration. The size of the AF1 was an upper bound on that of the EL2. Electrical compatibility with other expected applications was also a major influence on the impedance and dc power requirements for the microphone. The low cell capacitance results in a combined cell source impedance (Z_s/ξ) of 16 M Ω at 1 kHz. This high impedance necessitates the use of an internal preamp intimately associated with the cells to reduce loading of the output signal (V), and to minimize external electromagnetic field pickup. Potential use in line-powered applications required the preamplifier to be operational down to a supply voltage of 2 volts. Performance objectives were also stipulated to insure satisfactory operation in all anticipated applications. Electroacoustic conversion efficiency, frequency response, and mechanical and electromagnetic interference sensitivity were some of the operational characteristics specified.

The optimized parameters generated with the analytical model [see eqs. (18) and (19) for most physical design parameters] provided a framework for the physical design. The ultimate configuration reflects model results that were in turn influenced by physical, material, and telephone constraints. A major effort was made to insure simplicity and ease of assembly while stressing environmental reliability and long life. Twenty years service within an operating temperature range of -23°C to $+49^\circ\text{C}$ was the overall design objective.

3.2 Description

3.2.1 Subassembly

The heart of the microphone is the subassembly or "cartridge," shown in the exploded view in Fig. 5. The diaphragm is positioned and tensioned across the face of the backplate between the shim and clamping plate. The spring clip clamps these parts against the top of the backplate while holding the preamplifier in position on the underside. The backplate is the nucleus of this structure with the other parts

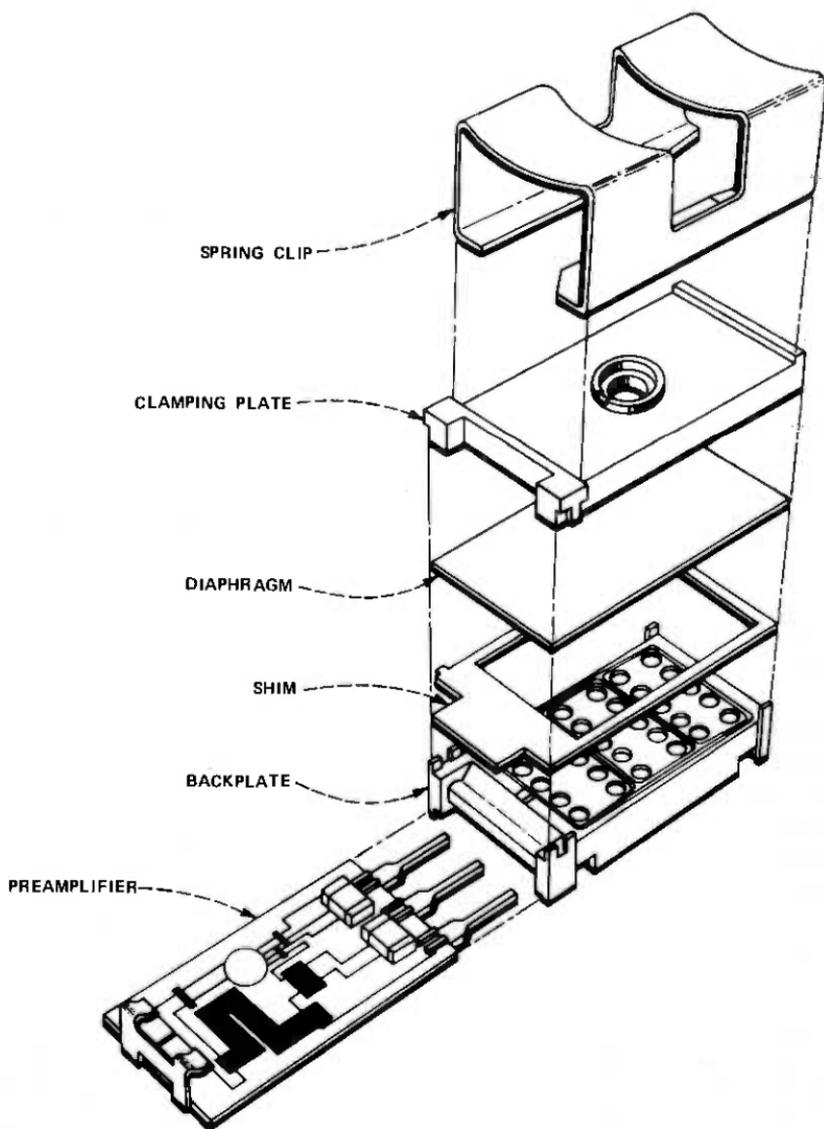


Fig. 5—EL2 subassembly.

assembled to it. An electroplatable grade of acrylonitrile butadiene styrene (ABS) is used to mold the backplate to facilitate selective metallization of the three microphone cell areas located on its upper surface. These conductive areas, which are electrically common, serve as the stationary electrode of the microphone configuration. A small diaphragm support rib $36 \mu\text{m}$ high (h) surrounds each metallized cell area. These supports are unmetallized to reduce stray capacitance. The diaphragm resting on these ribs creates the air film (h^1) between

the flexible diaphragm and the stationary electrode. The sensitivity of the microphone output to changes in this air film necessitates the use of precision molding techniques to maintain dimensional control of rib height to within $\pm 5 \mu\text{m}$ (± 0.0002 in). The holes through the backplate couple the air film under the diaphragm to the rear acoustic chamber and are spaced at a distance to effect proper air film damping in accordance with the model in Section II. They are located in rows at the outer edges of the cells adjacent to the ribs. This maximizes the amount of continuous electrode area (νA_e) in the center region where the greatest diaphragm displacement (u) occurs.

The preamplifier is a conventional source follower configuration with a JFET transistor and two thick film resistors on a ceramic substrate providing an effective impedance transformation and a voltage gain of -1 dBV. The input impedance (Z_p) is approximately 100 M ohms resistance in parallel with about a 5-pF junction, and 6-pF stray capacitance. The output impedance is nominally 1 kilohm. A contact spring soldered to one end of the substrate provides electrical continuity from the stationary electrode on the backplate to the input of the amplifier circuit, while mechanically clipping the preamp to the backplate. Three leads extending from the opposite end of the substrate provide for electrical bias, output, and common connections to the assembled microphone. Two small chip capacitors bridge these leads on the substrate to provide RFI suppression.

The electret diaphragm material is a 29- μm thick, stress-relieved, cast, poly(tetrafluoroethylene) (PTFE) film. It is metallized on one side with a titanium flash approximately 100 Å thick under 2000 Å of evaporated gold.¹³ The metallized side of the film functions as the movable electrode in the microphone. This particular Teflon film was selected for its superior charge retention characteristics. It is electrostatically charged to approximately $13 \times 10^{-5} \text{ C/m}^2$ to provide the electric field (E'_0) in the microphone assembly air film. The charge ($-\sigma$) on the polymer film eliminates the need for the external dc bias required in conventional condenser microphones. The diaphragm is tensioned longitudinally under fixed load $S = 122 \text{ N/m}$ for duration $\bar{t} = 60$ sec minimum prior to clamping it in the relative position indicated, and with the metallized side opposite the air film. The nominal value of \bar{t} chosen is influenced by two primary considerations. The duration of \bar{t} should be maximized to minimize its tolerance stringency [see the discussion following eq. (26)], whereas it should be minimized for the sake of manufacturing expediency.

A thin Mylar† shim is sandwiched between the backplate and

† Registered trademark of E. I. DuPont de Nemours.

clamping plate along with the diaphragm film. It provides a relatively compliant surface against which to clamp the film. "Hard" clamping the diaphragm without a shim would result in nonuniform clamping pressure, causing nonuniform tension across the width of the film. The tab on the end of the shim provides electrical insulation between the preamp contact spring and the enclosure in the final assembly.

The clamping plate is a metallized precision molded piece part which is indexed on the backplate and clamps the pretensioned diaphragm in position across the cell area. Both the clamping plate and backplate are molded from the same grade of ABS material so as to have a similar thermal expansion coefficient. This coefficient also approximates that of the electret film. The clamping plate is metallized to provide electrical continuity from the metallized side of the electret film to the spring clip, which is connected directly to the common terminal on the preamp.

The phosphor bronze spring clip maintains a compressive spring force of approximately 20 N on the assembled parts to keep the diaphragm properly tensioned. This force is sufficiently large to prevent slippage of the diaphragm in the cartridge assembly, but is limited to avoid overstressing and causing cold flow of the plastic parts. The clip also provides electromagnetic shielding by surrounding most of the subassembly and being electrically connected to the preamp common terminal. After clamping with the spring clip, the subassembly is a working microphone configuration. The spring clip design permits easy disassembly to facilitate repair or reassembly if required prior to final assembly.

The completed cartridge assembly is thermally stabilized, which involves "soaking" the units at an elevated temperature of 60°C for a period of 16 hours to accelerate the diaphragm stress relaxation. This preaging treatment results in a more uniform and stable product (see Fig. 4). Preliminary testing is done after stabilization to evaluate microphone performance prior to final assembly.

3.2.2 Final assembly

Final assembly involves ferruling the pretested subassembly into a rectangular aluminum enclosure and back cover. A stainless steel woven wire acoustic screen is included in front of the subassembly and a gasket across the back, as shown in Fig. 6. The screen, in conjunction with the input sound port, provides acoustic impedance elements R_F and M_F which modify the frequency response. The wire screen also serves as a dirt shield across the input sound port. The gasket provides electrical insulation between the preamp and back cover and functions as a shock-absorbing element in the microphone assembly. The external aluminum enclosure is electrically connected to the common ter-

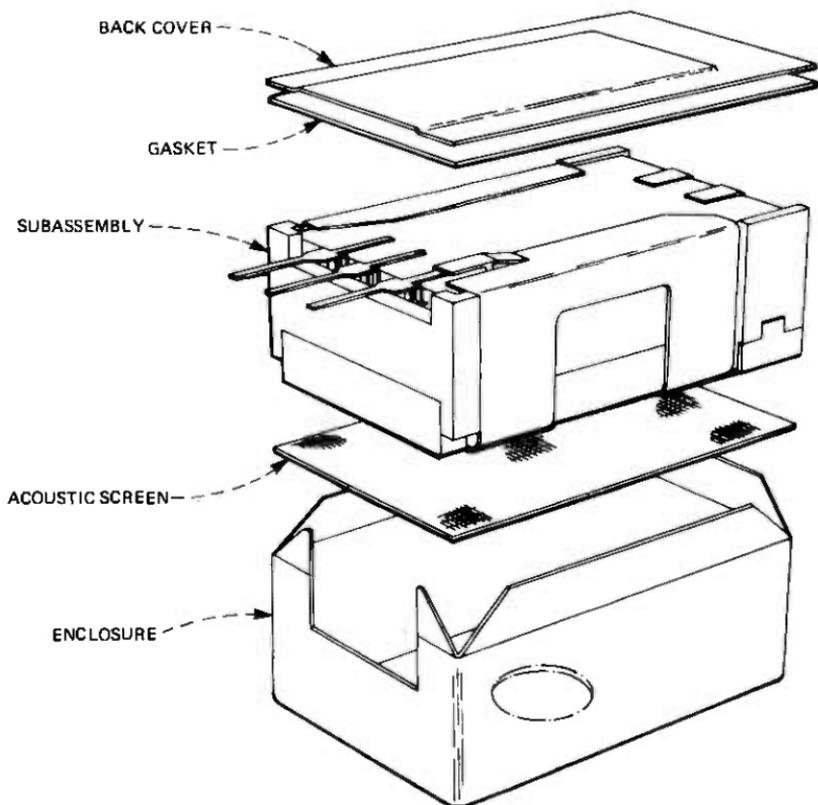


Fig. 6—EL2 final assembly.

minal on the preamp, thereby providing additional electromagnetic shielding. Acoustic leaks around the leads are sealed, and the complete unit is tested according to the final requirements. A cross section of the completed EL2 microphone is illustrated in Fig. 7 to provide an overall perspective of the assembly, which is 18 by 12 by 7.6 mm in depth.

IV. MICROPHONE PERFORMANCE

4.1 Performance characteristics

The EL2 operates on a bias supply of 2 to 16 V while drawing about 150 μ A. Figure 3 shows a typical measured frequency response of the telephone transducer compared to the calculated result from the model. The response is relatively flat in the lower frequency stiffness-controlled region and rises about 7 dB to the resonance peak at 3200 Hz, above which the level falls off rapidly. The typical electroacoustic sensitivity at 1 kHz is -32.0 dBV/N/m². Model results show that the output signal level is altered by (uncompensated) changes in a number

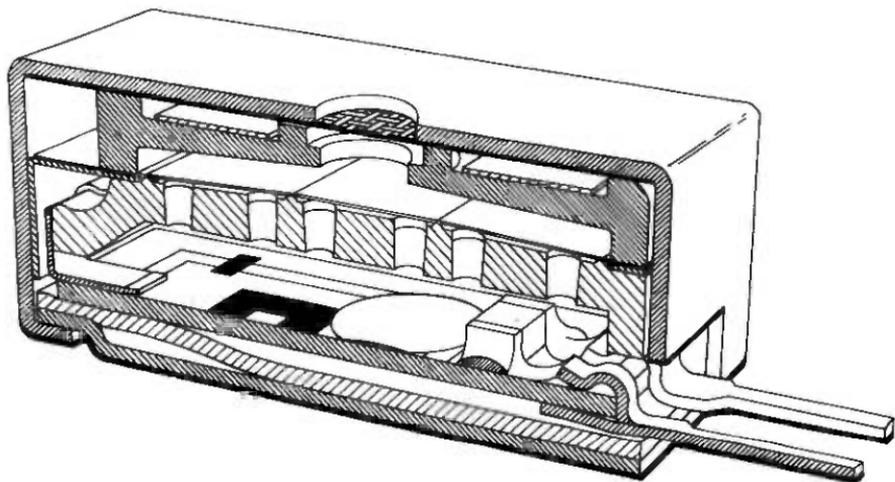


Fig. 7—EL2 Transmitter Unit.

of the basic microphone parameters. For example, the variation of output voltage versus charge for this microphone configuration is given by $0.7 \text{ dBV}/10^{-5}\text{C}/\text{m}^2$. This relationship was verified both analytically and empirically. EL2 sensitivity will also vary with changes in the air film thickness in accordance with the gradient $-1.4 \text{ dBV}/5\mu\text{m}$, which substantiates the need for precision molding of the diaphragm support ribs. The microphone output variation with applied tension is given by $-0.5 \text{ dBV}/8\text{N}/\text{m}$. The resonance frequency also varies with tension according to the ratio of $100 \text{ Hz}/8\text{N}/\text{m}$. It must be pointed out that these coefficients, which are linear approximations, apply only for minor variations about the nominals specified for the basic parameters.

The vibration sensitivity of the microphone is $< -40 \text{ dB re } 1 \text{ V}/g$ at 1000 Hz . The EL2 "signal voltage to spurious mechanical noise voltage" ratio is about 20 dB greater than that of the AF1 microphone. This improved vibration isolation is acquired due to the relatively low mass of the electret diaphragm. The RFI sensitivity is $< -75 \text{ dB re } 1 \text{ V}$ in a $10 \text{ V}/\text{m}$ field from 0 to 300 MHz . The hum or low-frequency electromagnetic pickup sensitivity is $< -95 \text{ dB re } 1 \text{ V}/\text{gauss}$ at 60 Hz . The hum interference rejection of the electret is 35 dB better than that of the AF1. In general, the electret microphone successfully met all the performance objectives initially outlined and performs well as a replacement for the AF1 in the 4A Speakerphone.

4.2 Reliability

The ruggedness and reliability of the EL2 design have been established by tests³⁰ which included mechanical and thermal shock, vibra-

tion, and combinations of temperature and humidity. The small changes which occurred during the course of the study tests had minimal effect on overall performance. The output was found to vary slightly as a function of temperature at a rate of $0.045 \text{ dB}/^\circ\text{C}$.³⁰ This is a reversible effect, resulting in no permanent change in output upon return to ambient temperature. The charge retention of the EL2 electret film, found to be very stable, is discussed in detail in a companion article.¹³ Moreover, earlier studies over 800 days of testing compared sensitivity variations (aging) of numerous commercial and Bell Labs' prototype electret microphones to transducers employing other mechanisms including granular carbon, electromagnetic, and piezoelectric. The aging conditions ranged between $25^\circ\text{C}/40\% \text{ R.H.}$ and $65^\circ\text{C}/90\% \text{ R.H.}$ The electret devices performed comparably to the other types of microphones in these tests.

V. SUMMARY

An EL2 electret transmitter has been developed which provides lower sensitivity to spurious electromagnetic and mechanical signals than the AF1 magnetic transmitter it has replaced in the 4A Speakerphone. It offers lower dc power consumption, smaller size, and lower intrinsic noise and distortion than the carbon transmitter. These attributes make it a strong candidate to replace the carbon transmitter in future electronic and special-purpose sets.

An electro-mechano-acoustic model of an electret transmitter governing its coupled electrostatic/dynamic operation has allowed EL2 parameters such as air film thickness, number of cells, and acoustic chamber volume to be collectively optimized. It is shown that, within bounds, sensitivity can actually be raised by designing with increased air film thicknesses. The analysis is novel in consistently treating a side condition governing the prescribed response resonance characteristics as an integral part of the optimization. Also, a thermal stabilization procedure used to minimize the effects of long-term stress relaxation (accompanied by charge decay) is described theoretically and optimized. Virtually constant sensitivity over the service life is shown possible.

These results, coupled with physical, material, and application constraints, are shown applied in the physical design to achieve desired performance and reliability. A metallized Teflon electret film is tensioned, supported, and clamped above a selectively metallized stationary electrode forming three cells acoustically and electrically in parallel. A preamplifier and input sound port complete the subassembly, which is then tested and housed in a rectangular aluminum enclosure, electrically shielding the transducer. The nominal EL2 electroacoustic sensitivity, output impedance, response resonance frequency, and re-

quired dc supply are -32 dBV/N/m², 1 k Ω , 3200 Hz, and $+2$ to 16 V, respectively. The response is relatively flat in the low-frequency stiffness-controlled region. A 3-dB cutoff occurs below 100 Hz.

For completeness, note that the EL2 is described here as it existed in the initial design and field trial units and does not reflect minor changes as a result of manufacturing experience. A companion article¹³ will treat the new technological aspects of the EL2.

VI. ACKNOWLEDGMENTS

The authors are thankful to J. E. Warren for stimulating discussions regarding the mathematical model and helpful suggestions during the overall EL2 design effort. S. R. Whitesell contributed to final design modifications, and R. S. Ball to measurements required in support of EL2 introduction to manufacture. T. T. Wang shared ideas regarding implications of the viscoelastic nature of the film and obtained required film data.

APPENDIX

Partial List of Symbols

$-\sigma$	Negative electret surface charge per unit area.
h^1	Air film thickness h^1 in the electrostatic state.
h	Rib height.
d	Polymer film thickness.
$\mu = \epsilon_0/\epsilon$	Inverse of relative permittivity.
M_e, A_e, K_e	Diaphragm effective mass, area, and stiffness per unit cell.
S	Membrane diaphragm's applied force per unit edge (tension).
ξ	Number of diaphragm cells.
V_R, K_R	Rear acoustic chamber volume, stiffness.
R_R	Air film damping coefficient
ν	That fraction of A_e metallized on the rear electrode.
$w, u(t)$	Diaphragm electrostatic and dynamic displacement.
V, Z_s	Cell open-circuit sensitivity and source impedance.
C^1	Electrostatic cell capacitance.
$Z_p = (1/R_p + j\omega C_p)^{-1}$	Preamplifier input impedance.
V_p^-	Transmitter (microphone) sensitivity.
ω_n, ω_d	Diaphragm natural and damped resonance frequency (in absence of front chamber).

M_M	Maximum of the dynamic magnification factor (in absence of front chamber).
$\mathcal{G}^\circ(t)$	Room temperature mechanical relaxation modulus of the metallized film diaphragm.
$\sigma^M(t), \epsilon^M(t)$	Diaphragm stress and strain.
t, θ	Real and reduced time.
\bar{t}	Time duration of diaphragm creep prior to clamping.
$\bar{\sigma}^M, \bar{\epsilon}^M$	Diaphragm stress and strain at time $t = \bar{t}$.

REFERENCES

- Alexander Graham Bell, "Researches in Telephony," Proc. Am. Acad. Arts Sci., 12 (1877), pp. 1-10.
- C. Flannagan, R. Wolf, and W. C. Jones, "Modern Theater Loud Speakers and Their Development," J. Soc. Mot. Pict. Eng., XXVIII, No. 3 (May 1937), pp. 246-263.
- F. S. Goucher, "The Carbon Microphone," B.S.T.J., 13, No. 2 (April 1934), pp. 163-194.
- J. E. Warren, J. F. Hamilton, and A. M. Brzezinski, "Capacitance Microphone Dynamic Membrane Deflections," J. Acoust. Soc. Am., 54, No. 5 (1973), pp. 1201-1213.
- H. J. Wintle, "Introduction to Electrets," J. Acoust. Soc. Am., 53, No. 6 (June 1973), pp. 1578-1588.
- J. K. Hilliard, "Electroacoustics to 1940," J. Acoust. Soc. Am., 61, No. 2 (February 1977), pp. 267-273. See Section VII.
- E. C. Wentz, "A Condenser Transmitter as a Uniformly Sensitive Instrument for the Absolute Measurement of Sound Intensity," Phys. Rev., 10 (July 1917), pp. 39-63.
- G. M. Sessler and J. E. West, "Electret Transducers: A Review," J. Acoust. Soc. Am., 53, No. 6 (June 1973), pp. 1589-1599.
- G. M. Sessler and J. E. West, "Condenser Microphone with Solid Dielectric," J. Audio Eng. Soc., 10, No. 3 (July 1962), pp. 212-215.
- G. M. Sessler and J. E. West, "Self-Biased Condenser Microphones with High Capacitance," J. Acoust. Soc. Am., 34, No. 11 (November 1962), pp. 1787-1788.
- C. W. Reedyk, "Noise-Cancelling Electret Microphone for Lightweight Head Telephone Sets," J. Acoust. Soc. Am., 53, No. 6 (June, 1973), pp. 1609-1615.
- J. C. Baumhauer, Jr., A. M. Brzezinski, J. E. Warren, and S. R. Whitesell, "Electroacoustic Transducer...", U.S. Patent 4,046,974, applied for October 1976, issued September 1977.
- S. P. Khanna and R. L. Remke, Bell Laboratories, to be published.
- G. M. Sessler, "Electrostatic Microphone with Electret Foil," J. Acoust. Soc. Am., 35, No. 9 (September 1963), pp. 1354-1357.
- J. C. Baumhauer and H. F. Tiersten, "Nonlinear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 54, No. 4 (1973), pp. 1017-1034. See eqs. (4), (6), (7), and (20a).
- J. C. Baumhauer, Jr., "Modeling and Analysis of Electret Microphones—An Example in Telephony," to be published.
- H. F. Tiersten, "On the Nonlinear Equations of Thermoelastoelectroelasticity," Int. J. Engng. Sci., 9 (1971), pp. 587-604. See eq. (3.19).
- J. E. Warren, J. F. Hamilton, and A. M. Brzezinski, "Capacitance Microphone Static Membrane Deflections," J. Acoust. Soc. Am., 52, No. 3 (1972), pp. 711-719.
- W. S. Griffin, H. H. Richardson, and S. Yamanami, "A Study of Fluid Squeeze-Film Damping," J. Basic Engineering, Trans. ASME, 88, No. 2 (June, 1966), pp. 451-456.
- K. Teer, "On the Optimum Configuration for a Condenser Microphone," Acoustica, 15, No. 5 (1965), pp. 256-263.
- Yu Chen, *Vibrations: Theoretical Methods*, Reading, Mass.: Addison-Wesley, 1966, eq. (1.3.18).
- F. W. Fraim and P. V. Murphy, "Electrets in Miniature Microphones," J. Acoust. Soc. Am., 53, No. 6 (June 1973), pp. 1601-1608.

23. L. L. Beranek, *Acoustics*, New York: McGraw-Hill, 1954. Material in both Sections 5.7 and 5.9 was employed.
24. T. T. Wang and S. Matsuoka, unpublished communication to J. C. Baumhauer, Jr., August 1975.
25. J. D. Ferry, *Viscoelastic Properties of Polymers*, New York: John Wiley, 1970, pp. 60-63.
26. During a visit to Northern Electric Co. (now Northern Telecom, Ltd.) in September 1974, author J. C. Baumhauer, Jr. was informed that the concept of high temperature conditioning for the purpose of relaxing stress and obtaining a more controlled stress level was in use by that company.
27. R. A. Schapery, "Stress Analysis of Viscoelastic Composite Materials," *J. of Composite Mat.*, 1, No. 3 (July, 1967), pp. 228-267.
28. W. Flügge, *Viscoelasticity*, New York: Springer-Verlag, 1975. See Ch. 1.
29. See Ref. 25, Ch. 11.
30. S. R. Whitesell and R. S. Ball, unpublished communication to A. M. Brzezinski in March/October 1977.

Comparison of Single Heterostructure and Double Heterostructure GaAs-GaAlAs LEDs for Optical Data Links

By A. K. CHIN, G. W. BERKSTRESSER, and V. G. KERAMIDAS

(Manuscript received February 22, 1979)

This paper presents a theoretical and experimental comparison of the performance of single heterostructure (SH) and double heterostructure (DH) GaAs-GaAlAs light-emitting diodes (LEDs). These LEDs are designed for optical data links operating at rates up to T3 (45 Mb/s). The SH LEDs were optimized with respect to active layer carrier concentration and thickness; similar DH LED active layer parameters have not yet been optimized. We find experimentally that the DH diodes launch approximately 8 times the SH power into a butt-coupled, 0.36 numerical aperture (NA), graded-index fiber. Using a diffusion model, we show that the power output of the SH LED is limited by surface recombination and reduced current crowding. These results demonstrate that DH LEDs are necessary for applications requiring high launched power.

I. INTRODUCTION

Burrus-type,¹ single heterostructure (SH) light-emitting diodes (LEDs) are presently fabricated in our laboratory for use in optical data links. Their performance requirements are 45 MHz bandwidth and a minimum butt-coupled power of 10 μ W into a 0.36 numerical aperture (NA), 55- μ m core, graded-index optical fiber at 60 mA forward current. SH LEDs were chosen for the simple growth procedure² and exceptional reliability.³ After optimization of the active layer width and carrier concentration for maximum efficiency while maintaining the required bandwidth, the SH LEDs butt-couple 10 to 15 μ W of power into the fiber. However, initially fabricated, unoptimized double heterostructure (DH) LEDs were found to couple 7 to 8 times more power than the optimized SH LEDs. The purpose of this paper is to explain the difference between the SH and DH LED performance. We begin with an

explanation of the factors influencing the efficiency and bandwidth of the two structures.

Figure 1 is a schematic of our Burrus-type SH and DH LEDs. These LEDs use current crowding by contact area restriction to increase current density for increased brightness. In addition, the localized emission region couples the light more efficiently into the fiber. The contact diameter for both structures is chosen to be $50\ \mu\text{m}$, slightly smaller than the fiber core, for our performance comparison.

Figure 1a shows the design of our SH LED. The active layer is confined on one side by the n-GaAlAs/p-GaAs heterojunction and on the other side by the p-ohmic contact. Since the injected carrier concentration is zero at the p-contact, the p-n junction (i.e., the minority carrier injection source) should be kept at least a minority carrier diffusion length from the contact to minimize nonradiative recombination. However, as the p-n junction is removed in distance from the p-contact, current crowding decreases. Thus, for the SH LED, the active layer width is a compromise between the effects of current crowding and nonradiative recombination at the contact. Five microm-

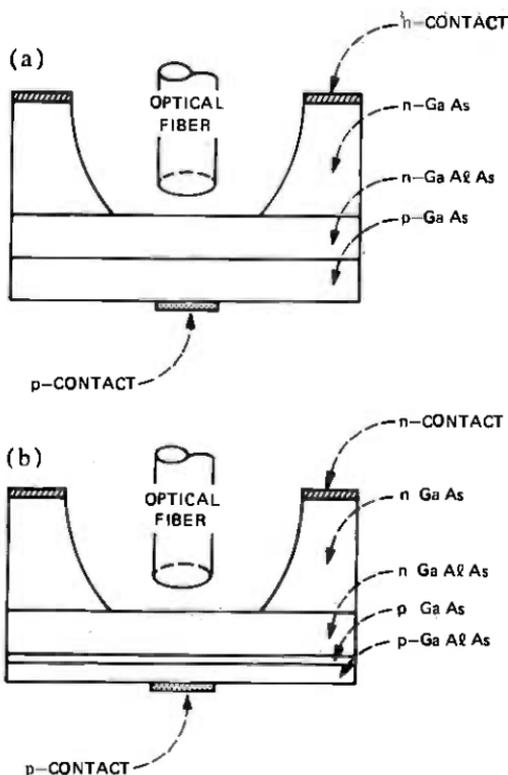


Fig. 1—(a) Schematic of Burrus-type SH LED. (b) Schematic of Burrus-type DH LED.

eters, roughly the electron diffusion length in the p-active layer, is found experimentally to be the optimum active layer thickness.

For the SH LED, the fundamental bandwidth is set by the active layer minority carrier lifetime which depends strongly on the doping density.⁴ For increased doping density, the minority carrier lifetime decreases (i.e., bandwidth increases), but the internal quantum efficiency of the LED also decreases.⁴ To optimize the SH LED performance, the active layer doping density may be chosen so that the minority carrier lifetime is longer than the value necessary to meet the bandwidth requirement. Nonradiative recombination at the p-contact, a result of the compromise in active layer thickness, increases the bandwidth to the required value.

Applying analysis similar to those used for the SH LED, the DH LED is found to be more efficient for the following reasons. First, as seen in Fig. 1b, the active layer of the DH LED is bounded by heterojunctions which have low interface recombination velocities. Second, very thin active and p-GaAlAs carrier confinement layers can be grown so as not to decrease current crowding. For the thickness of 0.7 μm for the active layer and 2 μm for the p-GaAlAs layer, the improved current crowding and carrier confinement of the DH LED results roughly in a factor of 2.5 superior performance compared to the SH LED. Next, when the injected carrier density exceeds the active region doping density, the radiative recombination rate increases with the current density.⁵ This condition, referred to as conductivity modulation or bimolecular recombination, can easily occur only in the DH LED with adequate carrier confinement. As a result, the DH active layer doping density can be chosen at a value well below that necessary to meet the speed requirement to take advantage of the higher internal efficiency; conductivity modulation is used to obtain the higher bandwidth. At the lower doping density ($1 \times 10^{17} \text{ cm}^{-3}$), the DH LED has twice the internal quantum efficiency of the higher doped ($1 \times 10^{19} \text{ cm}^{-3}$) SH LED.⁴ Our analysis thus accounts for a total factor of 5 between the performance of the SH and DH LEDs.

Detailed discussion of the above points is contained in the subsequent sections.

II. MATERIAL-RELATED PARAMETERS

In this section, we discuss the selection of active layer hole concentration for ~ 70 -MHz bandwidth SH and DH LEDs and the effect of the hole concentration on the internal efficiency of the LEDs.

2.1 Lifetime

For T3 data rate (45-Mb/s) optical communication systems, the LED should have an approximate 70-MHz bandwidth (pulse decay time $\tau \approx$

2.3 ns).^{*} The higher speed is needed to allow for receiver response time and dispersion in the optical fiber.⁶ To obtain this large bandwidth, the active layer must be chosen with a minority carrier lifetime equivalent to the required bandwidth, or the device structure must be appropriately designed to increase the speed. For a SH LED, the device time constant can only be reduced by nonradiative recombination, e.g., surface recombination. On the other hand, nonradiative recombination or bimolecular recombination (conductivity modulation) may be used to increase the bandwidth of a DH LED. This mechanism also minimizes the loss due to interface recombination by reducing the diffusion length.

Figure 2 shows the decay time of DH LEDs as a function of active layer hole concentration. This figure is taken from Nelson.⁴ The decay times shown are essentially structure-independent, since the effects of surface recombination and conductivity modulation were minimized. These values are somewhat larger than the minority carrier lifetime because of the effects of photon recycling, but they can be considered to be minimum bulk lifetimes.⁴

From Fig. 2, the active layer hole concentration of SH LEDs should be chosen in the range $p \approx 8 \times 10^{18} \text{ cm}^{-3}$ to have a 70-MHz bandwidth. Due to the trade-offs between current crowding and surface recombination, the range $p \approx 6 \times 10^{18} \text{ cm}^{-3}$ is allowable. In optimizing the active layer width, an increase in bandwidth and a reduction in efficiency due to surface recombination is also obtained. For 70-MHz DH LEDs, a lower active layer hole concentration range ($p \approx 10^{17} \text{ cm}^{-3}$) can be used. A hole concentration below 10^{18} cm^{-3} is desirable for increasing the internal efficiency, while bimolecular recombination which maintains bulk efficiency is used to decrease the decay times from the values shown in Fig. 2.

2.2 Internal efficiency

Figure 3 shows the bulk efficiency of p-GaAs decreasing rapidly for $p > 10^{18} \text{ cm}^{-3}$ at both high and low carrier injection. This figure is also taken from Nelson,⁴ and the details of obtaining the data may be found there. For both low and high electron injection, the bulk efficiency at $p = 1 \times 10^{17} \text{ cm}^{-3}$ is approximately twice that at $p = 1 \times 10^{19} \text{ cm}^{-3}$. The decrease in bulk efficiency at high doping density is presumably due to the introduction of nonradiative recombination centers.⁴ Thus, since the active layer of SH LEDs must be doped approximately 10^{19} cm^{-3} to obtain a 70-MHz bandwidth while only 10^{17}

^{*} The pulse decay time (τ) is the time for the LED light output to decay to $1/e$ of its peak value. τ is related to the 90-percent to 10-percent fall time $\tau_{90\%-10\%}$ by the relation: $\tau_{90\%-10\%} = 2.2\tau$.

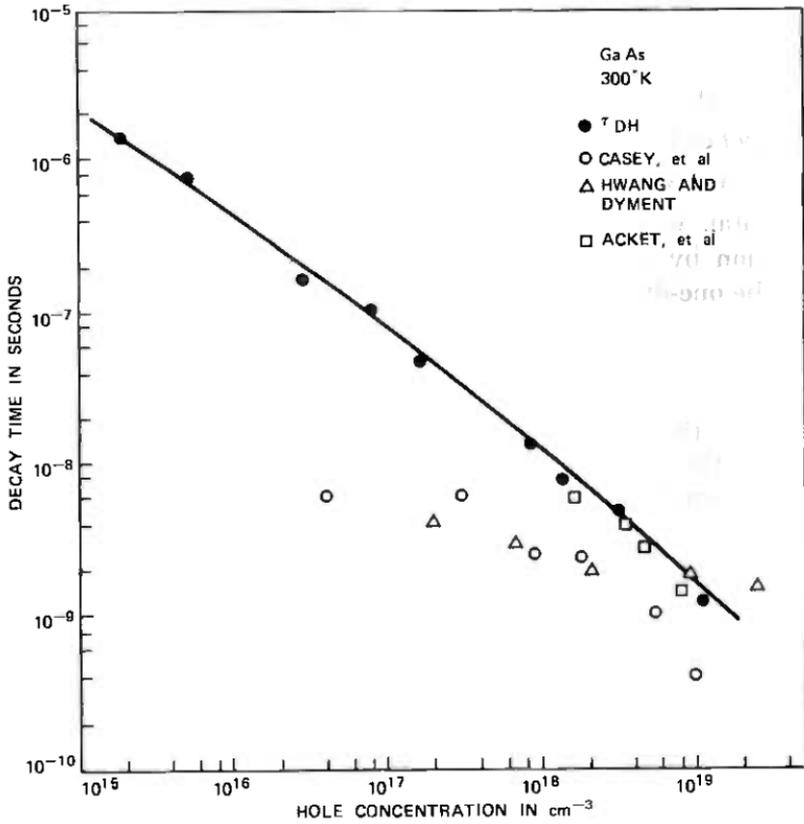


Fig. 2—Variation of photo-luminescent decay times with doping level (from Ref. 4).

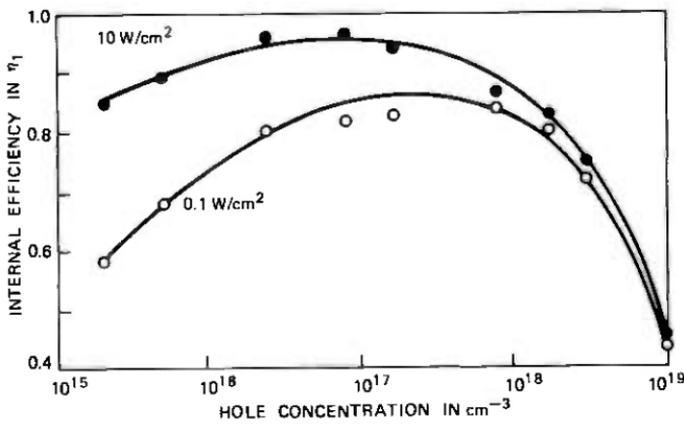


Fig. 3—Variation of internal quantum efficiency with doping level at two different excitation rates (from Ref. 4).

cm^{-3} hole concentration is required for DH LEDs, the DH LED should have twice the efficiency of the SH LED.

III. STRUCTURAL PARAMETERS

3.1 Effect of surface recombination and self absorption

Figure 4 shows a DH structure and two SH structures. Following Lee and Dentai,⁷ we estimate the effects of surface recombination and self-absorption by using a one-dimensional diffusion model. At steady state, the one-dimensional continuity equation is given by

$$D \frac{d^2 n}{dx^2} - \frac{n}{\tau} = 0, \quad (1)$$

where n is the excess electron density, D is the electron diffusivity, and τ is the bulk electron lifetime.⁷ As shown in the figure, the boundary conditions for each of the three structures are

DH

$$-\left. \frac{dn}{dx} \right|_{x=0} = \frac{J}{eD} - \frac{s}{D} n(0) \quad (2)$$

$$-\left. \frac{dn}{dx} \right|_{x=w} = \frac{s}{D} n(w) \quad (3)$$

SH₁

$$-\left. \frac{dn}{dx} \right|_{s=0} = \frac{J}{eD} \quad (4)$$

$$-\left. \frac{dn}{dx} \right|_{x=w} = \frac{s}{D} n(w) \quad (5)$$

SH₂

$$-\left. \frac{dn}{dx} \right|_{x=0} = \frac{J}{eD} - \frac{s}{D} n(0) \quad (6)$$

$$n(w) = 0, \quad (7)$$

where w is the active layer width, s is the interfacial recombination velocity at the heterojunction interfaces, and J is the current density at the p-n junction. Table I lists $n(x)$, the solution to the diffusion equation, for the three structures. Taking into account self-absorption in the active region, the light intensity (P) from the p-n junction side of the LED is given by

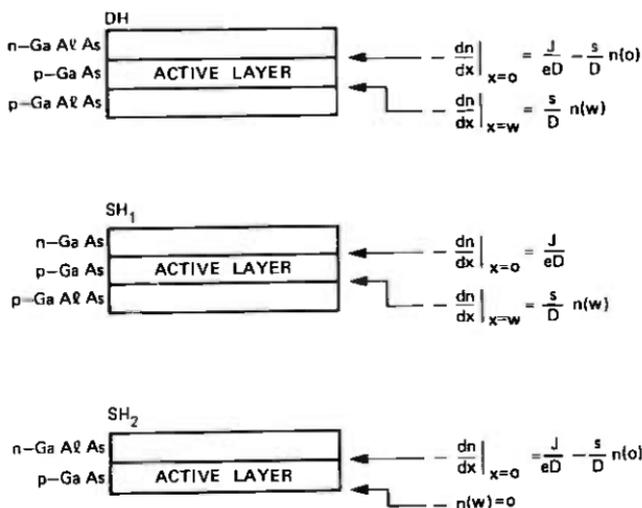


Fig. 4—Boundary conditions for a DH LED and two SH LEDs.

Table I

	DH, SH ₁	SH ₂
$n(x)$	$\frac{JL}{eD} \left\{ \frac{\cosh\left(\frac{w-x}{L}\right) + S \sinh\left(\frac{w-x}{L}\right)}{DEN} \right\}$	$\frac{JL}{eD} \frac{\sinh\left(\frac{w-x}{L}\right)}{DEN}$
$\frac{\tau_{eff}}{\tau}$	$\frac{\sinh\left(\frac{w}{L}\right) + S \left[\cosh\left(\frac{w}{L}\right) - 1 \right]}{DEN}$	$\frac{\cosh\left(\frac{w}{L}\right) - 1}{DEN}$
η_x	$\frac{1}{2 DEN} \left\{ \frac{1+S}{1+\alpha L} \left[1 - e^{-\frac{w}{L}(1+\alpha L)} \right] e^{\frac{w}{L}} - \frac{1-S}{1+\alpha L} \left[1 - e^{\frac{w}{L}(1-\alpha L)} \right] e^{-\frac{w}{L}} \right\}$	$\frac{1}{2 DEN} \left\{ \frac{1}{1+\alpha L} \left[1 - e^{-\frac{w}{L}(1+\alpha L)} \right] e^{\frac{w}{L}} - \frac{1}{1-\alpha L} \left[1 - e^{\frac{w}{L}(1-\alpha L)} \right] e^{-\frac{w}{L}} \right\}$

$$S = \frac{Ls}{D}$$

$$DEN_{DH} = (S^2 + 1) \sinh\left(\frac{W}{L}\right) + 2S \cosh\left(\frac{W}{L}\right)$$

$$DEN_{SH_1} = \sinh\left(\frac{W}{L}\right) + S \cosh\left(\frac{W}{L}\right)$$

$$DEN_{SH_2} = \cosh\left(\frac{W}{L}\right) + S \sinh\left(\frac{W}{L}\right)$$

$$P = \frac{1}{2} \frac{h\nu}{\tau_r} \left\{ \int_0^w n(x)e^{-\alpha x} dx + Re^{-2w\alpha} \int_0^w n(x)e^{\alpha x} dx \right\}. \quad (8)$$

The second term in the equation is the reflected power from the contact. Absorption losses due to the additional layer for the DH and SH₁ case are neglected. The emission frequency is ν , the contact reflectivity is R , the radiative lifetime is τ_r , and the absorption coefficient is α . The external efficiency of the LED is defined by

$$\eta_s = \frac{P}{J} \left(\frac{e}{h\nu} \right) \frac{\tau_r}{\tau}. \quad (9)$$

Table I also lists η_s for the three structures.

Comparing the solutions for SH₁ and DH diodes, only a small difference is noted for $sL/D \ll 1$. In the range of thickness to which these equations are applied, the discussions concerning interface recombination and self-absorption for the DH device refer also to the SH₁ device. The one disadvantage of the SH₁ structure relative to the DH structure is the lack of carrier confinement at the homojunction. Thus SH₁ is equivalent to DH under conditions where conductivity modulation is not important, but becomes less efficient at high carrier injection. The SH LED discussed in the following sections refers only to SH₂.

Figure 5a displays the calculated SH external efficiency as a function of the active layer width for several diffusion lengths; Fig. 5b shows similar curves for the DH case. The calculation parameters are listed on the figures. Values for $1 \times 10^{17} \text{ cm}^{-3}$ and $1 \times 10^{19} \text{ cm}^{-3}$ p-doping are chosen for the DH and SH case, respectively, for comparison with fabricated devices.

The fabricated SH LEDs are analyzed using the 5- μm SH efficiency curve in Fig. 5a, since a 5- μm minority carrier diffusion length was measured on 10^{19} cm^{-3} p-type GaAs using electron-beam-induced current (EBIC).⁸ A peak in the SH efficiency results from the high p-contact surface recombination for small w and self-absorption for large w . The peak efficiency of 0.48 at $w = 12 \mu\text{m}$ is inconsistent with the experimentally determined optimum layer thickness of 5 μm . Better agreement between calculation and experiment is found when current crowding is considered.

For the DH case, an 11- μm diffusion length is derived from the device decay time of ~ 3 ns. The high output power of the DH LED with $w = 0.5 \mu\text{m}$ is consistent with the high calculated efficiency shown in Fig. 5b. This consistency is maintained when current crowding is considered.

3.2 Device bandwidths

The effective device lifetime can be obtained from the carrier concentration by averaging over the active region:⁷

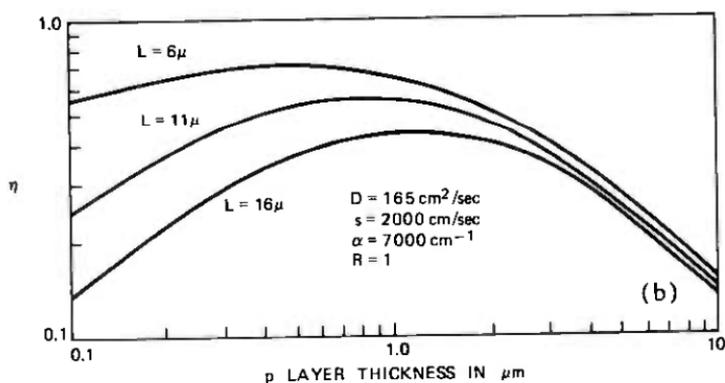
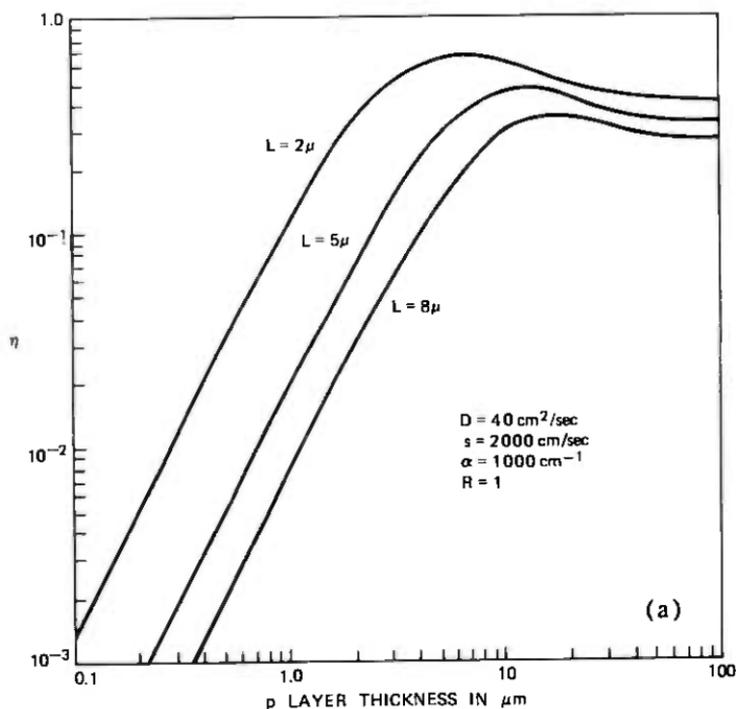


Fig. 5—(a) SH efficiency vs p-layer thickness for indicated electron diffusion lengths. Material parameters are for 10^{19} cm^{-3} doping. (b) DH efficiency vs p-layer thickness for indicated electron diffusion lengths. Material parameters are for 10^{17} cm^{-3} doping.

$$\bar{n} = \frac{1}{w} \int_0^w n(x) dx \equiv \frac{J \tau_{eff}}{e w} \quad (10)$$

τ_{eff}/τ is listed in Table I for the three structures of Fig. 3.

Figure 6a plots the calculated values of τ_{eff}/τ as a function of p-layer thickness for the SH₂ LED. The material parameters in the calculation are those of 10^{19} cm^{-3} Ge-doped GaAs for comparison with fabricated SH LEDs. The diffusion length is varied to show the effect of surface

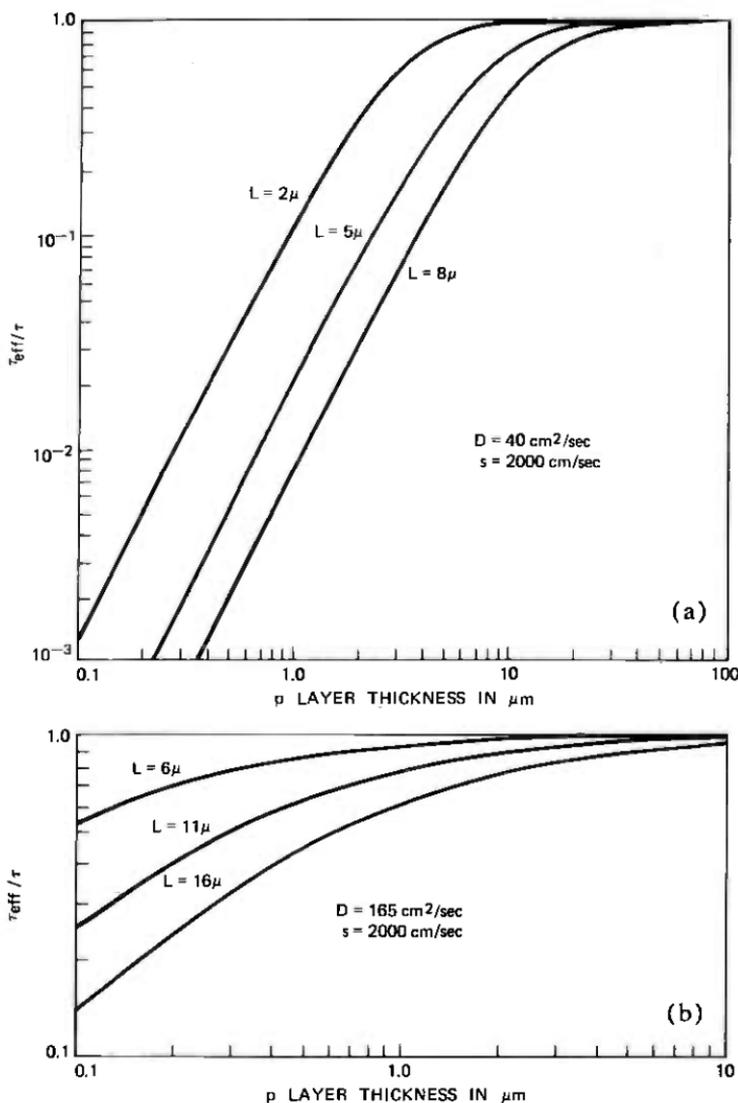


Fig. 6—(a) Normalized $s\tau$ effective decay time vs p-layer thickness for indicated electron diffusion lengths. Material parameters are for 10^{19} cm^{-3} doping. (b) Normalized $s\tau$ effective decay time vs p-layer thickness for indicated electron diffusion lengths. Material parameters are for 10^{17} cm^{-3} doping.

recombination. For a 5- μm diffusion length and p-layer thickness of 5 μm , surface recombination reduces the decay time by 0.35. Using the 1- to 2-ns decay time for 10^{19} cm^{-3} GaAs from Fig. 2, a device decay time of 0.35 to 0.7 ns is obtained. This value appears to be in disagreement with the measured value of 2.5 ns, but the device lifetime is probably the result of photon recycling in the thick p-layer.⁴

Figure 6b plots the calculated values of τ_{eff}/τ as a function of p-layer thickness for the DH LED. The material parameters of 10^{17} cm^{-3} Ge-doped GaAs are used in the calculation for comparison with fabricated devices. The diffusion length is again varied to show the effect of surface recombination. A diffusion length of $11 \mu\text{m}$ is obtained from the bimolecular recombination time constant of 7.3 ns and a diffusivity of $165 \text{ cm}^2/\text{s}$.⁹ The bimolecular recombination time for our DH LEDs is given by⁷

$$\tau_{BR} = B \left(\frac{J}{ewB} \right)^{-1/2}, \quad (11)$$

where $B = 5 \times 10^{-11} \text{ cm}^3/\text{s}$ is the recombination probability, $w = 0.5 \mu\text{m}$ is the active layer width, and $J = 3000 \text{ A/cm}^2$ is the current density. τ_{BR} is used in the analysis rather than the decay time in Fig. 1, since the injected carrier density

$$\Delta n = \left(\frac{J}{ewB} \right)^{1/2} = 2.8 \times 10^{18} \text{ cm}^{-3} \quad (12)$$

is an order of magnitude greater than the doping concentration. From Fig. 6b, at $0.5\text{-}\mu\text{m}$ active layer width, τ_{BR} is reduced by a factor of 0.63 due to interface recombination to give a device decay time of $\sim 4.5 \text{ ns}$. This value is consistent with the $\sim 3.5\text{-ns}$ decay time measured on DH LEDs at 60 mA ($J = 3000 \text{ A/cm}^2$) forward bias.

3.3 Current crowding

The simplest method of coupling a LED to an optical fiber is to butt the fiber end directly to the LED. To maximize this coupling, the light-emitting area should be limited to the fiber core size. The graded-index fiber used for the present study has a $55\text{-}\mu\text{m}$ diameter core. The SH and DH LEDs were fabricated with a $50\text{-}\mu\text{m}$ contact to localize the light-emitting area.

The SH and DH LEDs are Burrus-type diodes where a well has been opened in the opaque n-GaAs substrate to access the generated light. The current is confined to flow through the $50\text{-}\mu\text{m}$ contact by a SiN_x insulating coating. The current density and distribution in the active region is determined by the active layer width and the sheet resistivity of the active layer. A thicker layer width or lower sheet resistivity results in a lower current density above the contact and a larger emitting area. This current spreading reduces the coupling efficiency and, in the DH case, the bimolecular recombination. An improvement may be made by reducing the contact dimension, but a lower limit of $25 \mu\text{m}$ is set by the use of an evaporation mask to define the contacts.

Using the technique described by Joyce and Wemple,¹⁰ the calcu-

lated current density (J) at the p-n junction above the contact is plotted as a function of p-layer thickness in Fig. 7. The calculation parameters are indicated on the figure. The current density is a strong function of the p-layer thickness; the current density is half the maximum value of 2500 A/cm^2 by $w \sim 1.5 \mu\text{m}$. The other two curves in the figure are the product of the current density times the calculated external efficiency ($\eta \times J$) of the SH and DH LEDs. The effect of the bottom confinement layer for the DH structure is neglected in the current density calculation, since the figure is meant to show only the approximate behavior with thickness.

The calculation of the SH efficiency uses the 10^{19} cm^{-3} Ge-doped GaAs parameters (listed on the figure) for comparison with fabricated devices. The absorption coefficient and interface recombination velocity are from Ref. 7, and the electron diffusivity is obtained from Ref. 9. The electron diffusion length was measured using the electron beam induced current technique.⁸ As shown in Fig. 7, the $\eta_{\text{SH}} \times J$ curve shows a peak for an active layer width between 2 and $3 \mu\text{m}$. $\eta_{\text{SH}} \times J$ is limited by surface recombination for thin layers and self-absorption or current spreading for thick layers. The calculated optimum active

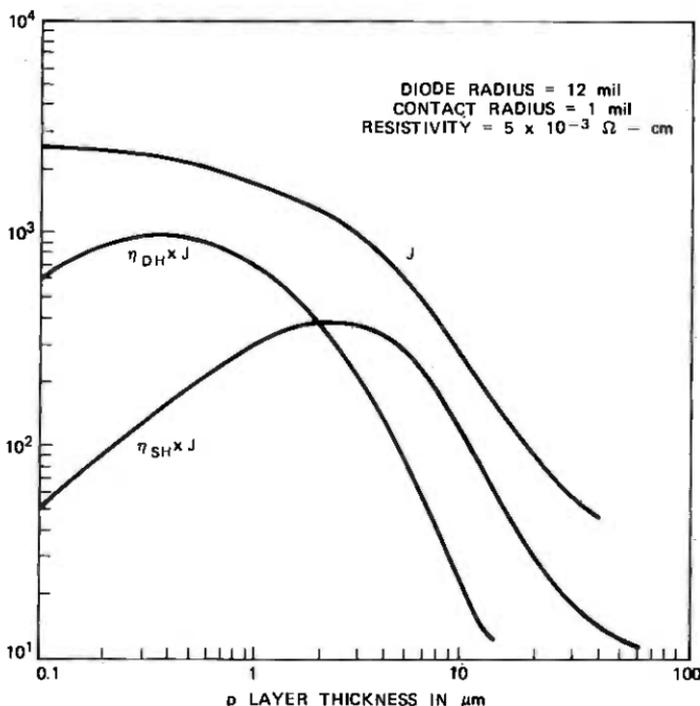


Fig. 7— J , $\eta_{\text{DH}} \times J$, and $\eta_{\text{SH}} \times J$ vs p-layer thickness. J is the current density for the diode parameters indicated at a forward current of 60 mA, η_{DH} is the DH efficiency for an 11- μm diffusion length from Fig. 5b, and η_{SH} is the SH efficiency for a 5- μm diffusion length from Fig. 5a.

layer thickness is lower than the experimentally determined value of $\sim 5 \mu\text{m}$ for SH LEDs. However, small variations in the many variables that went into the calculation may easily shift the optimum value to $\sim 5 \mu\text{m}$.

In the $\eta_{\text{DH}} \times J$ calculation, the optimum active layer thickness is determined to be between 0.2 and 0.7 μm . The material parameters corresponding to 10^{17}cm^{-3} Ge doping were used. The diffusion length was corrected for the effects of bimolecular recombination. A peak occurs in $\eta_{\text{DH}} \times J$ for the reasons given in the SH case. The 0.5- μm active layer width of the fabricated DH LEDs lies within the calculated optimum range. From Fig. 7, $\eta_{\text{DH}} \times J$ for $w = 0.5 \mu\text{m}$ is approximately 2- $\frac{1}{2}$ times the peak value of $\eta_{\text{SH}} \times J$.

IV. SUMMARY

We have compared SH and DH LEDs designed for optical communications at T3 data rate. Our analysis shows a factor-of-2 loss in the internal efficiency of the SH LED due to the introduction of nonradiative recombination centers at the high active layer doping level required by the SH devices. An additional factor-of-2.5 loss was found in the external efficiency of the SH LED, resulting from surface recombination at the p-contact and the self absorption and current spreading in the thicker active layer of the SH LED. The total factor-of-5 difference between SH and DH LEDs based on the arguments in this paper is in reasonable agreement with the experimentally determined value of 7 to 8, since the calculations involved many variables. These results demonstrate that DH LEDs are required for data links limited by power launched into the optical fiber.

V. ACKNOWLEDGMENTS

One author (AKC) is indebted to O. G. Lorimor and R. J. Nelson for many discussions on diode parameters. We would also like to acknowledge helpful discussions with R. H. Saul, T. P. Lee, H. J. Leamy, W. H. Hackett, F. R. Nash, and S. E. Haszko.

REFERENCES

1. C. A. Burrus and B. I. Miller, *Opt. Commun.*, **4** (1971), p. 307.
2. V. G. Keramidis, G. W. Berkstresser, and C. L. Zipfel, unpublished work.
3. C. L. Zipfel, R. H. Saul, and V. G. Keramidis, paper presented at DRC, Boulder, Colorado (1979).
4. R. J. Nelson and R. G. Sobers, *J. Appl. Phys.*, **49** (1978), p. 6103.
5. H. Namizaki, H. Kay, M. Ishii, and A. Ito, *Appl. Phys. Lett.*, **24** (1974), p. 486.
6. H. Temkin, private communication.
7. T. P. Lee and A. G. Dentai, *IEEE J. Quantum Elec.*, **14** (1978), p. 150.
8. A. K. Chin, unpublished work.
9. M. Ettenberg, H. Kressel, and S. L. Gilbert, *J. Appl. Phys.*, **44** (1973), p. 827.
10. W. B. Joyce and S. H. Wemple, *J. Appl. Phys.*, **41** (1970), p. 3818.

Adaptive Echo Cancellation/AGC Structures for Two-Wire, Full-Duplex Data Transmission

By D. D. FALCONER and K. H. MUELLER

(Manuscript received March 2, 1979)

Three different receiver arrangements are studied, all of which incorporate provision for joint adaptive echo cancellation and gain adjustment to provide two-wire full-duplex data communication. In each case, the canceler consists of a data-driven transversal filter, but the architectures differ in the way the gain adjustment is provided. For all architectures, we investigate the properties of a joint adaptive LMS algorithm based on the receiver's decisions on far-end data symbols and present appropriate computer simulations. We show that an arrangement where the gain control adjusts the reference level after the decision detector output performs significantly better than AGC schemes attempting to adjust the level of the analog signal.

I. INTRODUCTION

High-speed full-duplex data communication on a single channel is of immense practical interest. Data transmission via the DDD telephone network and the possibility of future digital subscriber lines are two of the most challenging applications. Techniques for achieving this goal fall in essentially three categories: Frequency Division Multiplexing (FDM), Time Division Multiplexing (TDM), and echo cancellation. Only echo cancellation allows full-bandwidth continuous use of the channel in each direction. This scheme therefore offers the highest potential bit rates.

The transmitter and receiver are jointly coupled to a two-wire line via a hybrid. In an environment of changing channel characteristics (e.g., switched network), the hybrid balancing, if fixed, will at best provide a compromise match to the channel. In this mode, a vestige of the local transmitted signal, leaking through the hybrid, can be expected to interfere with the incoming signal from the far-end simultaneously operating transmitter. Figure 1 shows the system under discussion, and Fig. 2 models the signals entering and leaving a two-wire

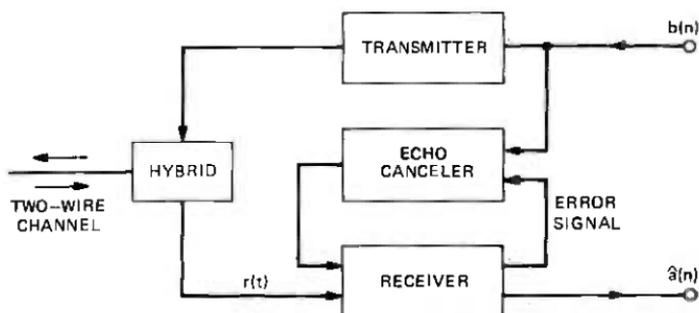


Fig. 1—Basic system configuration.

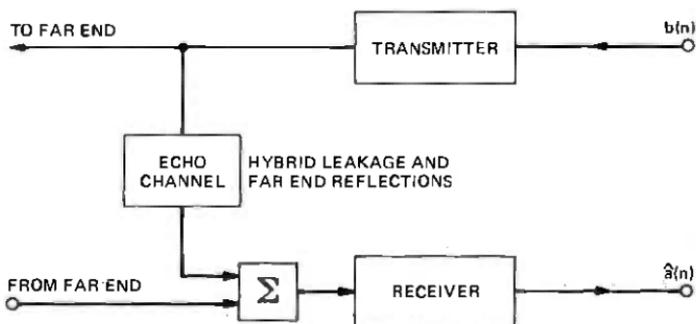


Fig. 2—System model (without canceler).

full-duplex modem. The local transmitter transmits a sequence of data symbols $\{b(n)\}$ at T -second intervals as a PAM data waveform. The received waveform $r(t)$ consists of a PAM data waveform with data symbols $\{a(n)\}$ transmitted also at T -second intervals from the *distant* end, plus noise, plus the interfering vestige of the locally transmitted signal. This interfering signal (which we shall refer to as the *echo signal* or *echo component*) may have power comparable to or even greater than that of the desired far-end signal component.

Decisions on the $\{a(n)\}$ are made by quantizing the sampled receiver output to ± 1 in the case of binary data, or to one of M values in the case of M -level data. A typically encountered echo component arising in a system with a conventional compromise balanced hybrid will cause an unacceptably high error rate.

To remove the interfering echo component, the local receiver must perform echo cancellation; that is, estimate the echo signal and subtract it from the incoming signal prior to making decisions, as shown in Figs. 1 and 3a. The estimate is a transversally filtered version of the local data symbols $\{b(n)\}$ * as proposed in Ref. 1. If the $\{b(n)\}$ are

* The $\{b(n)\}$ may be different from the user data since they are defined to include such operations as differential encoding or scrambling.

binary, the implementation is simple, requiring only additions and subtractions. The transversal filter tap coefficients $\{p_m\}$ should approximate the samples of the impulse response of the combination of the local transmitter and the echo path.

Equivalently,¹ the tap coefficients should be chosen to minimize, in a mean-square sense, the measured receiver error signal which is the difference between the actual receiver output $y(n)$ and the ideal output. This error is available at each sampling instant of the received data. The subtraction of the (decision-directed) reference $\{\hat{a}(n)\}$ is, of course, what makes it possible to adapt quickly even in the presence of doubletalk. Such adaptation allows tracking time-varying components of the echo channel or coping with larger call-to-call variations in a switched system. However, since the level of the received signal is likely to vary significantly under those conditions, it is essential that proper scaling be done when the error signal is computed. Such scaling involves a gain adjust device which must operate jointly and adaptively with the echo canceler. This paper deals with this joint adaptation

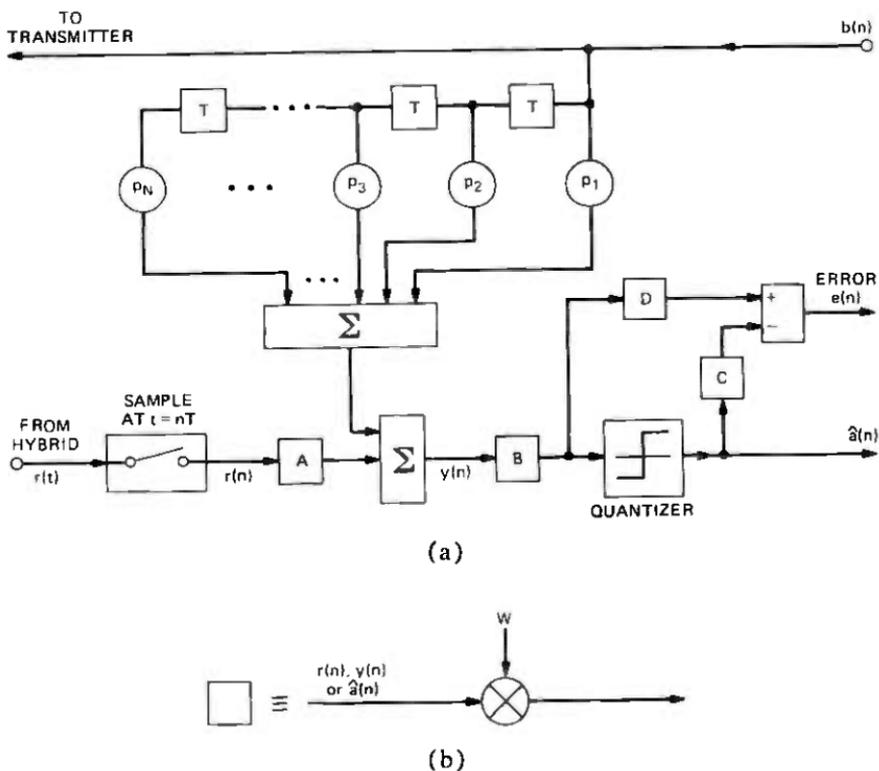


Fig. 3—Receiver structure with echo canceler and automatic gain control. (a) Overall system showing various locations for AGC function. (b) Basic AGC functions in one of above marked locations A, B, C, or D.

problem. Although we perform our study for a system operating at baseband, our results can be applied to passband structure via appropriate redefinitions.

As shown in Fig. 3a, the error $e(n)$ is the difference between the receiver's decision $\hat{a}(n)$ (assumed correct) and the receiver's output $y(n)$. But if the end-to-end channel gain is α , $y(n)$ consists of $\alpha a(n)$ plus possibly noise and uncanceled echo. Intersymbol interference is not treated in this study. It is expected to be of secondary concern in typical subscriber cable systems, but we realize that it cannot be neglected for high-speed DDD applications. The error $e(n)$ in the absence of gain control would contain the term $(\alpha - 1)a(n)$, which is relatively large if α differs significantly from unity. Therefore, a small steady-state mean-squared error, with consequent minimal fluctuation of the canceler tap coefficients $\{p_n\}$ and low error rate, is possible *only if the nonunity gain α is compensated for by an AGC adjusted to provide a gain approximating $1/\alpha$* . The gain of the AGC, denoted w , will be considered to be adjusted jointly with the echo canceler tap coefficients to minimize the mean squared error.

Shown in Fig. 3a are four possible locations, A, B, C, and D, for placement of the AGC whose functional form is depicted in Fig. 3b. Since D is identical to B as long as the signal is binary and the quantizer is ideal, only three variations need to be examined. The corresponding three receiver architectures, labeled naturally A, B, and C, may differ significantly in their adaption speeds. The main theme of this paper is the convergence of each of the three receiver arrangements, and how it is influenced by channel parameters, such as the relative powers of echo and far-end components.

The study concludes that arrangement C which features an automatic reference control (ARC) at the quantizer output offers the fastest convergence rate, together with the simplest implementation. Arrangements A and B may suffer slower rates of convergence due to coupling interaction between the AGC and echo canceler tap coefficient adaptation. However, for arrangement B, a simple step-size modification is proposed, which on the average decouples the AGC and echo canceler adaptations, thus improving its convergence rate. Readers interested in C, the "best" arrangement, may skip Sections III to V, which deal with A and B.

Mention of earlier related work is in order at this point. Adaptive echo cancellation without a jointly adapting AGC or equalizer for two-wire full duplex data communication has been treated in Refs. 1 to 4. References 1 and 4 treat essentially the echo-cancellation system proposed here, assuming α is known, and thus omit the AGC. This type of scheme, in which the echo canceler's input consists only of local data symbols, offers obvious simplicity of implementation. In Ref. 2, a

voice-type canceler is investigated, and Ref. 3 discusses an external data-driven structure which cancels the entire waveform by computing compensation samples at a high enough rate.* Reference 5 summarizes work reported in Ref. 1 and 3, and discusses a further receiver arrangement comprising an adaptive echo canceler and adaptive equalizer for mitigating end-to-end linear distortion. Arrangement A considered here, the so-called "convex canceler," is a special case of this, since an AGC can be considered a one-tap linear equalizer. Reference 6 describes an echo-cancelling receiver structure incorporating decision feedback equalization. Finally, our arrangement C combined with decision feedback equalization has recently been proposed in Ref. 7.

II. SYSTEM MODELING

We shall examine the convergence properties of several adaptation strategies for the three receiver arrangements. Each follows a decision-directed approach; AGC and echo canceler parameters are adjusted once per symbol interval, based on the observed error between the unquantized receiver output and the decision $\hat{a}(n)$ (for arrangements A and B) or $w\hat{a}(n)$ (for arrangement C). For purposes of analysis, the decisions $\hat{a}(n)$ are assumed equal to the actual data symbols $a(n)$. The convergence of decision-directed adaptive receivers making small adjustments at each iteration has been found to be negligibly affected by occasional decision errors.

The analysis is based on a simple linear model of the end-to-end channel and leakage paths: the signal $r(t)$ entering the receiver from the hybrid will be written as

$$r(t) = \sum_n a(n)g_A(t - nT) + \sum_n b(n)g_B(t - nT) + \nu(t). \quad (1)$$

The first summation represents the signal from the far end, and $g_A(t)$ is the end-to-end channel impulse response, including receiver front-end filtering. The second summation is the echo signal, and $g_B(t)$ is the impulse response of the echo path. The symbol $\nu(t)$ is a waveform of additive white noise. The symbol interval T is equal in both summations. This is tantamount to assuming that the near- and far-end transmitters are synchronized in clock frequency.¹

Decisions $\hat{a}(n)$ are made by quantizing samples of the receiver's output to ± 1 in the case of binary data. The end-to-end channel impulse response and the phase of the receiver's sampling clock are assumed ideal, so that no intersymbol interference is present in the samples of $r(t)$; i.e. the n th sample is of the form

* In Ref. 2, the canceler input is the sampled transmitted waveform. In Ref. 3, it is the $\{b(n)\}$.

$$r(nT) = a(n) \alpha + \sum_k b(k) \beta_{n-k} + \nu(nT), \quad (2)$$

where $\alpha \equiv g_A(0)$, $g_A(nT) = 0$ for $n \neq 0$, and $\beta_n \equiv g_B(nT)$. In subsequent notation, we write $r(nT)$ and $\nu(nT)$ as $r(n)$ and $\nu(n)$, respectively. The binary (± 1) data symbols $\{a(n)\}$ and $\{b(n)\}$ are statistically independent. The noise samples $\nu(n)$ are assumed to be independent with zero mean and variance σ^2 .

III. ARRANGEMENT A

We recall from Fig. 3a that arrangement A forms the receiver output as the sum of the echo canceler and AGC outputs. The receiver output is a linear function of the echo canceler tap coefficients and AGC gain. Thus the mean-squared error at the receiver's output is a convex quadratic function of the receiver parameters, and a simple gradient algorithm can be used with confidence to adjust the parameters jointly. As mentioned before, if the AGC is replaced by an adaptive linear equalizer, the arrangement A generalizes to the jointly adaptive echo canceler equalizer structure discussed in Ref. 5.

Given an AGC gain w and a set of N echo canceler tap coefficients $\{p_k\}_{k=1}^N$, the receiver's unquantized output sample $y(n)$ is

$$y(n) = w r(n) + \sum_{k=1}^N p_k b(n-k). \quad (3)$$

Define the N -dimensional vectors

$$\mathbf{p} \equiv \begin{bmatrix} p_1 \\ \cdot \\ \cdot \\ \cdot \\ p_N \end{bmatrix} \quad \text{and} \quad \mathbf{b}(n) \equiv \begin{bmatrix} b(n-1) \\ \cdot \\ \cdot \\ \cdot \\ b(n-N) \end{bmatrix},$$

and the $(N+1)$ -dimensional vectors in partitioned form as

$$\mathbf{c} \equiv \begin{bmatrix} w \\ \mathbf{p} \end{bmatrix} \quad \text{and} \quad \mathbf{z}(n) \equiv \begin{bmatrix} r(n) \\ \mathbf{b}(n) \end{bmatrix}.$$

Then (3) is written more compactly as

$$y(n) = \mathbf{c}^\dagger \mathbf{z}(n), \quad (4)$$

where \dagger denotes transpose. The vector \mathbf{c} is the set of receiver parameters to be adaptively adjusted, and $\mathbf{z}(n)$ is the current set of inputs stored by the receiver.

The ideal output at time nT would be $a(n)$, and the error is

$$e(n) = y(n) - a(n). \quad (5)$$

The expression for the mean-squared error is

$$\langle e(n)^2 \rangle = \mathbf{c}^\dagger A \mathbf{c} - 2\mathbf{c}^\dagger \mathbf{x} + 1, \quad (6)$$

where A is a $(N + 1)$ by $(N + 1)$ covariance matrix

$$A \equiv \langle \mathbf{z}(n)\mathbf{z}(n)^\dagger \rangle, \quad (7)$$

and \mathbf{x} is an $(N + 1)$ -dimensional vector

$$\mathbf{x} = \langle a(n)\mathbf{z}(n) \rangle. \quad (8)$$

By rewriting (6) as

$$\langle e(n)^2 \rangle = (\mathbf{c} - A^{-1}\mathbf{x})^\dagger A (\mathbf{c} - A^{-1}\mathbf{x}) + 1 - \mathbf{x}^\dagger A^{-1}\mathbf{x}, \quad (9)$$

and recognizing that A by definition is positive semidefinite, it is clear that the mean-squared error has its minimum value

$$e_{\min}^2 = 1 - \mathbf{x}^\dagger A^{-1}\mathbf{x}, \quad (10)$$

when

$$\mathbf{c} = \mathbf{c}_{\text{opt}} = A^{-1}\mathbf{x}. \quad (11)$$

Using the independence assumptions for the data symbols and noise, and the expression (2) for $r(n)$, we readily find that A can be written as

$$A = \begin{bmatrix} A_{00} & \beta^\dagger \\ \beta & I \end{bmatrix}, \quad (12)$$

where

$$A_{00} \equiv \langle r(n)^2 \rangle = \alpha^2 + \sum_k \beta_k^2 + \sigma^2, \quad (13)$$

$$\beta \equiv \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix} \quad (14)$$

denotes the sampled echo impulse response truncated to N samples, and I is the N -dimensional identity matrix. Note that, because of the truncation involved in defining β , $|\beta|^2 \leq \sum_k \beta_k^2$. Similarly, the vector \mathbf{x} is

$$\mathbf{x} = \begin{bmatrix} \alpha \\ \mathbf{0} \end{bmatrix}, \quad (15)$$

where $\mathbf{0}$ is an N -dimensional all-zero vector.

Adaptive adjustment of the receiver parameter vector \mathbf{c} can be accomplished by employing the Widrow-Hoff LMS algorithm⁸ just as in adaptive mean-square equalization.⁹ The current value of \mathbf{c} at time nT , $\mathbf{c}(n)$, is then updated according to

$$\mathbf{c}(n+1) = \mathbf{c}(n) - \gamma e(n)\mathbf{z}(n), \quad (16)$$

where γ is a constant step size. The average value of the correction term $-\gamma e(n)\mathbf{z}(n)$ is proportional to the negative of the gradient of the mean-squared error with respect to \mathbf{c} . Expression (16) portrays the joint updating of the AGC gain and echo canceler taps:

$$w(n+1) = w(n) - \gamma e(n)r(n). \quad (17a)$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma e(n)\mathbf{b}(n). \quad (17b)$$

Adaptation of the echo canceler *alone*, according to (17b) with w fixed at 1, has been analyzed by Mueller.¹ The rate of convergence of $\mathbf{p}(n)$ to β for an optimum choice of γ was shown to be determined only by the number of taps, rather than by the detailed characteristics of the echo path.

The joint updating algorithm (16) resembles equalizer adaptation algorithms, whose convergence behavior has been extensively studied.⁷⁻¹⁶ The convergence of the more general version of (16), for joint echo cancellation and equalization, was discussed in Ref. 5. These theoretical studies have rested on an untrue assumption of independence of successive equalizer or canceler contents. However, a more rigorous analysis in Ref. 16, in addition to experimental results, suggests that the independence assumption does not cause serious error. The aforementioned studies have revealed that, for a fixed step size coefficient γ , the speed of convergence is largely governed by the spread of the eigenvalues of the matrix A defined by (12). Without elaborating on the details, we can say that a ratio of maximum-to-minimum eigenvalues which is close to unity leads to relatively fast convergence, while slow convergence is associated with a maximum-to-minimum eigenvalue ratio which is much greater than unity. If the step size coefficient γ is chosen to effect a judicious compromise between speed of convergence and noise due to random tap fluctuations, then a system with a small eigenvalue spread would typically converge in a number of iterations equal to a small multiple of the number of adjustable tap coefficients.

The $N+1$ eigenvalues of matrix A defined by (12), (13), and (14) are readily found to consist of $N-1$ unit eigenvalues plus λ_{\max} and λ_{\min} , given by

$$\lambda_{\min}^{\max} = 1/2 [1 + \alpha^2 + \sigma^2 + \sum_k \beta_k^2 \pm \sqrt{(1 + \alpha^2 + \sigma^2 + \sum_k \beta_k^2)^2 - 4(\alpha^2 + \sigma^2 + \sum_k \beta_k^2 - |\beta|^2)}], \quad (18)$$

where the plus is associated with λ_{\max} and the minus with λ_{\min} . It is straightforward to show that

$$\lambda_{\min} < 1 < \lambda_{\max}.$$

It is interesting to point out that expression (18) for λ_{\max} and λ_{\min} coincides with the expressions for *bounds* on the maximum and minimum eigenvalues found in Ref. 5 for the more general canceler/equalizer combination. Table I shows values of λ_{\max} and λ_{\min} for various values of $(\alpha^2 + \sigma^2)$ (the power of the far-end signal component plus noise) and $\sum_k \beta_k^2$ (the power of the echo component), assuming $\sum_k \beta_k^2 = |\beta|^2$, so that perfect echo cancellation is possible.

The ratio $\lambda_{\max}/\lambda_{\min}$ increases rapidly as $\alpha^2 + \sigma^2$ decreases. Thus a system with a far-end signal component which is much weaker than the near-end echo component would be expected to converge much more slowly than one with a relatively strong far-end component. This sensitivity of the convergence behavior to the relative strengths of far-end and near-end signal components was also noted in Ref. 5 for the canceler/equalizer system.

IV. ARRANGEMENT B

In this arrangement, combining the received signal and echo canceler output is done ahead of the AGC. This may appear as a more natural arrangement, since the intent is to cancel the echo component at the sampling instants before they enter the portion of the receiver devoted to estimating $a(n)$. The local data symbols are processed by the echo canceler and AGC in tandem; the output is not a linear function of the canceler tap coefficients and AGC gain, and the mean-squared error is not a convex function of these receiver parameters. We can thus call this second arrangement a "nonconvex canceler."

The receiver output is

Table I

Far-End Signal Power Plus Noise Power $\alpha^2 + \sigma^2$	Echo Power $ \beta ^2$	Eigenvalue		Ratio λ_2/λ_1
		λ_{\min}	λ_{\max}	
0.1	1.0	0.0487	2.015	41.4
0.35	1.0	0.1598	2.190	13.7
0.50	1.0	0.2187	2.281	10.4
1.0	1.0	0.3820	2.618	6.85
2.0	1.0	0.5858	3.414	5.83
1.0	2.0	0.2680	3.732	13.9

$$y(n) = w[r(n) + \mathbf{p}^\dagger \mathbf{b}(n)]. \quad (19)$$

With $r(n)$ modeled as in eq. (2), the mean-squared error for given values of w and \mathbf{p} is

$$\begin{aligned} \langle e(n)^2 \rangle &= \langle (y(n) - a(n))^2 \rangle \\ &= w^2 [\alpha^2 + \sum_k \beta_k^2 + \sigma^2 + |\mathbf{p}|^2 + 2 \mathbf{p}^\dagger \boldsymbol{\beta}] \\ &\quad - 2 w \alpha + 1, \end{aligned} \quad (20)$$

which can be written as

$$\begin{aligned} \langle e(n)^2 \rangle &= (w\alpha - 1)^2 + |\mathbf{p} + \boldsymbol{\beta}|^2 w^2 \\ &\quad + w^2 [\sigma^2 + \sum_k \beta_k^2 - |\boldsymbol{\beta}|^2]. \end{aligned} \quad (21)$$

The final term in brackets represents the effect of additive noise and uncancellable echo, if any. Expression (21) can also be written as

$$\langle e(n)^2 \rangle = (\alpha^2 + \delta^2)(w - w_{\text{opt}})^2 + w^2 |\mathbf{p} - \mathbf{p}_{\text{opt}}|^2 + \frac{\delta^2}{\alpha^2 + \delta^2}, \quad (22)$$

where

$$\delta^2 \equiv \sigma^2 + \sum_k \beta_k^2 - |\boldsymbol{\beta}|^2 \quad (23a)$$

and

$$w_{\text{opt}} \equiv \frac{\alpha}{\alpha^2 + \delta^2} \quad (23b)$$

$$\mathbf{p}_{\text{opt}} = -\boldsymbol{\beta} \quad (23c)$$

are the parameter values which minimize $\langle e(n)^2 \rangle$. In practical systems, δ is small and $w_{\text{opt}} \sim 1/\alpha$.

It is instructive to plot contours of constant mean-squared error in the plane whose coordinates are the AGC error

$$e_w \equiv w - w_{\text{opt}} \quad (24)$$

and canceler error magnitude

$$|\mathbf{e}_p| = |\mathbf{p} - \mathbf{p}_{\text{opt}}|, \quad (25)$$

respectively; i.e., curves satisfying

$$\alpha^2 e_w^2 + \left(e_w + \frac{1}{\alpha} \right)^2 |\mathbf{e}_p|^2 = \text{MSE} \quad (26)$$

(assuming δ is negligible), for various positive values of MSE. Such contour plots are shown in Figs. 4 and 5 for values of the end-to-end gain α of 0.5 and 1. Each plot shows 10 contours for MSE ranging from

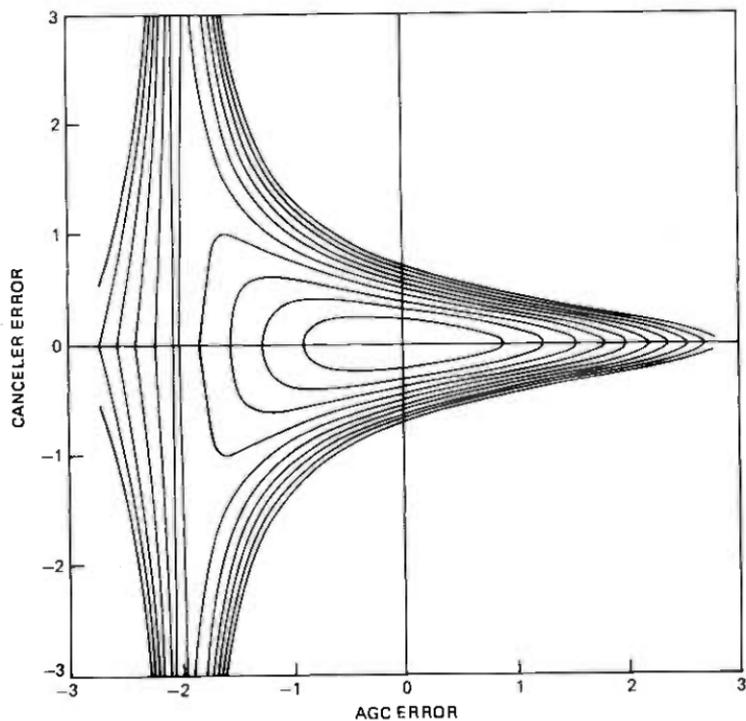


Fig. 4—Contours of constant mean-squared error for arrangement B ($\alpha = 0.5$).

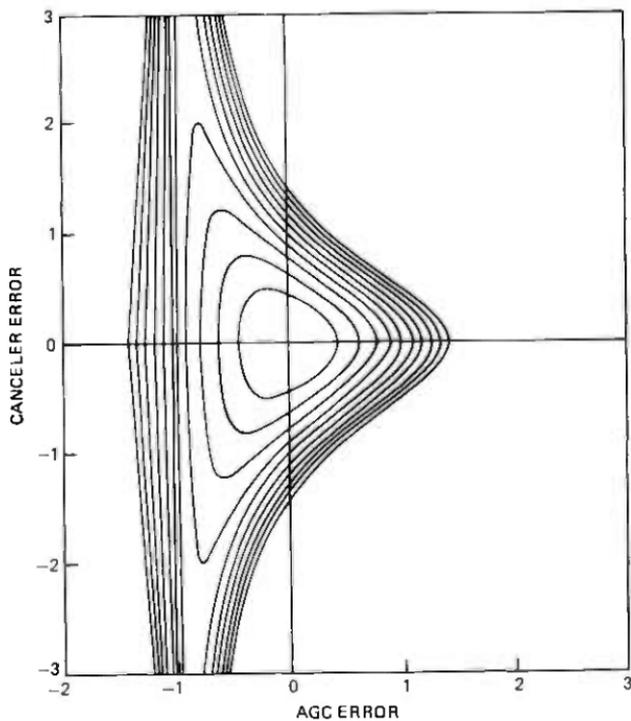


Fig. 5—Contours of constant mean-squared error for arrangement B ($\alpha = 1.0$).

0.2 to 2.0 in steps of 0.2. In each case, there is a unique minimum ($e_w = |e_p| = 0$), but the MSE surface is not convex. Moreover, it exhibits a kind of trough running along the line $e_w = -1/\alpha$. The floor of this trough slopes very gradually toward the origin for large values of $|e_p|$. The shape of the contours for this system differs radically from the ellipsoidal contours that are characteristic of the convex system. Those contours are plotted in Figs. 6 and 7 for the same values of MSE and of α and $|\beta| = 1$, as for the nonconvex system. The ellipses are plotted for convenience for error values mapped onto the two eigenvectors μ_1 and μ_2 of the matrix A . The eccentricity of the ellipse is the ratio $\lambda_{\max}/\lambda_{\min}$. Similarly, the convex system would converge slowly, starting from zero-valued parameters if α is small (implying $\lambda_{\max}/\lambda_{\min} \gg 1$).

A gradient procedure (LMS algorithm) for jointly adjusting w and p should, with proper choice of step size, converge to the optimum parameters. It would, on the average, follow a path perpendicular to the contours that it crosses. Thus very slow convergence of the nonconvex system would be expected if the initial value of all parameters is zero; i.e., starting at $e_w = -(1/\alpha)$, $|e_p| = |\beta|$.

An LMS algorithm for the nonconvex system is obtained by making the correction terms proportional to the negative gradients of the

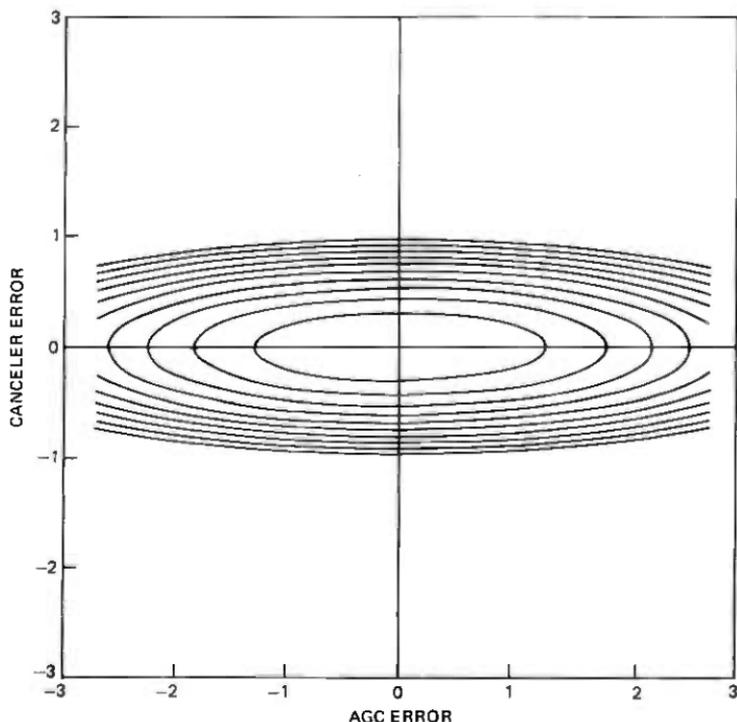


Fig. 6—Contours of constant mean-squared error for arrangement A ($\alpha = 0.5$, $|\beta| = 1$).

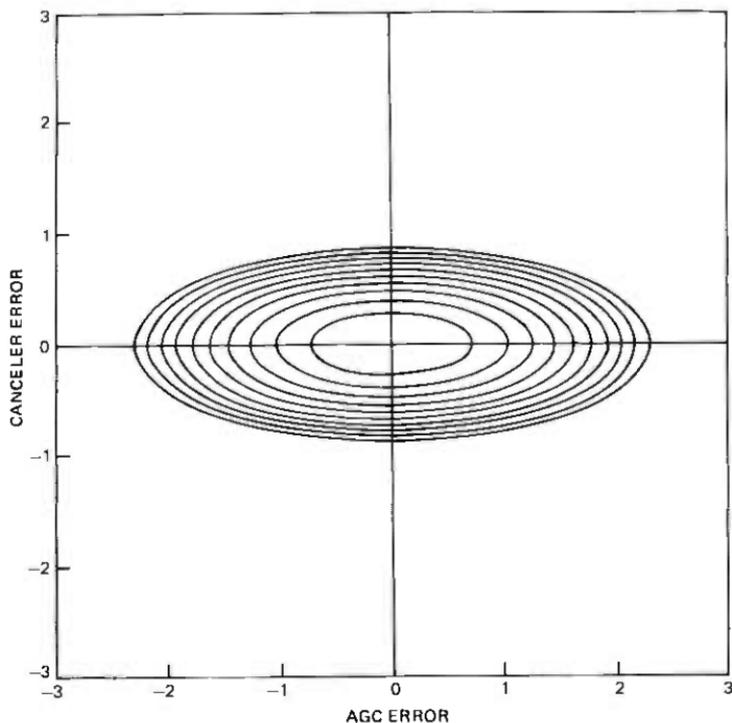


Fig. 7.—Contours of constant mean-squared error for arrangement A ($\alpha = 1, |\beta| = 1$).

squared error. The resulting joint AGC/canceler updating algorithm would then be

$$w(n+1) = w(n) - \gamma e(n)(r(n) + \mathbf{b}(n)^\dagger \mathbf{p}(n)) \quad (27a)$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma e(n)w(n)\mathbf{b}(n), \quad (27b)$$

where

$$e(n) = w(n)(r(n) + \mathbf{b}(n)^\dagger \mathbf{p}(n)) - a(n). \quad (27c)$$

V. REDUCING COUPLING EFFECTS ON THE NONCONVEX SYSTEM'S ADAPTATION

Combining (27b) and (27c) results in the following algorithm for updating $\mathbf{p}(n)$:

$$\begin{aligned} \mathbf{p}(n+1) = & \mathbf{p}(n) - \gamma w(n)^2 \mathbf{b}(n)[r(n) + \mathbf{b}(n)^\dagger \mathbf{p}(n)] \\ & + \gamma w(n)a(n)\mathbf{b}(n). \end{aligned} \quad (28)$$

It would be desirable to reduce or eliminate the coupling effect of $w(n)$ in this algorithm. Consider modifying the algorithm by replacing the

constant step size γ with $\gamma/w(n)^2$, where the second γ is a constant. Then (28) becomes

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \gamma \mathbf{b}(n)[r(n) + \mathbf{b}(n)^\dagger \mathbf{p}(n)] + \frac{\gamma}{w(n)} a(n) \mathbf{b}(n). \quad (29)$$

The average value of the correction term for fixed $\mathbf{p}(n)$ is then $-\gamma[\beta + \mathbf{p}(n)]$, which is the same as the average correction term for an adaptive echo canceler with no AGC. Thus by the simple choice of step size $\gamma/w(n)^2$, we can on the average remove the influence of AGC adaptation on echo canceler adaptation. The only remaining coupling stems from the term $[\gamma/w(n)]a(n)\mathbf{b}(n)$, whose mean value is zero. The mean of the correction term in (27a) for updating the AGC gain is

$$\begin{aligned} & -\gamma \langle e(n)(r(n) + \mathbf{b}(n)^\dagger \mathbf{p}(n)) \rangle \\ & = -\gamma \alpha (w(n)\alpha - 1) - \gamma w(n) |\mathbf{p}(n) + \beta|^2 \\ & \quad - \gamma w(n) (\sigma^2 + \sum_k \beta_k^2 - |\beta|^2), \quad (30) \end{aligned}$$

in which coupling from the echo canceler adaptation is evident in the middle term $\gamma w(n) |\mathbf{p}(n) + \beta|^2$. There is no simple way to eliminate this coupling, apart from observing that $\mathbf{p}(n) + \beta$ is the error in the echo canceler's tap coefficients, which eventually dies away.

In summary, we propose the following algorithm for jointly updating the AGC gain $w(n)$ and the echo canceler tap coefficient vector $\mathbf{p}(n)$:

$$w(n+1) = w(n) - \gamma_1 e(n)(r(n) + \mathbf{b}(n)^\dagger \mathbf{p}(n)) \quad (31a)$$

$$\mathbf{p}(n+1) = \mathbf{p}(n) - \frac{\gamma_2}{w(n)} e(n) \mathbf{b}(n), \quad (31b)$$

where γ_1 and γ_2 may be different constants. Algorithm (31b) is, on the average, uncoupled from (31a) and equivalent to the echo canceler algorithm operating in solitude with step size γ_2 . The latter algorithm has been examined by Mueller,¹ who determined an optimum step size γ equal to the reciprocal of the number of taps, and demonstrated favorable convergence characteristics, independent of echo path characteristics. The choice of γ_1 would best be made by experiment. While the above decoupling modification was only heuristically motivated, the simulations reported in Section VI confirm its usefulness.

VI. ARRANGEMENT C

Arrangement C, shown in Figs. 3a and 8a, simply omits the AGC for purposes of making a decision on the binary symbol $a(n)$; the quantizer input, which is the algebraic sum of the channel output sample and the echo canceler output, is hardlimited to ± 1 . Note that, if the $\{a(n)\}$ are binary symmetric ($a(n) = \pm 1$) data symbols, then the attenuation

of the end-to-end channel is irrelevant and no explicit AGC is necessary for making a decision. This comment also applies to other baseband data symbol formats such as diphase, and also to phase-modulated signals, in which case the data symbols $a(n)$ are numbers lying on the unit circle in the complex plane.

To enable adaptive adjustment of the echo canceler in arrangement C, the receiver's decision $\hat{a}(n)$ is scaled by an adjustable coefficient w before being subtracted from the unquantized output to form the error which is used to update the tap coefficients. The coefficient w thus has the role of an automatic reference control (ARC), rather than an AGC, since it adjusts the receiver's reference signal to a level commensurate with the attenuation of the end-to-end channel. It is adjusted jointly with the echo canceler tap coefficients. Moreover, it multiplies a discrete-valued data symbol, not a continuous-valued channel output, and so digital implementation of the receiver is simplified.

This receiver arrangement can also be applied to multi-amplitude data formats as shown in Fig. 8b. In the multi-amplitude case, the quantizer compares the analog signal with reference levels which require proper scaling in relation to its amplitude. The quantity w provides this information and can thus directly serve as a reference input to the quantizer as shown in Fig. 8b. A less attractive alternative

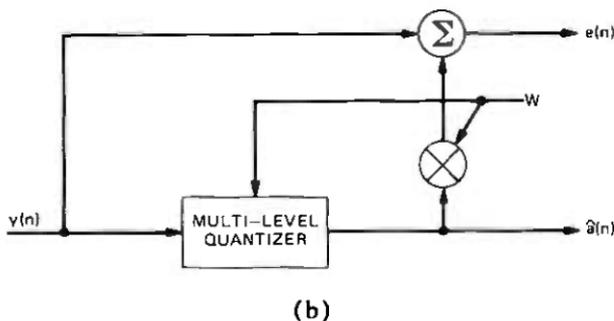
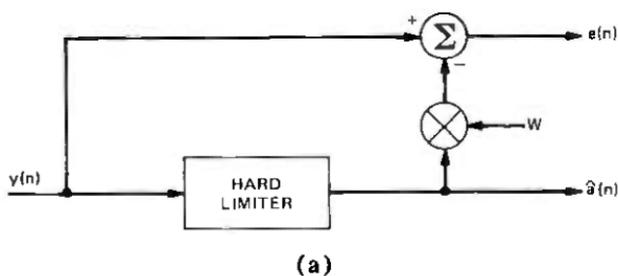


Fig. 8—Decision-making and automatic reference control in arrangement C. (a) Case of binary data symbols. (b) Case of multilevel data symbols.

would be to multiply the input signal by $1/w$ to bring it to a constant level (as with the AGC schemes) and then apply it to a quantizer with a fixed reference.

Given a set of N echo canceler tap coefficients $\{p_k\}^N$, the receiver's output sample $y(n)$, which will subsequently be quantized to form the output data symbol $\hat{a}(n)$, is

$$y(n) = r(n) + \sum_{k=1}^N p_k b(n-k). \quad (32)$$

Ideally, the coefficients $\{p_k\}_{k=1}^N$ should approximate the negatives of the echo channel's impulse response samples $\{\beta_k\}_{k=1}^N$, so that the echo is canceled and $y(n)$ consists of $\alpha a(n)$ plus noise.

The desired value of $y(n)$ is $wa(n)$, where w is an estimate of the end-to-end channel gain α . The error between the actual and desired receiver outputs, measured at the n th symbol intervals, is then

$$e(n) \equiv r(n) + \sum_{k=1}^N p_k b(n-k) - wa(n). \quad (33)$$

The ARC parameter w implicitly assumes the role of an AGC, multiplying the desired receiver output data symbol rather than the receiver input.* The parameters $\{p_k\}$ and w are to be jointly adjusted with the aim of minimizing the mean-squared value of $e(n)$.

For notational compactness, define the N -dimensional vectors

$$\mathbf{p} \equiv [p_1, p_2, \dots, p_N]^{\dagger} \quad (34a)$$

and

$$\mathbf{b}(n) \equiv [b(n-1), b(n-2), \dots, b(n-N)]^{\dagger}, \quad (34b)$$

and the $(N+1)$ -dimensional partitioned vectors

$$\mathbf{c} \equiv [\mathbf{p} : -w]^{\dagger} \quad (35a)$$

and

$$\mathbf{u}(n) \equiv [\mathbf{b}(n) : a(n)]^{\dagger}. \quad (35b)$$

Then the unquantized receiver output is written

$$y(n) = r(n) + \mathbf{p}^{\dagger} \mathbf{b}(n), \quad (36)$$

and the error is

$$e(n) = r(n) - \mathbf{c}^{\dagger} \mathbf{u}(n). \quad (37)$$

Squaring both sides of (37) and taking the expectation, we find that the mean-squared error is

* In arrangements A and B, the ideal value of w equals $1/\alpha$; in arrangement C, this value equals α .

$$\langle e(n)^2 \rangle = \sigma^2 + |\epsilon|^2, \quad (38)$$

where $\epsilon \equiv \mathbf{c} - \mathbf{s}$ and

$$\mathbf{s} \equiv \langle r(n)\mathbf{u}(n) \rangle = \begin{pmatrix} \beta \\ \dots \\ \alpha \end{pmatrix}. \quad (39)$$

The minimum mean-squared error is σ^2 , and the excess mean-squared error at time nT is defined as $|\epsilon(n)|^2$, where

$$\epsilon(n) \equiv \mathbf{c}(n) - \mathbf{s} \quad (40)$$

is the difference between the tap coefficient vector $\mathbf{c}(n)$ at time nT and its optimum value \mathbf{s} .

A simple gradient algorithm for updating $\mathbf{c}(n)$ is

$$\mathbf{c}(n+1) = \mathbf{c}(n) + \gamma e(n)\mathbf{u}(n), \quad (41)$$

where γ is a constant. (Note from expression (37) that $e(n)\mathbf{u}(n)$ is proportional to the gradient of $e(n)^2$ with respect to \mathbf{c} .) We shall examine the convergence of the excess mean-squared error ($|\epsilon(n)|^2$), as in earlier studies, assuming that successive input vectors $\mathbf{u}(n)$ are statistically independent. Subtracting \mathbf{s} from both sides of (41), we have

$$\epsilon(n+1) = \epsilon(n) + \gamma e(n)\mathbf{u}(n). \quad (42)$$

Since the $N+1$ components of $\mathbf{u}(n)$ are ± 1 ,

$$|\mathbf{u}(n)|^2 = N+1. \quad (43)$$

Also the average of the inner product

$$\begin{aligned} \langle \epsilon(n)^\dagger \mathbf{u}(n) e(n) \rangle &= \langle \epsilon(n)^\dagger \mathbf{u}(n) (r(n) - \mathbf{u}(n)^\dagger \mathbf{c}(n)) \rangle \\ &= \langle \epsilon(n)^\dagger (\mathbf{s} - \mathbf{c}(n)) \rangle \\ &= -\langle |\epsilon(n)|^2 \rangle, \end{aligned} \quad (44)$$

where we have used the assumption that $\mathbf{u}(n)$ is independent of $\mathbf{u}(n-1)$, and therefore of $\mathbf{c}(n)$. Using (38), (43), and (44), after squaring and averaging both sides of (42) we get an equation describing the evolution of the excess mean-squared error

$$\langle |\epsilon(n+1)|^2 \rangle = \langle |\epsilon(n)|^2 \rangle [1 - 2\gamma + \gamma^2(N+1)] + \gamma^2(N+1)\sigma^2. \quad (45)$$

This expression is readily iterated to yield

$$\begin{aligned} \langle |\epsilon(n)|^2 \rangle &= [1 - 2\gamma + \gamma^2(N+1)]^n \langle |\epsilon(0)|^2 \rangle \\ &\quad + \frac{\gamma(N+1)\sigma^2 [1 - (1 - 2\gamma + \gamma^2(N+1))^n]}{2 - \gamma(N+1)}. \end{aligned} \quad (46)$$

To guarantee convergence, the constant γ must be such that

$$0 < \gamma < \frac{2}{N + 1}.$$

Then the first term in (46) is a transient, which eventually decays to zero and the steady-state value $\lim_{n \rightarrow \infty} \langle |\epsilon(n)|^2 \rangle$ is

$$\frac{\gamma(N + 1)\sigma^2}{2 - \gamma(N + 1)}. \quad (47)$$

The steady-state total mean-squared error is this plus σ^2 , or

$$\lim_{n \rightarrow \infty} \langle e(n)^2 \rangle = \frac{2\sigma^2}{2 - \gamma(N + 1)}. \quad (48)$$

Expressions (46) and (47), describing the evolution of the excess mean-squared error of arrangement C, are of the same form as those derived in Ref. 1 for the echo canceler alone, with no AGC. The only difference is that $(N + 1)$ replaces N (as a result of the extra AGC coefficient w). Note that the convergence rate of $\langle e(n)^2 \rangle$ is independent of all channel parameters, in contrast to the dependence of the con-

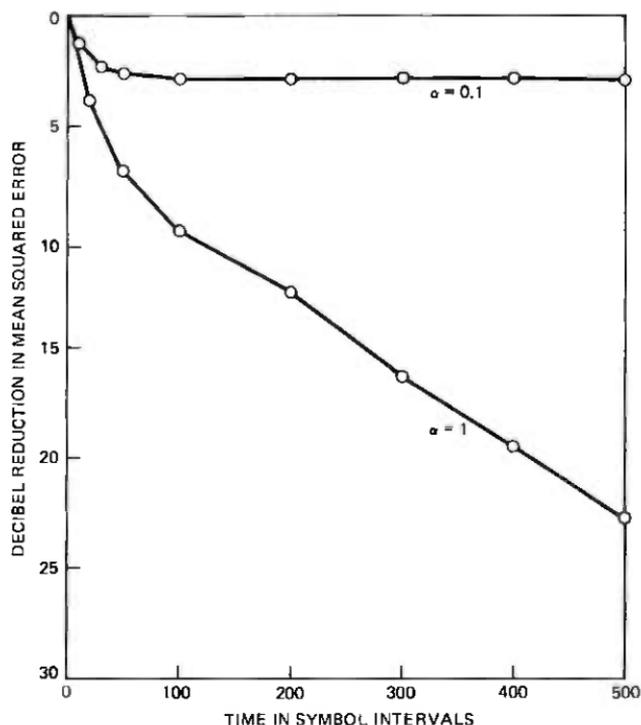


Fig. 9—Reduction of mean-squared error. Arrangement A: $\omega_0 = 1$, $\gamma = 0.01$, 16 taps.

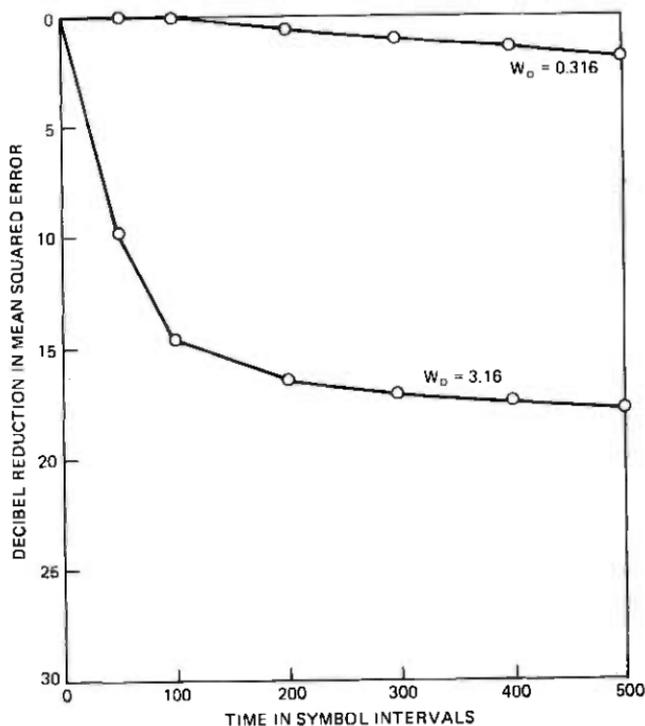


Fig. 10—Reduction of mean-squared error. Arrangement A: $\alpha = 0.316$, $\gamma = 0.01$, 16 taps.

vergence rates of arrangements A and B on the relative echo and far-end signal powers.

The choice of the adaptation coefficient γ reflects a compromise between fast adaptation and small steady-state mean-squared error. A suitable choice, proposed in Refs. 1 and 11, is

$$\gamma = \frac{1}{N + 1},$$

which yields a steady-state mean-squared error of twice the minimum mean-squared error σ^2 . With this choice of γ substituted in (46), we find that the excess mean-squared error decreases toward its minimum value (47) at a rate of $4.34/(N + 1)$ dB per adjustment.

VII. SIMULATION

Computer simulations afforded a comparison of the actual convergence behavior of the three arrangements. Results of these simulations are shown in Figs. 9 through 15. In each case, the same echo channel is used in combination with a 16-tap canceler and $\gamma = 0.01$. The echo power is normalized to unity, and the sampled echo response is

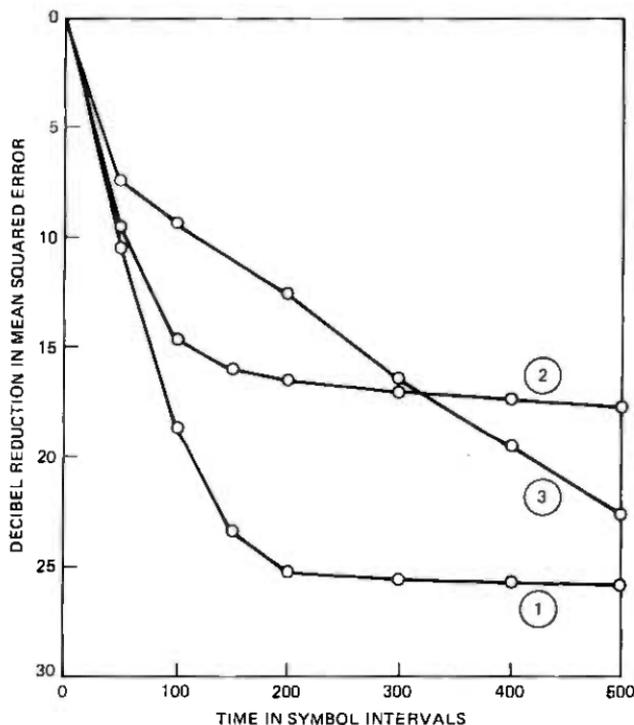


Fig. 11—Reduction of mean-squared error. Arrangement A:

- | | |
|-------------------------------------|-------------------------------------------------------------------|
| ① $\alpha = 0.1, \omega_0 = 10$ | } Initial value of AGC is $1/\alpha$; $\gamma = 0.01$, 16 taps. |
| ② $\alpha = 0.316, \omega_0 = 3.16$ | |
| ③ $\alpha = 1, \omega_0 = 1$ | |

truncated after 16 samples, so that perfect cancellation could be obtained via a set of proper coefficients. For the nonconvex arrangement B, only the decoupled updating algorithm (31) has been used since some initial runs without this modification showed convergence problems for a variety of parameter choices.

In both arrangements A and B, a uniform, nice exponential convergence seems to be the exception rather than the rule. For weak received signals, the mean-squared error often initially reduces rapidly up to a certain point, after which convergence can become very slow. This problem appears to exist even in the case where the AGC is preset to the reciprocal of the received signal. One explanation for this behavior is that, in those cases where the initial mean-squared error is dominantly caused by a misadjustment in only one loop, the overall behavior at first approximates that of a system where either only w or p is the only parameter to be adjusted. This provides a relatively fast reduction of the initial error to the point where a joint improvement

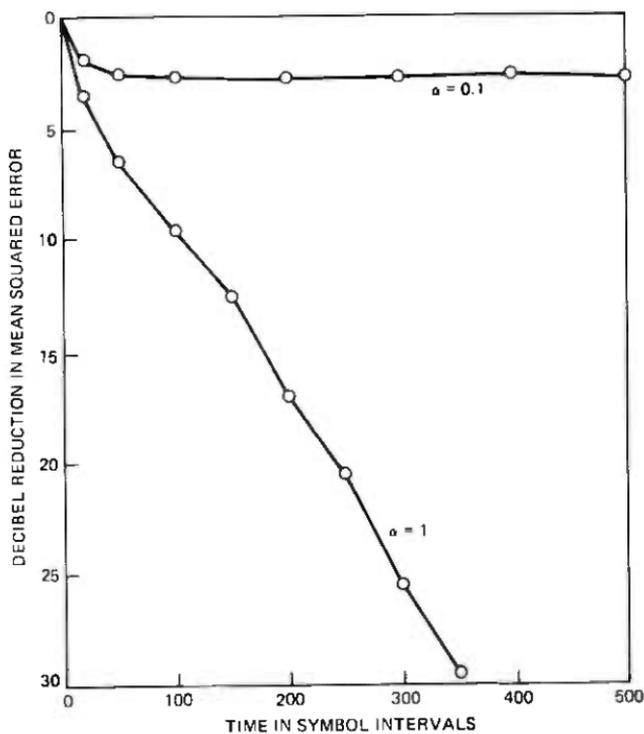


Fig. 12—Reduction of mean-squared error. Arrangement B: $\omega_0 = 1$, $\alpha = 0.01$, 16 taps.

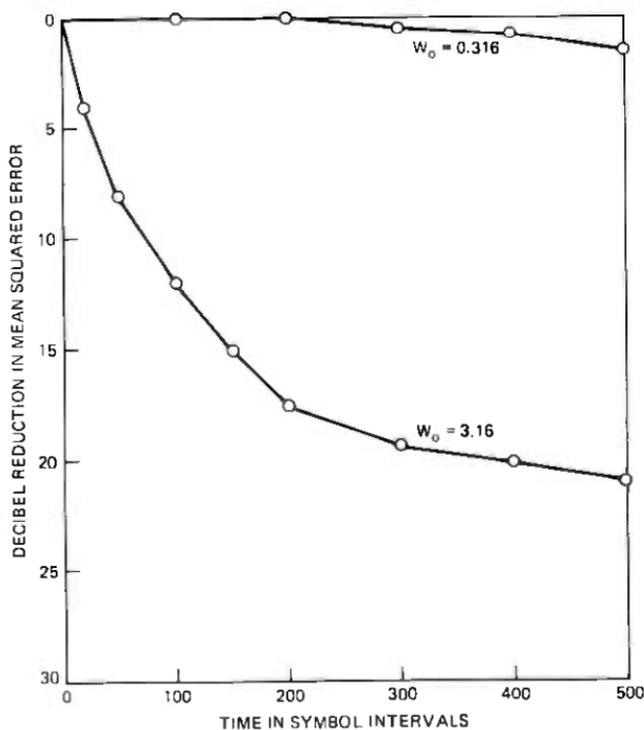


Fig. 13—Reduction of mean-squared error. Arrangement B: $\alpha = 0.316$, $\gamma = 0.01$, 16 taps.

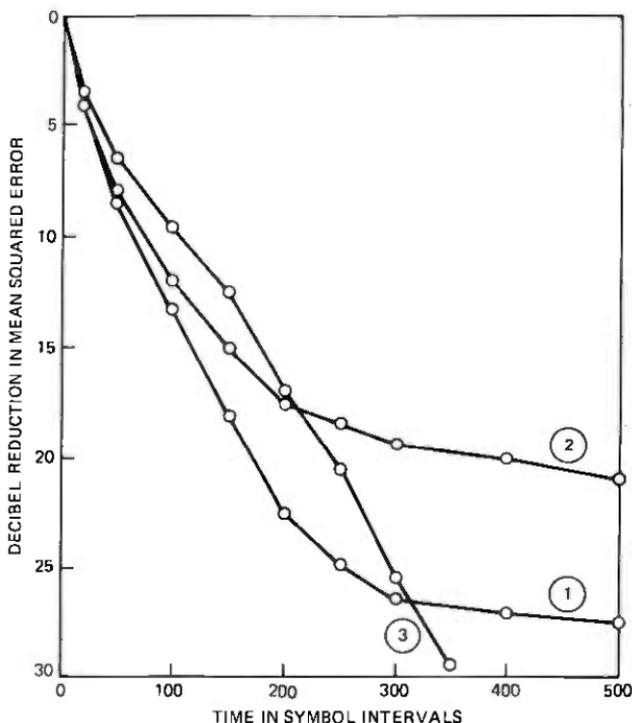


Fig. 14—Reduction of mean-squared error. Arrangement B:

- | | |
|-------------------------------------|-------------------------------------------------------------------|
| ① $\alpha = 0.1, \omega_0 = 10$ | } Initial value of AGC is $1/\alpha$; $\gamma = 0.01$, 16 taps. |
| ② $\alpha = 0.316, \omega_0 = 3.16$ | |
| ③ $\alpha = 1, \omega_0 = 1$ | |

of w and p is required. Once this situation is realized, a much slower convergence rate is expected (as we have pointed out in earlier sections), in particular with weak received signals. In the latter case, the value of w must become large and any inaccuracies in p are magnified.

Although we do not fully understand the dynamics of these systems, we feel it is worthwhile to present these results to point out the inherent problems to communication system designers. Further work would be required to provide a more thorough insight into these systems, but it is our feeling that this may be of more academic interest since our analysis has already shown that arrangement C provides as attractive a solution to the problem as one could possibly wish. This is demonstrated in Fig. 15. Only a single curve is presented, but actual simulations have shown that channel attenuation and initial ARC value can be varied by orders of magnitude and all such resulting curves would be essentially identical to the one shown in Fig. 15. The fact that simulations tend to deliver more rapid convergence than theoret-

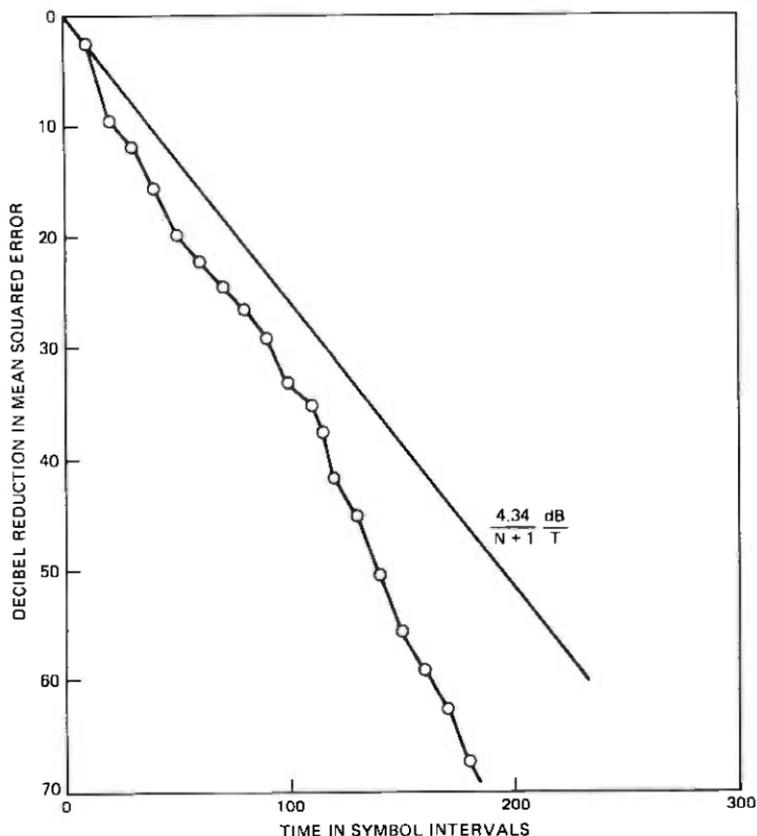


Fig. 15—Reduction of mean-squared error. Arrangement C: 16 taps, $\gamma = 1/(N + 1)$.

ically predicted is, of course, due to the programmed “ideal randomness” of the data sequence used in these simulations, whereas the analytic result is an average which includes many sequences that will not provide convergence at all (e.g., constant zeros, ones, or a dotting pattern).

VIII. DISCUSSION AND SUMMARY

Three arrangements of joint adaptive echo cancellation and gain control for full duplex data transmission have been examined. Each corresponds to a different AGC location. The third, arrangement C, has proven superior in two respects:

(i) Its convergence rate depends only on the adaptation coefficient and on the number of adjustable tap coefficients. On the other hand, arrangement A suffers slower convergence as the ratio of echo power to distant signal power increases. Arrangement B’s echo canceler can, by proper choice of adaptation coefficient, on the average be made to

converge at a rate independent of channel parameters, but the convergence of its gain control still depends on the gain of the forward channel.

(ii) Arrangement C offers simpler digital implementation; the gain w multiplies a data symbol, instead of a finely quantized channel output or receiver output. Multiplication in arrangement C becomes addition or subtraction in the case of binary symbols.

Although the foregoing analyses presuppose binary data symbols, each receiver arrangement accommodates multilevel symbols.

Severe intersymbol interference may necessitate forward equalization. If so, arrangement A with the single coefficient $w(n)$ replaced by an adaptive transversal filter is necessary. This case is discussed in Ref. 5. The favorable convergence properties and hardware simplicity of arrangement C may be retained if adaptive decision feedback equalization with a fixed forward equalizer is used.^{6,7}

REFERENCES

1. K. H. Mueller, "A New Digital Echo Canceller for Two-Wire Full-Duplex Data Transmission," *IEEE Trans. Commun.*, COM-24, No. 9 (September 1976), pp. 956-962.
2. V. G. Koll and S. B. Weinstein, "Simultaneous Two-Way Data Transmission over a Two-Wire Circuit," *IEEE Trans. Commun.*, COM-21, No. 2 (February 1973), pp. 143-147.
3. S. B. Weinstein, "A Passband Data-Driven Echo Canceller for Full-Duplex Transmission on Two-Wire Circuits," *IEEE Trans. Commun.*, COM-25, No. 7 (July 1977), pp. 654-666.
4. H. C. Van den Elzen, P. J. Van Gerwen, and W. A. M. Sniijders, "Echo Cancellation in Two-Wire Full Duplex Data Transmission with Bipolar Encoding," Proceedings of National Telecommunications Conference, Dallas, December 1976.
5. D. D. Falconer, K. H. Mueller, and S. B. Weinstein, "Echo Cancellation Techniques for Full-Duplex Data Transmission on Two-Wire Lines," Proceedings of National Telecommunications Conference, Dallas, December 1976.
6. K. H. Mueller, "Combining Echo Cancellation and Decision Feedback Equalization," *B.S.T.J.*, 58, No. 2 (February 1979), pp. 491-500.
7. P. T. Nielsen and M. W. Gram, "A Digital Hybrid for Two-Wire Digital Subscriber Loops," Proceedings of National Telecommunications Conference, Birmingham, December 1978.
8. B. Widrow and M. Hoff, Jr., "Adaptive Switching Circuits," in *IRE WESCON Conv. Rec.*, Pt. 4, 1960, pp. 96-104.
9. A. Gersho, "Adaptive Equalization of Highly Dispersive Channels for Data Transmission," *B.S.T.J.* 48, No. 1 (January 1969), pp. 55-70.
10. R. W. Chang, "A New Equalizer for Fast Start-Up Digital Communication," *B.S.T.J.*, 50, No. 6 (July-August 1971), pp. 1969-2014.
11. R. D. Gitlin, J. Mazo, and M. G. Taylor, "On the Design of Gradient Algorithms for Digitally-Implemented Adaptive Filters," *IEEE Trans. on Circuit Theory*, CT-20, No. 2 (March 1973), pp. 125-136.
12. G. Ungerboeck, "Theory on the Speed of Convergence in Adaptive Equalizers for Digital Communications," *IBM J. Res. Devel.*, November 1972, pp. 546-555.
13. K. H. Mueller and D. A. Spaulding, "Cyclic Equalization—A New Rapidly Converging Equalization Technique for Synchronous Data Communication," *B.S.T.J.*, 54, No. 2 (February 1975), pp. 369-406.
14. J. G. Proakis, "Advances in Equalization for Intersymbol Interference," in *Advances in Communication Systems*, A. J. Viterbi, ed., New York: Academic Press, 1975.
15. K. H. Mueller, "A New, Fast-Converging Mean-Square Algorithm for Adaptive Equalizers with Partial-Response Signaling," *B.S.T.J.*, 54, No. 1 (January 1975), pp. 143-153.
16. J. E. Mazo, "On the Independence Theory of Equalizer Convergence," *B.S.T.J.*, 58, No. 5 (May-June 1979), pp. 963-993.

Laser Transmitters for 70-MHz Entrance Links

by F. S. CHEN, M. A. KARR, and P. W. SHUMATE

(Manuscript received March 8, 1979)

Lightwave transmitters providing an amplitude-modulated light output from injection lasers have been designed and tested. Six transmitters were assembled using GaAlAs double-heterostructure lasers and displaying (i) third-order intermodulation products >26 dB below the fundamental at 80-percent modulation index and 0.6 to 0.8 mW average optical output; (ii) excess noise ratios <1.5 dB; and (iii) frequency response within ± 0.6 dB from 0.1 to 100 MHz. These transmitters were burned-in at room temperature for 300 h and used with entrance links in a satellite experiment.

I. INTRODUCTION

An experiment to evaluate lightwave technology as a means for providing inexpensive, wideband, reliable entrance links between remote satellite earth stations and video/telephone operating centers is being performed at Bell Laboratories locations at Holmdel, N.J. and Naperville, Ill.¹ In this experiment, either light-emitting diodes (LEDs) or injection-laser diodes (ILDs) are intensity-modulated with a 70-MHz electrical signal that is frequency-modulated by baseband signals (IM/FM). Analog receivers using either p-i-n photodiodes or avalanche photodiodes (APDs) convert the modulated light back into electrical format. The frequency-modulated 70-MHz carrier occupies the band 70 ± 20 MHz. There is also a possibility of sending several narrowband FM carriers occupying the same band. The baseband signal would be either a single 6-MHz diplexed TV channel or 1200 multiplexed voice channels occupying a similar bandwidth. Others have also reported the application of fiber optics to satellite entrance links.^{2,3}

The design objective for the transmitter was to achieve a carrier-to-noise ratio (CNR) of at least 35 dB. If an injection laser were used as the source, a simple analysis of signal-to-noise ratio showed that, to meet this requirement, an average power of -10 dBm must be launched into the optical fiber.⁴ This assumes 50-percent modulation index, 10 dB of

fiber and connector loss, 10-dB excess noise from the laser, and a p-i-n photodiode as a detector. The receiver amplifier is of the transimpedance type with a transimpedance of $4\text{ k}\Omega$ and a noise resistance of $2\text{ k}\Omega$. Because of the large laser excess noise assumed in this analysis, the use of an APD would not improve the receiver sensitivity. A launched power level of -10 dBm can be easily met using injection lasers. On the other hand, if an LED were used as a source and an APD as a detector, the required power level would be -15 dBm . Burrus-type LEDs can easily meet this requirement. For a single-carrier signal, distortion arising from nonlinearities in the lasers or photodetectors is relatively unimportant. However, for a multi-carrier signal, distortion products will have the same effect on the FM detection process as noise, and thus the sum of noise and distortion must be kept 35 dB below the carrier level. The linearity requirements could be met using selected injection lasers and Burrus-type LEDs, although the distortion may increase with aging in the case of lasers.⁵ Nevertheless, it was decided to use lasers as sources because they could launch much larger power into the fiber and because of their long lifetimes.⁶

This paper first describes the transmitter design, including circuitry and packaging. Next, intermodulation (IM) distortion and excess-noise measurements on the transmitters are presented. Finally, observations of transmitter performance during burn-in are described.

We designed and assembled nine analog transmitters using $12\text{-}\mu\text{m}$ stripe, GaAlAs double-heterostructure injection lasers.⁷ These transmitters were burned-in for 300 hours and six of the nine were used for the field experiment. The receivers, not described in this paper, were equalized versions⁸ of a previous design,⁹ with a response of $\pm 0.1\text{ dB}$ from a few kilohertz to 100 MHz.

II. TRANSMITTER CIRCUITRY AND PACKAGING

An injection laser is a threshold device with a light-current (L-I) transfer characteristic as shown in Fig. 1. Below threshold, light output is spontaneous LED light. Above threshold, the light is a coherent, lasing output. For analog modulation, the laser clearly must be biased above threshold, for example at point B in Fig. 1, in the center of a linear lasing region. The optical output then follows amplitude variations in the device current.

Transmitter design is complicated, however, by the fact that the threshold, and hence the operating point, are strong functions of temperature and device aging. Therefore, the operating point B must be stabilized under feedback control of some sort.

A digital circuit previously designed incorporating such feedback control¹⁰ has been modified for high-frequency analog modulation. The modified circuit consists of two distinct parts: the driver and the

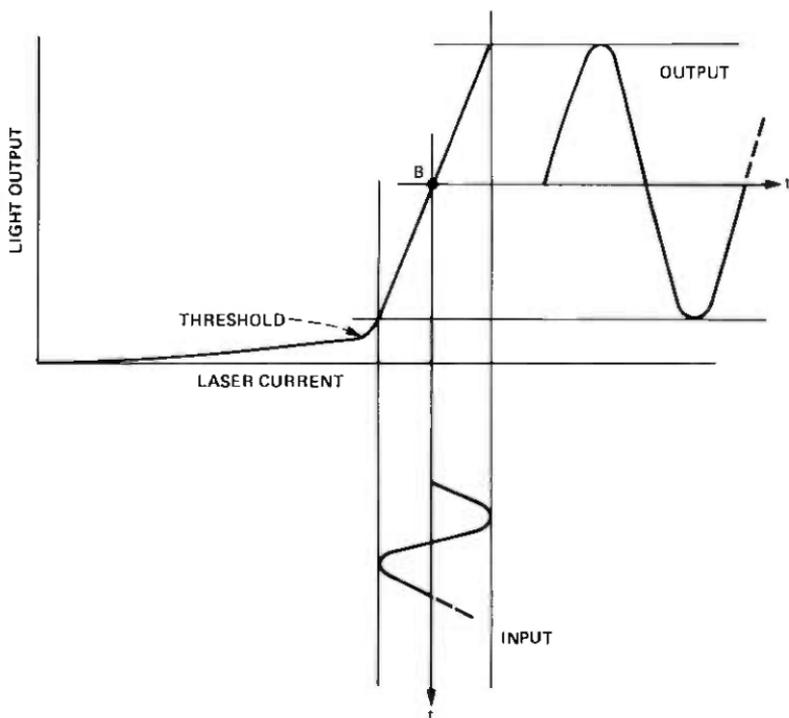


Fig. 1—Analog modulation of an injection laser.

feedback control itself (see Fig. 2). The driver is a one-transistor circuit which converts the input signal voltage to a collector current which flows through the laser. Voltage measured across the $10\ \Omega$ emitter resistor of Q_1 at a test point allows monitoring of the modulation current. The frequency response of the modulated light output using this driver was within ± 0.6 dB from 0.1 to 100 MHz. The second- and third-order intermodulation products in the drive current were at least 42 dB and 52 dB respectively, below the fundamental. Thus it is seen in the next section that the distortion of the circuit can be ignored in comparison with distortions arising in the laser.

An optical-fiber tap¹¹ samples the average optical output for feedback control. In a standard closed-loop configuration, the laser bias is controlled so as to maintain the level of this sample constant on a 3-ms time scale. Therefore, the bias does not change in response to modulation of frequencies above 300 Hz, but does limit the low-frequency response of the transmitter. (Experiments have shown, however, that the time constant can be made much longer so that frequency response is attainable down to tens of hertz.)

The laser was mounted in a hermetically sealed package.¹² The laser emission was coupled through a pigtail fiber with a hemispherical lens

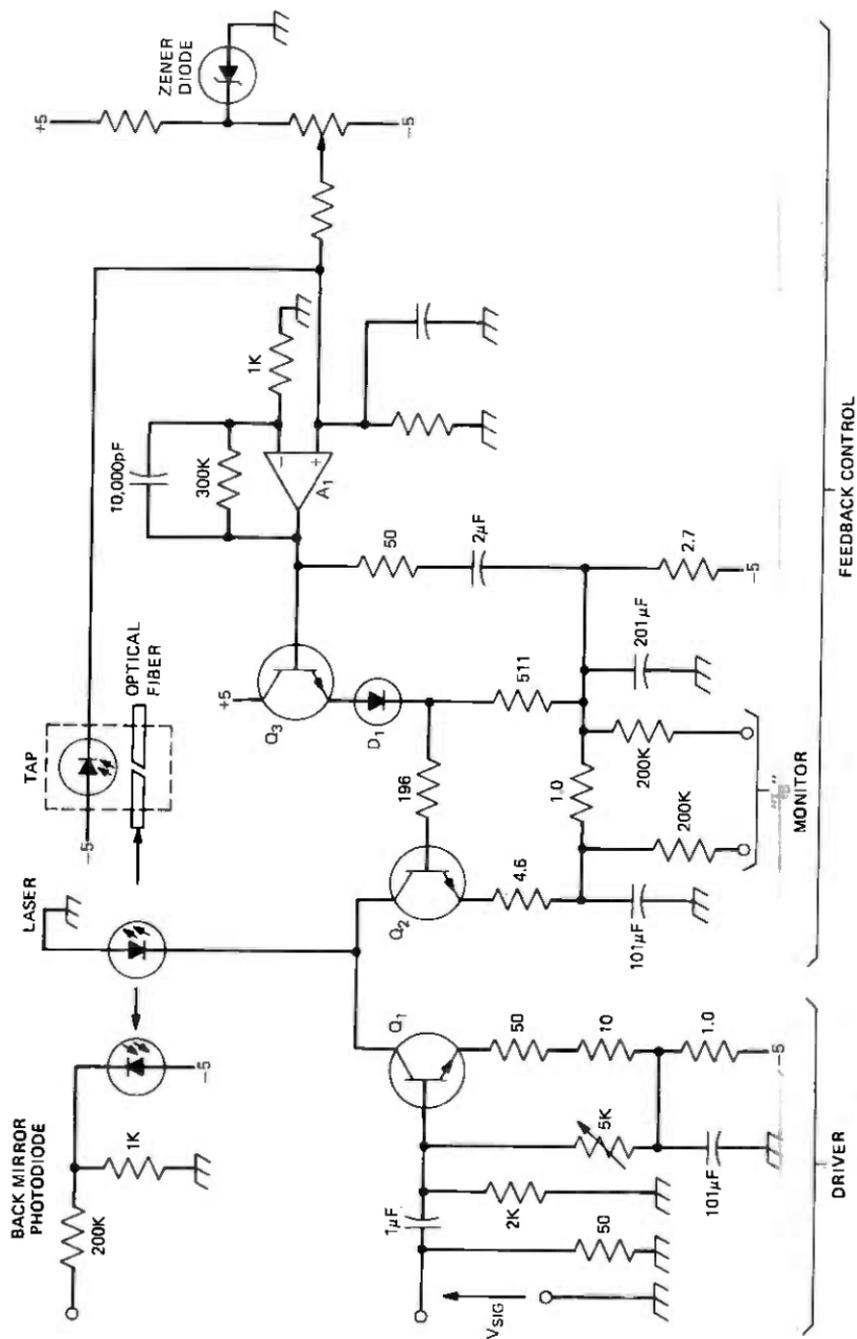


Fig. 2—Laser transmitter circuit.

at its tip to improve the coupling efficiency (about 30 to 50 percent). The optical tap in the pigtail assembly diverted about 10 percent of the power in the fiber to a p-i-n photodiode for controlling the laser operating point. There was also a p-i-n photodiode in the package for monitoring the back-mirror laser emission. The laser package and the electronic circuits were mounted on a printed-circuit board.

III. INTERMODULATION DISTORTION

Intermodulation products were measured by applying two signals of equal amplitude at frequencies $f_1 = 25$ MHz and $f_2 = 30$ MHz to the laser through a transistor driver (Q_1 in Fig. 2). The laser was dc-biased to emit 2 mW of optical power from its front mirror (P_{FM}) although, infrequently, measurements were made at higher average power levels. The modulated light was detected using a p-i-n photodiode and its output power was measured with a spectrum analyzer. The fundamental power, either f_1 or f_2 components, one of the second-order products $f_1 + f_2 = 55$ MHz, and one of the third-order products $2f_2 - f_1 = 35$ MHz were measured as a function of modulation index m . Intermodulation products from the laser could be measured to about 50 dB below either f_1 or f_2 components (56 dB below the sum of f_1 and f_2 components) using this apparatus. Results for one laser having a fairly linear $L-I$ characteristic are shown in Fig. 3, where V_{sig} is the peak input signal voltage per tone to the transistor driver. One notices that, at $P_{FM} = 2$ mW (shown in solid lines), the fundamental component (either f_1 or f_2) of the output power of the photodiode is increased by 6 dB, the second-order component by 12 dB, and the third-order component by 18 dB as V_{sig} is doubled. This is typical of a nonlinear transfer characteristic that can be expressed as a Taylor series. This relationship did not hold at $P_{FM} = 3$ mW (shown in broken lines) where a large swing of V_{sig} extended to a region of greater $L-I$ nonlinearity.

The second-order components can be excluded from the useful band by choosing a proper frequency-multiplexing scheme. However, the third-order components fall within the band and they degrade the effective CNR. From Fig. 3, the third-order component ($2f_2 - f_1$) at $m = 0.8$ was 37 dB below the f_1 component; therefore, the carrier-to-distortion ratio (CDR) is $10 \log P(2f_2 - f_1)/2P(f_1) = 42$ dB. The largest third-order products are of the form $f_1 + f_2 - f_3$ and are 6 dB larger than the type $2f_2 - f_1$.¹³ Thus the worst CDR of the distortion products of this laser would still be 37 dB, exceeding the requirement for the multicarrier FM applications (35 dB). However, the uncertainty in the aging behavior of these lasers makes them less suitable for these applications. It has been observed that nonlinearities in the $L-I$ characteristic are not always stable over extended periods of time.

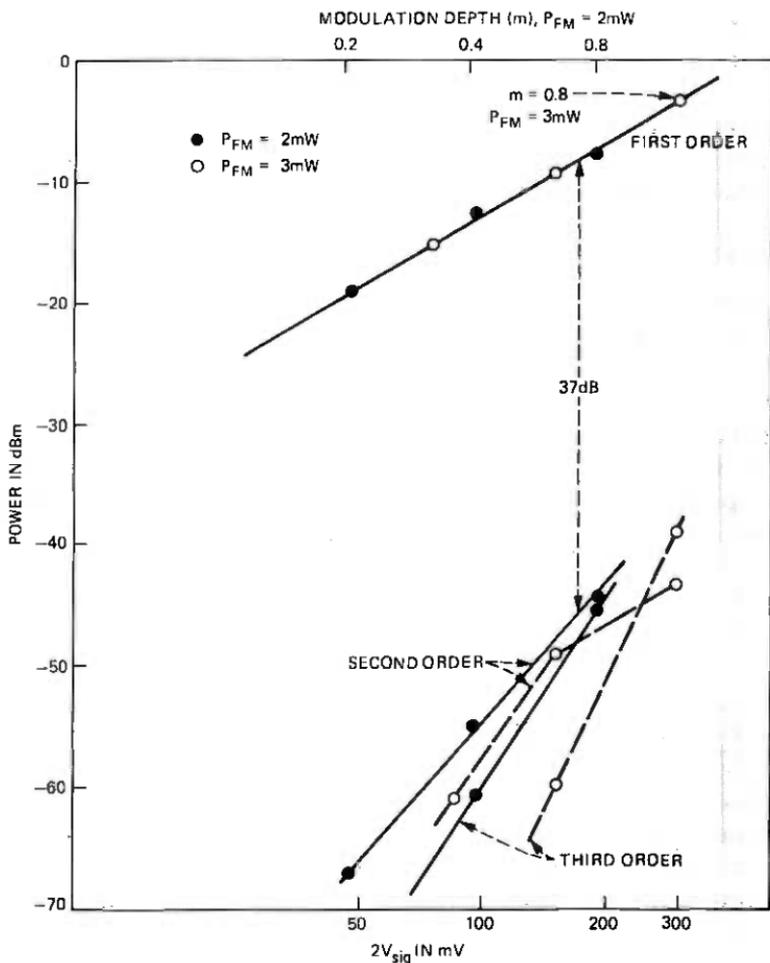


Fig. 3—Fundamental, second-, and third-order intermodulation components vs $2V_{sig}$ at $P_{FM} = 2\text{ mW}$ and 3 mW for a $12\text{-}\mu\text{m}$ stripe laser. V_{sig} is the peak input voltage per tone to the transistor driver. The slopes of the lines are 6 dB/octave, 12 dB/octave, and 18 dB/octave. Modulation depth m is also shown.

Both the severity of a nonlinearity and its position on the $L-I$ characteristic can change during periods of hundreds to thousands of hours.¹⁴

For other $12\text{-}\mu\text{m}$ stripe lasers with various degrees of nonlinearity in their $L-I$ characteristics, the third-order components varied from 17 to 46 dB (average 33 dB) below the fundamental at 2 mW average powers (P_{FM}) and $m = 0.8$.

The second-order intermodulation components of these lasers varied from 16 to 40 dB (average 29 dB) below the fundamental at $P_{FM} = 2\text{ mW}$ and $m = 0.8$.

Intermodulation products from a linear 8- μm stripe laser¹⁵ were also measured for average powers up to 5 mW. The results are shown in Fig. 4. The distortion for a given depth of modulation improved as the average laser power was increased. At 2.5 mW and $m = 0.8$, the third-order products were about the same as for linear 12- μm stripe lasers (~ 40 dB below fundamental).

Intermodulation products from a strip-buried-heterostructure (SBH) laser¹⁶ were also measured. At an average power of 4 mW and $m = 0.8$, the third-order component was down by about 48 dB. Further improvement in the sensitivity of the measuring setup is necessary to determine such small distortion more accurately. In addition to having a linear $L-I$ characteristic extending up to 100 mW/mirror under pulsed

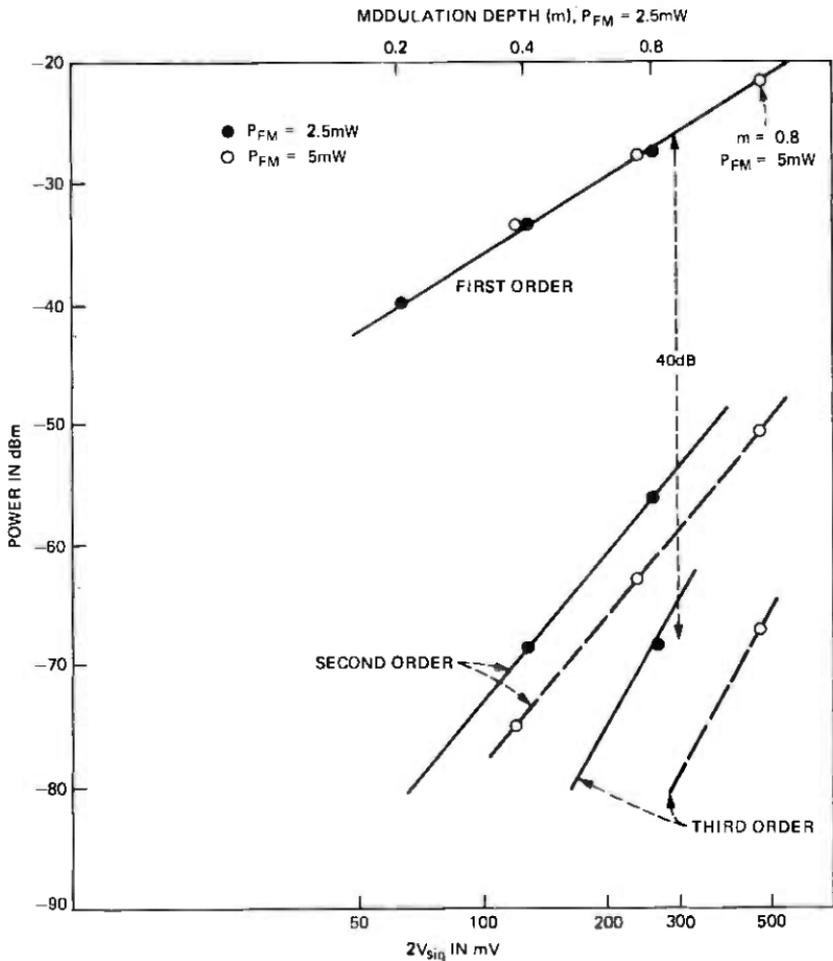


Fig. 4—Fundamental, second-, and third-order intermodulation components vs $2V_{sig}$ at $P_{FM} = 2.5$ mW and 5 mW for an 8- μm stripe laser. Modulation depth m is also shown.

conditions, the SBH laser also has the advantages of a narrowbeam divergence and a single longitudinal mode of operation. Thus it is well suited for applications requiring amplitude modulation.

IV. EXCESS NOISE

Since the laser is a regenerative amplifier saturated by spontaneous emission noise, the noise present in its output intensity rises above the shot-noise level near threshold and, ideally, it falls back to the shot-noise level above threshold. This excess-noise behavior has been observed in some junction lasers,¹⁷ but in others, especially those with nonlinear $L-I$ characteristics, excess noise persisted above threshold. Since most of our lasers showed some degree of nonlinearity, we measured the noise behavior of these lasers as part of a process before packaging.

The excess-noise ratio is defined as the ratio of intensity noise from the laser to the shot noise expected for the same average photocurrent. (This was called relative-noise ratio in Ref. 17.) Since the noise from an injection laser is known to be independent of frequency from a few megahertz up to its resonance frequency near 1 GHz, the noise was measured at 30 MHz using the apparatus shown in Fig. 5. First, the total power emitted from the front mirror of the laser (P_{FM}) was measured as the drive current was increased. The output voltage of the p-i-n photodiode packaged with the laser to monitor the emission from the back mirror (V_{BM}) was recorded at discrete levels of P_{FM} . This measurement provided sufficient data to infer P_{FM} from the back-mirror signal, even in the presence of drift in threshold. Next, the laser was driven in its LED regime ($P_{FM} = 0.2$ mW) and the front-mirror

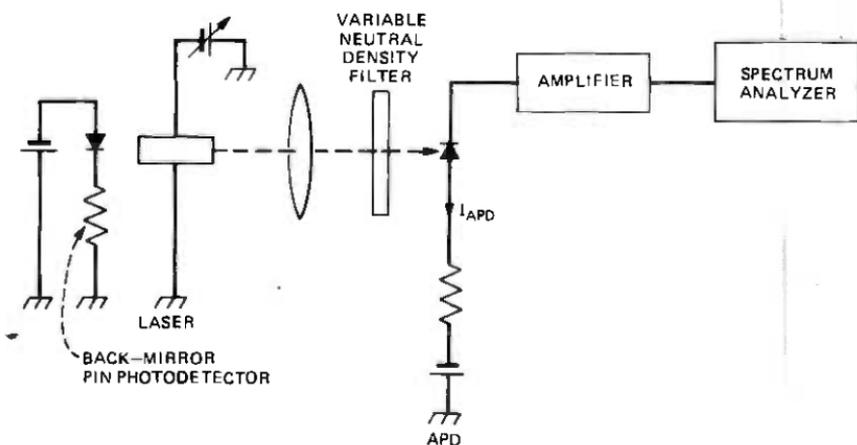


Fig. 5—Diagram of the apparatus used to measure excess noise from the junction laser.

emission was focused through a neutral-density filter onto an APD. The APD output was amplified and measured with a spectrum analyzer. The spectrum analyzer measured the sum of the avalanche-multiplied noise from the laser light and the dark current from the APD and the thermal noise from the amplifier. The noise from the latter was at least 15 dB below the other noise sources. Thus the amplifier thermal noise could be neglected in calculating the excess-noise ratios. The drive current to the laser was then increased, and, at discrete levels of P_{FM} , the neutral-density filter was adjusted to attenuate the optical signal until the average photocurrent of the APD (i_{APD} in Fig. 5) was the same as that when the laser was driven as an LED. The difference in noise power (expressed in dBm) read off the spectrum analyzer is the excess-noise ratio in decibels.

The excess-noise ratios of two lasers, one with a linear and the other with a nonlinear $L-I$ characteristic, are shown in Fig. 6. The excess noise peaked near lasing threshold. A small secondary peak appeared near the power level where the $L-I$ characteristic became nonlinear, similar to what has been observed by others.^{17, 18} Many lasers selected for this experiment showed fairly linear $L-I$ characteristics up to $P_{FM} = 4$ mW, and no secondary noise peak was observed within this power level. For the 17 lasers measured, the excess-noise ratio decreased to less than 1.5 dB when the lasers were driven at $P_{FM} = 2$ mW (the average power for all the packages). However, with a large modulation index (≈ 0.8), the laser will be driven down toward threshold for a fraction of a modulation cycle, and more excess noise will be added to the system during this period. How this effect will influence the overall system performance is not clear at this time.

V. BURN-IN RESULTS

Nine lasers were selected for packaging based on their relatively low IM distortion, small excess-noise ratios, and absence of severe kinks in their $L-I$ characteristics at both mirrors below $P_{FM} = 4$ mW.

All completed packages were mounted in a life-test rack and operated for 300 h at room temperature (18 to 26°C). The photocurrents of the p-i-n diode in the tap, the p-i-n diode in the laser package monitoring the back laser mirror, and a third p-i-n diode monitoring the actual fiber output were monitored. In addition, the drive and bias currents were recorded semiautomatically in a digital format during burn-in. The percent changes of the output power from the fiber ($\Delta P_{\text{fiber}}/P_{\text{fiber}}$), of the back-mirror p-i-n voltages ($\Delta V_{\text{bm}}/V_{\text{bm}}$) and the bias current (I_B) are shown in Table I, together with the second- and the third-order intermodulation products and excess-noise ratios measured at $P_{FM} = 2$ mW. Among the nine packages completed, six showed less than 2-percent change in the bias current (I_B) during burn-in. Part

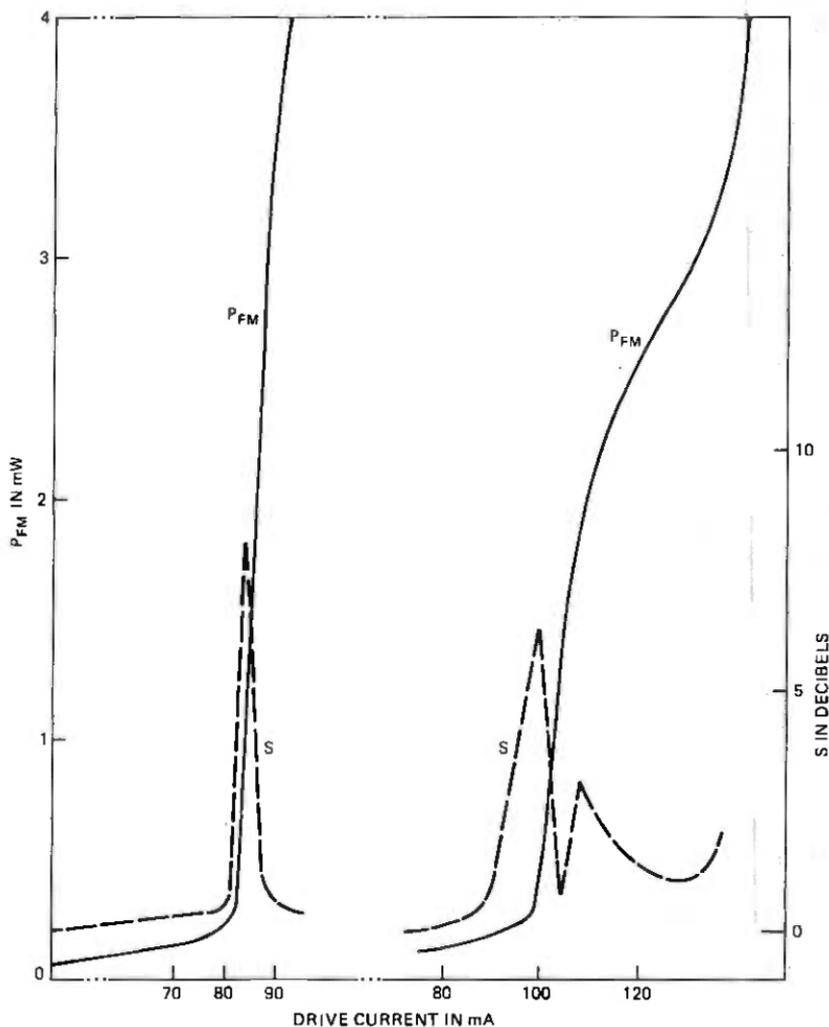


Fig. 6—Laser power (P_{FM}) and the excess-noise ratio (S) vs drive current.

of the change in I_B was due to variations in the ambient temperature and is not necessarily indicative of deterioration of the lasers. Consequently, they were used in the field experiment.

One package ($L-2$) developed large changes in fiber output, back-mirror emission, and bias current, while the tap photocurrent was kept constant by the feedback circuit. The cause of this instability was probably mechanical instability in the tap. The remaining two packages failed in less than seven hours for unknown reasons.

Table I—Results of 300 h burn-in at room temperature, and second- and third-order intermodulation products and excess noise ratios at $P_{FM} = 2 \text{ mW}$

Package No.	$\frac{\Delta P_{\text{fiber}}}{P_{\text{fiber}}}$ ($\pm\%$)	$\Delta V_{\text{BM}}/V_{\text{BM}}$ ($\pm\%$)	I_B (mA)	2nd-order inter-modulation (-dB)	3rd-order inter-modulation (-dB)	Excess-noise ratio (dB)
L-1	2	5.3	82-84	38	26	1.2
L-2	7.5	8.6	80-89	32	17	1.1
L-3	1.4	2.6	97-101	31	40	0.8
L-4	0.8	1.1	76-78	34	46	1.5
L-5	1.4	10.2	86-87	29	34	1.3
L-7	1.2	1.9	110-112	20	35	1.4
L-9	1.8	2.1	83-85	16	30	0.1

VI. SUMMARY AND DISCUSSION

Lightwave transmitters capable of delivering frequency-modulated subcarrier 1F signals ($70 \text{ MHz} \pm 20 \text{ MHz}$) were developed for use in satellite entrance links. Double-heterostructure GaAlAs injection lasers with $12\text{-}\mu\text{m}$ stripes were used. At 2-mW laser power, the excess-noise ratio from the lasers was not more than about 1.5 dB. Second-order IM products were 16 to 40 dB and third-order IM products were 17 to 46 dB below the fundamental at a modulation index of 0.8. The average power coupled into the FT3 fiber ($NA = 0.23$ core dia = $55 \mu\text{m}$) varied from 0.6 to 0.8 mW.

The third-order IM products observed for many $12\text{-}\mu\text{m}$ stripe lasers were sufficiently low for possible applications in multi-carrier transmission. However, the magnitude of distortion may increase as the lasers age, since they tend to develop nonlinearities in their $L-I$ characteristics. In principle, improvements in reducing the nonlinear distortion from direct modulation of injection lasers can be approached from two directions: electronic compensation for the nonlinearity and/or modification of the laser structure. Use of a predistorted signal derived from one laser and applied to a second laser¹⁹ to minimize the distortion products has been successful with LEDs, but it may be difficult with injection lasers since two lasers cannot be expected to develop the same nonlinearity at the same time. The approach using improved laser structures has been more successful. Already various structures have been reported in the literature^{16, 20-22} showing extremely linear characteristics. With these new, linear structures, ultimately the magnitude of the IM products and the useful bandwidth for analog modulation will be determined by the inherent resonance of these lasers, which is in the vicinity of 1 GHz. Due to these resonances, distortion would be expected to appear at subcarrier frequencies of

several hundred megahertz even for a laser with perfectly linear $L-I$ characteristics²³ as measured at dc (or low frequencies).

VII. ACKNOWLEDGMENTS

We wish to thank I. Camlibel for packaging the lasers, H. W. Reinbold for building the life-test rack, G. Moy for providing the printed circuits, R. Pawelek for assistance in making the taps, and M. DiDomenico for critical reading of the manuscripts. We also thank members of B. C. DeLoach's department for providing the lasers used in this work, and W. T. Tsang and R. A. Logan for the strip-buried-heterostructure laser.

REFERENCES

1. A. Albanese and H. F. Lenzing, "Video Transmission Tests Performed on Intermediate-Frequency Lightwave Entrance Links," *SMPTE J.* 87 (December 1978), pp.821-824.
2. J. J. Pan and D. E. Halley, "Fiber-Optic Links for Microwave and Satellite-Communication Terminals," 1976 Nat. Telecom. Conf.
3. P. G. File, "Fiber-Optics Links for Satellite Terminal Interconnectors," 1978 Int. Conf. on Commun.
4. P. W. Shumate and R. G. Smith, unpublished work.
5. T. L. Paoli, "Nonlinearities in the Emission Characteristics of Stripe-Geometry (AlGa) As Double-Heterostructure Junction Lasers," *IEEE J. Quant. Elec., QE-12* (December 1976), pp. 770-776.
6. R. L. Hartman, N. E. Schumaker, and R. W. Dixon, "Continuously Operated (Al,Ga) As Double-Heterostructure Lasers with 70°C Lifetimes as Long as Two Years," *Appl. Phys. Lett.*, 31 (December 1977), pp. 756-759.
7. R. L. Hartman and R. W. Dixon, "Reliability of DH GaAs Lasers at Elevated Temperatures," *Appl. Phys. Lett.*, 26 (March 1975), pp. 239-242.
8. P. W. Dorman, private communication.
9. R. G. Smith, C. A. Brackett, and H. W. Reinbold, "Optical Detector Package," *B.S.T.J.*, 57, No. 6, Part I (July-August 1978), pp. 1809-1822.
10. P. W. Shumate, F. S. Chen, and P. W. Dorman, "GaAlAs Laser Transmitter for Lightwave Transmission Systems," *B.S.T.J.*, 57, No. 6, Part I (July-August 1978), pp. 1823-1836.
11. M. A. Karr, T. C. Rich, and M. DiDomenico, Jr., "A Lightwave Fiber Tap," *Appl. Opt.*, 17 (August 1978), pp. 2215-2218.
12. I. Camlibel, T. C. Rich, and D. R. Mackenzie, unpublished work.
13. K. Miya, "Satellite Communications Engineering," Japan: Lattice Co., Ltd., 1975, pp. 131-142.
14. T. L. Paoli, "Changes in the Optical Properties of CW (AlGa)As Junction Lasers During Accelerated Aging," *IEEE J. Quant. Elec., QE-13* (May 1977), pp. 351-359.
15. R. W. Dixon, F. R. Nash, R. L. Hartman, and R. T. Hepplewhite, "Improved Light-Output Linearity in Stripe-Geometry Double-Heterostructure (AlGa)As Lasers," *Appl. Phys. Lett.* 29 (September 1976), pp. 372-374.
16. W. T. Tsang, R. A. Logan, and M. Hegems, "High-Power Fundamental-Transverse-Mode Strip Buried Heterostructure Lasers with Linear Light-Current Characteristics," *Appl. Phys. Lett.* 32 (March 1978), pp. 311-314.
17. T. L. Paoli, "Noise Characteristics of Stripe-Geometry Double-Heterostructure Junction Lasers Operating Continuously—I. Intensity Noise at Room Temperature," *IEEE J. Quant. Elec. QE-11* (June 1975), pp. 276-283.
18. R. Lang, K. Minemura, and K. Kobayashi, "Low-Frequency Intensity Noise in C. W. (GaAl)As D. H. Lasers with Stripe Geometry," *Elec. Lett.*, 13 (April 1977), pp. 228-230.
19. J. Strauss, "Linearized Transmitters for Analog Fiber Optical Systems," *CLEOS Digest of Tech. Papers*, 1978, p. 18.
20. K. Aiki, M. Nakamura, T. Kuroda, and J. Umeda, "Channeled-Substrate Planar Structure (AlGa)As Injection Lasers," *Appl. Phys. Lett.*, 30 (June 1977), pp. 649-651.

21. D. R. Scifres, W. Streifer, and R. D. Burnham, "Curved Stripe GaAs: GaAlAs Diode Lasers and Waveguides," *Appl. Phys. Lett.*, **32** (February 1978), pp. 231-234.
22. H. Kumabe, T. Tanaka, H. Namizaki, M. Ishii, and W. Susaki, "High Temperature Single Mode CW Operation with a Junction up TJS Laser," *Appl. Phys. Lett.*, **33** (July 1978), pp. 38-39.
23. W. Harth and D. Siemsen, "Modulation Characteristics of Injection Lasers Including Spontaneous Emission—1. Theory," *AEU*, **30** (1976), pp. 343-348.

Adaptive Aperture Coding for Speech Waveforms—I

By N. S. JAYANT and S. W. CHRISTENSEN

(Manuscript received December 29, 1978)

In aperture coding, one refrains from encoding waveform samples until the waveform crosses an appropriately wide aperture centered around the last encoded value. If the waveform is slowly varying in some sense, the above procedure can be a basis for bit rate reduction. The identification of aperture-crossing samples can be either explicit or implicit, and it is the latter case that this paper mainly addresses. We follow a finite length, converging-aperture procedure proposed recently for picture waveforms, and show that it can be used for speech coding as well if the aperture width is designed to be syllabically adaptive. We also describe, for Nyquist-sampled speech, desirable designs for aperture shape and aperture length L . The special case of $L = 1$ corresponds to ternary delta modulation with a constant encoding rate of $\log_2 3 \sim 1.6$ bits/sample. Using longer apertures (e.g., $L = 2, 3$), we show that it is possible to obtain average encoding rates as low as 1.2 bits/sample without significantly changing output speech quality. With 8- to 12-kHz sampling, the average bit rate would then be 9.6 to 14.4 kb/s. At these transmission rates, adaptive aperture coding, used in conjunction with a simple (first-order) adaptive predictor, can provide communications quality speech.

I. INTRODUCTION

The encoding technique described in this paper is intended to be a simple time-domain approach for encoding speech waveforms at transmission rates like 9.6 or 16 kb/s. The digital speech output resulting from this technique, or simple modifications thereof,⁶ is expected to be of *communications* quality: less than *toll* quality, but nevertheless adequate for many applications.

The notion of aperture coding, per se, is not new. It has been considered extensively for digitizing telemetry data, with a view to

exploiting their slowly changing characteristics.¹⁻³ The point of this paper is that aperture coding can be useful for low-bit-rate digitizations of speech waveforms as well, provided the coding procedure is designed to be properly adaptive to the changing statistics of speech inputs. In fact, an important contribution of this paper is the specification of a rather carefully designed syllabic adaptation algorithm for aperture width.

Adaptive aperture coding is inherently a variable rate procedure, and for use with a transmission channel that expects a constant-rate output, one would need an appropriate buffer at the coder output. Typical buffer lengths and consequent encoding delays can be several tens of milliseconds. This will be of no concern when aperture coding is used for digital speech storage but, for transmission applications, the encoding delay will be an important consideration.

II. APERTURE CODING

The basic notion can be explained with reference to Fig. 1. Assume that the waveform sample at time 0 has been encoded and transmitted. The idea now is to view the immediate future of $X(t)$ through an aperture of width $2A$, centered on the circle that represents the transmitted value at time 0; and to refrain from transmitting samples that lie within this aperture; the next transmission will therefore occur at time 3, after which the process continues with an updated aperture. Here, and in the next figure, open circles represent transmitted values, while solid dots denote samples deemed redundant. In reconstructing the waveform, redundant samples can be assigned amplitudes equal (for example), to that of the last transmitted sample, as shown by the dashed horizontal running through the aperture. This procedure en-

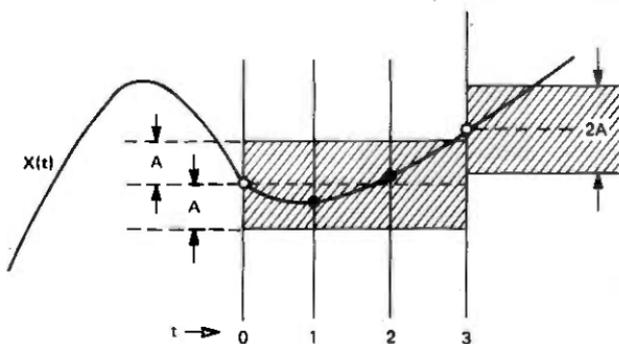


Fig. 1—Illustration of the aperture coding concept.

tails a distortion that can be referred to as aperture noise. As one increases A , aperture noise increases, but so does the proportion of samples that need not be encoded/transmitted. The tradeoff between noise and transmission probability depends on how slowly the input waveform varies, and for nonstationary inputs such as speech waveforms, the best tradeoffs are realized in schemes where one adapts A to changing input statistics. [With nonadaptive aperture schemes for Nyquist-sampled speech, a transmission probability of 1 out of 2 (or 2.5, 3, 4, 5) samples implies typical signal-to-aperture-noise ratios of about 33 (or 21, 18, 14, 11) dB, assuming that the only silences present in the speech input are naturally occurring microsilences, and not explicit pauses.]

Practical aperture schemes present two considerations which have not been introduced in Fig. 1. First, the "transmitted" samples have to be digitized somehow, so that the quality of reproduced speech will be characterized by this digitization—or quantization—noise, in addition to the aperture noise mentioned earlier. Second, the decoder at the receiving end has to know which of the input samples have been deemed redundant by the encoder, and which of them have been explicitly digitized. Most aperture coding literature¹⁻³ assumes explicit transmission of the above "timing" information. For example, the encoder can transmit, for each input sample, a binary number which tells the decoder whether that sample is being encoded or deemed redundant. If the probability of a nonredundant sample is p and if such a sample is further encoded using B bits, the average transmission rate is $[p \cdot B + 1]$ bits/sample, where the term 1 is due to the constant timing information bit; and the savings, relative to a zero-aperture scheme, are $[B(1 - p) - 1]$ bits/sample. This formula suggests that B has to be large enough (for a given p) so that the savings is positive in spite of the timing information. On the other hand, in low bit rate applications, values of p (that are compatible with a tolerable amount of aperture + quantization noise) may be such that the savings due to aperture coding are either insufficient or negative—unless, of course, the timing information overhead can be avoided altogether. An aperture scheme which does precisely this was described recently by Murakami, Tachibana, Fujishita, and Omura⁴ in the context of picture coding, and the purpose of this paper is to describe our modification of that scheme for encoding speech waveforms with $B = 1.2$ to 1.6 bits/sample, a range of bit rates which clearly cannot afford explicit transmission of timing information.

Succeeding sections describe our findings concerning aperture characteristics that are desirable for low bit rate speech coding. These characteristics include aperture shape, aperture length (to be defined presently), and adaptation algorithms for aperture width A .

Table I—Aperture characteristics for $L = 3$ and $J = 0.5$

Run Length R (First time that crossing is observed)	Updating Relative to Predicted Value
1	$\pm A_0$
2	$\pm A_0 / \sqrt{2}$
3	$> A_0 / 2$
">3" } No crossing observed	0

The use of a "zero" output level implies the use of a finite-length aperture. In fact, the aperture length can be defined as the time at which the aperture is truncated with a zero output level Z . In Fig. 2, $L = 3$, and in the particular example that has been sketched, $R = 3$ as well.

The number of "output" points on the aperture characteristic of Fig. 2 was 7 ($3P$'s + $3N$'s + $1Z$). In general, the aperture characteristics in our scheme are described by $(2L + 1)$ distinct outputs, and corresponding transmission code words.

Relative amplitudes on the characteristic are determined by the shape of the aperture. We have found that converging apertures that are appropriate for speech can be conveniently formalized into the class

$$A(t) = A_0 \cdot 2^{-J \cdot t}, \quad (1)$$

where $A(t)$ is the aperture width at time t . We have further found that a desirable range for J is $(0.5 \leq J \leq 1)$. (We have also looked at shapes described by complete convergence, $A(L) = 0$, with corresponding $(2L + 1)$ -point characteristics, but we have found them to be less useful than those described by the exponential decay above.)

Table I defines the quantization characteristics of aperture coding for the illustrative case of $L = 3$ and $J = 0.5$. Notice that the output (quantized) amplitudes are defined relative to a "predicted value." In the examples of Figs. 1 and 2 we have assumed that all predicted values are equal in amplitude to the last (explicitly) transmitted amplitude, as indicated by the dashed horizontals running through the aperture areas. The situation corresponds, formally, to a first-order predictor with a coefficient a_1 equal to unity.* In general, however, one can use a speech-specific predictor $a_1 = 0.85$,⁵ or a higher-order predictor (for example, $a_1 = 1.10$, $a_2 = -0.28$, $a_3 = -0.08$; see Ref. 5) to reconstruct redundant samples, and to predict nonredundant samples prior to updating, as in Table I. The coding procedures in these more general cases would be qualitatively described by Fig. 3. Further, the predictor can also be adaptive, to follow the changing spectral char-

* Nonpredictive aperture coding results if $a_1 = 0$.

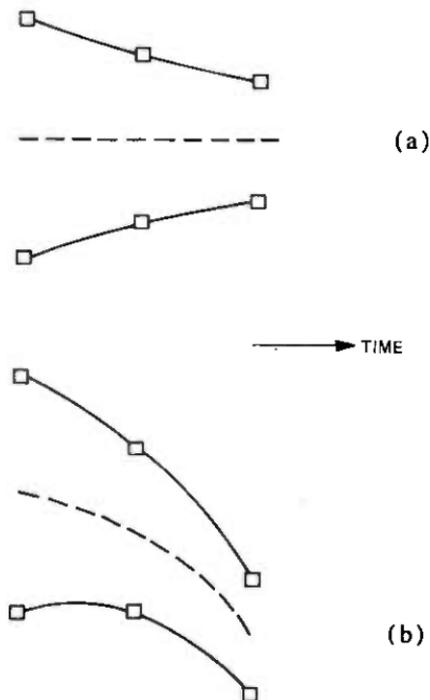


Fig. 3—Aperture coding with (a) first-order prediction: $\alpha_1 = 1$, and (b) more general prediction.

acteristics of input speech. The adaptive predictors considered in this paper are of first order, in the interest of simplicity: the adaptive predictor coefficient is simply set equal to the one-sample-lag autocorrelation value c_1 of the speech input. The parameter c_1 is updated once for every 256-sample input block. Explicit transmission of c_1 values to a receiver will typically entail an additional information transmission of about 100 to 200 bits per second. This extra transmission can be entirely avoided in schemes where c_1 is estimated from a past history of reconstructed, rather than input, speech.⁶

IV. ADAPTIVE APERTURES

Nonadaptive and adaptive apertures are sketched in Fig. 4. The figures show the time evolution, if any, of the maximum aperture width $A_0(A(0))$. For a nonstationary signal such as speech, it is critical to have an adaptive procedure such as in Fig. 4c. The adaptations would let A_0 follow changing input statistics and provide individually tailored arrangements for encoding high-level voiced segments, low-level sounds such as fricatives, and zero-level microsilenences. The results of

this paper assume that the aperture shape as described by J in (1) is fixed, and that only the width A_0 is adaptive.

We studied many adaptation algorithms, including those that can be described as instantaneous, periodic, and syllabic. The best results were obtained with syllabic adaptations as typified by the algorithm

$$A_0^{(r+1)} = G_1 \cdot A_0^{(r)} + G_2 \cdot (\text{ADAPT})^{(r)}$$

$$G_1 = 1 - \epsilon^2; \quad \epsilon \rightarrow 0$$

$$(\text{ADAPT})^{(r)} = 1 \quad \text{if} \quad \sum_{s=1}^4 R_{r-s} < K$$

$$= 0 \quad \text{otherwise.} \quad (2)$$

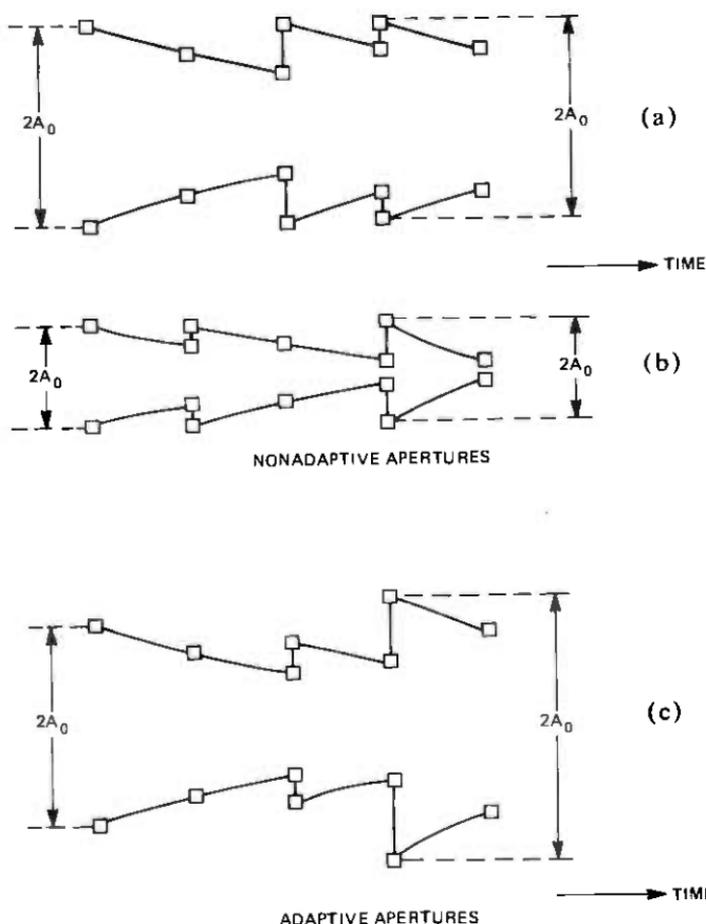


Fig. 4—Nonadaptive and adaptive apertures.

In the above algorithm, r indexes successive new apertures, and not successive input samples. In other words, going from r to $r + 1$ could mean an interval up to L input samples. The parameter R refers to the run length of redundant samples. By defining a "Z" event as a run of length $L + 1$, the parameter R is seen to have a range ($1 \leq R \leq L + 1$). Briefly, the above adaptation logic uses a succession of four small runs as a cue for increasing A_0 ; in the absence of such a cue, the logic lets A_0 decrease slowly, at a rate given by G_1 . Our experiments have shown that desirable values of G_1 for 8-kHz-sampled speech are between 0.95 and 0.99 (corresponding to syllabic time constants of 2 to 8 ms for aperture decreases); while good choices for the threshold K are 6, 6, and 8 for aperture lengths of $L = 1, 2,$ and 3 respectively. The most interesting parameter in (2) by far was the quantity G_2 that determines the nature of aperture width increases, and we shall come back to this parameter presently.

Meanwhile, Figs. 5, 6, and 7 provide illustrative descriptions of the adaptive procedure described in (2). Figure 5 shows how the aperture width A_0 (5b) tracks input speech power (5a), Fig. 6 provides a typical histogram of A_0 samples, showing a microsilence-related concentration at very low values of A_0 , and Fig. 7 compares typical aperture-crossing sequences (redundant and nonredundant samples, versus time) in nonadaptive (7b) and adaptive (7c) schemes. In the 2-state sequences of Figs. 7b and 7c, a zero state represents a redundant sample, while a nonzero state denotes an aperture crossing, or nonredundant sample.

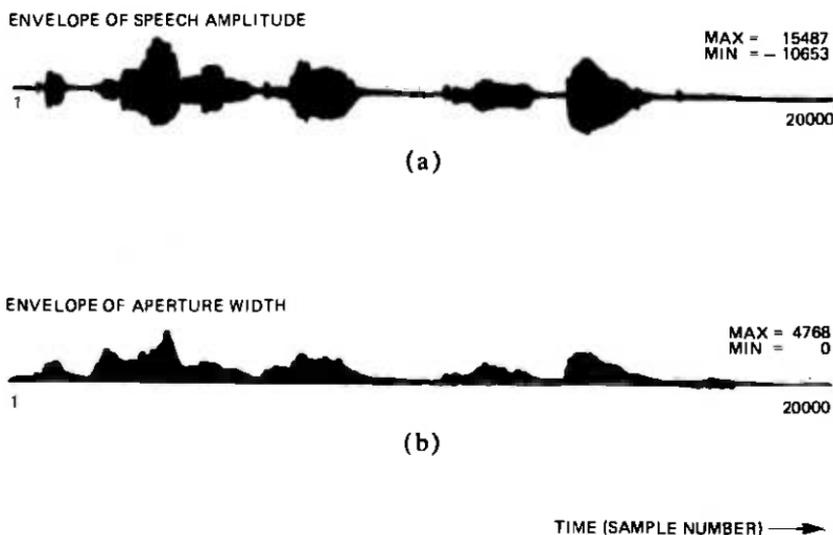


Fig. 5—Syllabically adaptive apertures: envelopes of (a) sentence-length speech (b) aperture width A_0 .

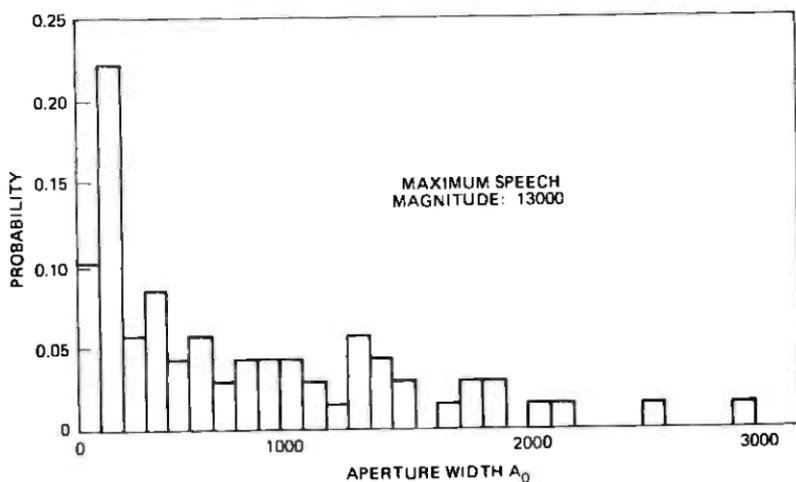


Fig. 6—Histogram of aperture width samples.

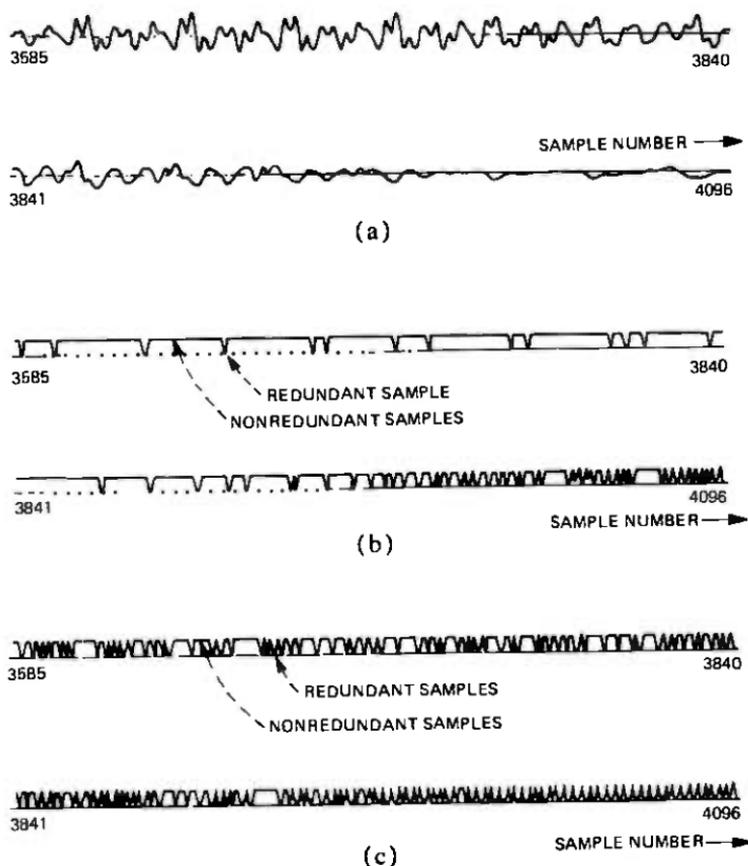


Fig. 7—Aperture crossings in (b) nonadaptive and (c) adaptive schemes, corresponding to a (a) speech waveform segment.

V. PERFORMANCE OF APERTURE CODING AS A FUNCTION OF L AND G_2

The more interesting results of our experiments (computer simulations) are summarized in Figs. 8 and 9. These results apply to a bandlimited (200 to 3200 Hz) female utterance, "THE CHAIRMAN CAST THREE VOTES," and the nonadaptive third-order predictor [$a_1 = 1.10$, $a_2 = -0.28$, $a_3 = -0.08$] mentioned earlier. The case of adaptive prediction is discussed briefly in Section VI.

5.1 Segmental signal-to-noise ratios

The objective speech quality measure used in Fig. 8 is the segmental signal-to-noise ratio SNRSEG obtained by computing the s/n ratios in 256-sample (32 ms) blocks, expressing the values in decibels, and taking the average of local decibel values over the length of the sentence-length input—a procedure which reflects low-level speech rendition better than the conventional average s/n ratio. It is significant that the maximum performance with $L = 3$ is nearly 1 dB below the peak performance with $L = 1$ and $L = 2$; and that for a given value of G_2 , $L = 2$ tends to perform better than $L = 1$ (except if $G_2 < 25$). It will be seen, on the other hand, that, transmission-rate-wise, interesting values of G_2 are quite different for different values of L , and we presently reexamine the results of Fig. 8 taking the above fact into account.

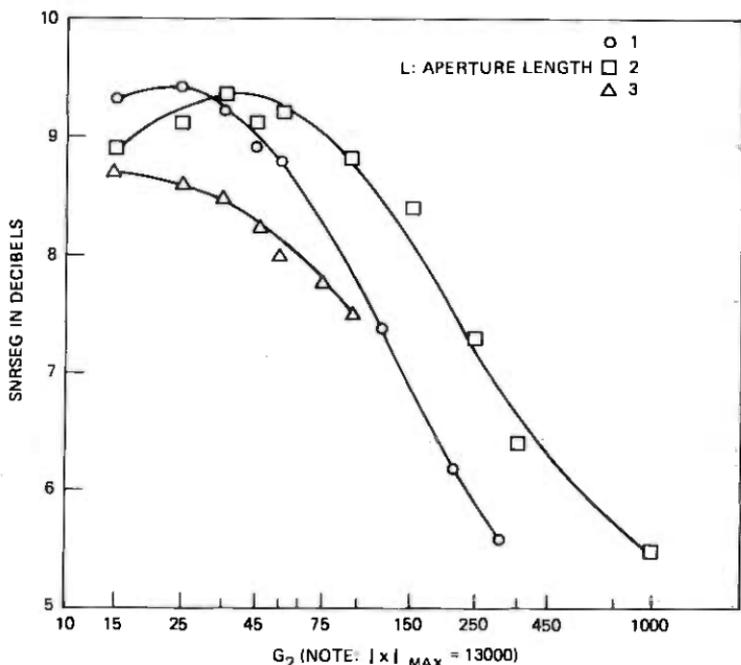


Fig. 8—Variation of speech quality SNRSEG with adaptation parameter G_2 , for $L = 1, 2$, and 3 (8-kHz sampling; nonadaptive prediction).

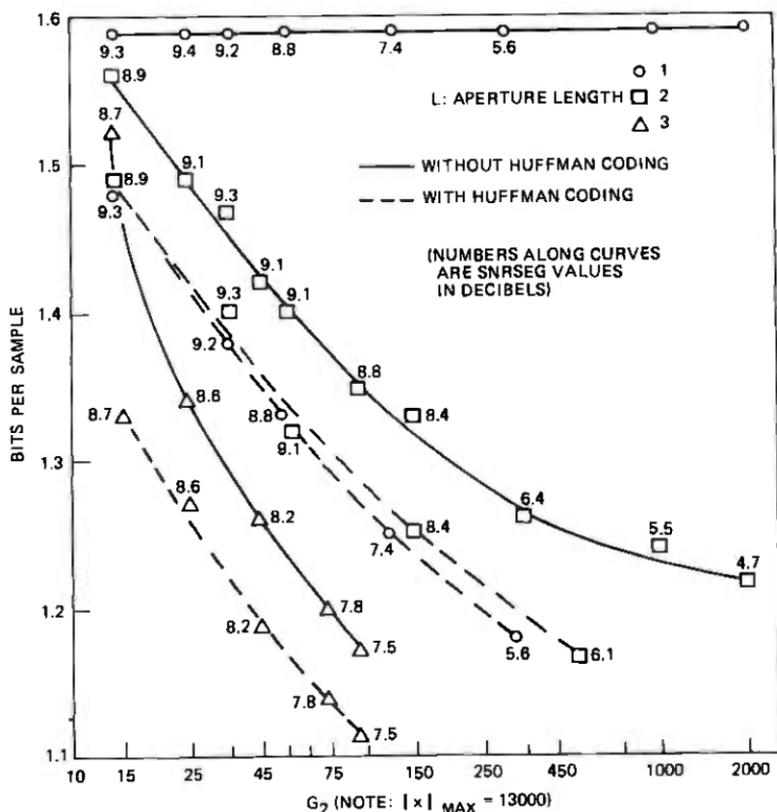


Fig. 9—Variation of transmission rate (bits/sample) with adaptation parameter G_2 , for $L = 1, 2$, and 3 (8-kHz sampling; nonadaptive prediction).

Meanwhile, it should be noted that, for a given value of L , and in the neighborhood of an SNRSEG-maximizing G_2 value, increasing G_2 tends to make the output speech more granular and harsh; while decreasing G_2 tends to make the speech more low-passy and muffled. Finally, the "design" parameter in Fig. 8 is strictly the ratio of G_2 to $|X|_{MAX}$, the maximum input speech magnitude, rather than the absolute value of G_2 .

5.2 Average transmission rates

The average (information) transmission rate in an aperture coding scheme is upper-bounded in the form

$$I(L) \leq p \cdot \log_2(2L + 1) \text{ bits/sample}, \quad (3)$$

where p is the probability of a nonredundant sample and $(2L + 1)$ is the number of distinct output points on the aperture characteristic.

The inequality above recognizes the fact that the $(2L + 1)$ output

code points, in general, have unequal probabilities of being used, so (for a given $p < 1$) further information compression can be achieved by assigning relatively short code words to frequent outputs and using relatively long code words for the infrequently occurring outputs. Thus, in the example of Table II, the variable-length Huffman coding results in an average bit rate of $1.34 < 1 \cdot \log_2 3 = 1.59$ bits/nonredundant sample. Note that for $L = 1$, $p = 1$ by definition because the 3-point aperture characteristic always puts out a nonredundant output corresponding to $P1$, $N1$, or Z . In fact, this special case is no more than *ternary (3-level) delta modulation*, and the output has a *constant* rate of $\log_2 3 \sim 1.6$ bits/sample. The effect of Huffman coding, however, is to make the output bit rate variable. With $L = 2$ or 3 , on the other hand, the output rate is variable even without Huffman coding, because $p < 1$ in general, for these cases. Information rates for $L = 1, 2$, and 3 are shown in Fig. 9 as a function of G_2 , with and without entropy (Huffman) coding in each case.

Figure 9 also includes, for convenience, the SNRSEG information from Fig. 8. For an average bit rate of 1.6 bits/sample, ternary delta modulation ($L = 1$) without Huffman coding is an obvious choice: there is no motivation for aperture coding and the attendant variability in the encoder output rate. For average bit rates of about 1.4 bits/sample, one has the choice: $L = 1$ with Huffman coding or $L = 2$ without Huffman coding. It is apparent that, for the greatest reductions of information rate (say, $I(L) = 1.2$ bits/sample), one needs to employ nontrivial ($L > 1$) aperture coding, an observation that is also suggested by the literature on adaptive asynchronous delta modulation.⁷ In our scheme, the justification for $L = 3$ comes directly from the fact that values of G_2 that realize 1.2 bits/sample encoding are far too suboptimal (SNRSEG-wise) in the cases of $L = 1$ and $L = 2$ (see Fig. 8).

VI. ADAPTIVE PREDICTION

We have studied the performance of an adaptive aperture coding scheme where the waveform predictor is also adaptive. In the interest of simplicity, we have confined our studies to the case of first-order prediction. In this case, the adaptive prediction procedure is simply to compute the adjacent sample correlation c_1 of input speech samples X

Table II—Huffman coding example ($L = 1$, $G_2 = 45$)

Sign	Run Length	Probability	Code Word
+	1	0.17	0 0
-	1	0.17	0 1
	>1	0.66	1
Transmission Rate	= $0.17 \cdot (2) + 0.17 \cdot (2) + 0.66 \cdot (1)$ = 1.34 bits/sample		

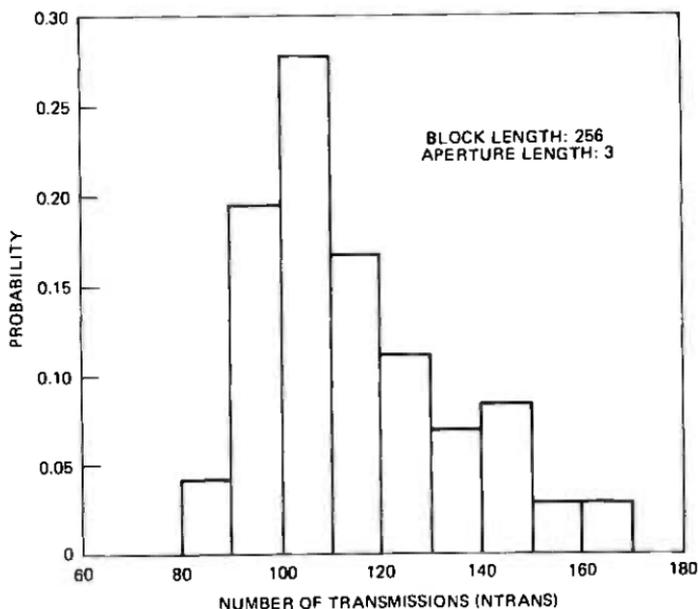


Fig. 10—Histogram of the number of transmissions in a 256-sample block ($L = 3$).

(typically, once for each input block of 256 samples), and to set the predictor coefficient a_1 equal to c_1 :

$$a_1 = c_1 = \frac{E[x_r x_{r-1}]}{E[x_r^2]}; \quad E(\cdot): \text{expected value.} \quad (4)$$

The use of adaptive prediction did not increase the SNRSEG values of Fig. 9 drastically, but perceptual improvements in coded speech quality were quite significant. The resulting speech output with 1.2 to 1.6 bits/sample aperture coding has *communications* quality: the degradation is obvious in a direct comparison with the input speech, but the quality should nevertheless be adequate for many communication purposes. The output speech quality also varies with input speech: with certain types of input, the output speech is highly *intelligible* even with nonadaptive prediction and 8-kHz sampling. The speech *quality*, however, improves significantly with adaptive prediction and faster sampling (say, 12kHz), and with adaptive low-pass filtering of the output.⁶ Finally, in informal comparisons with adaptive delta modulation (ADM) at a given bit rate, adaptive aperture coding is clearly better, as expected.

VII. VARIABILITY OF BIT RATE IN THE APERTURE CODER OUTPUT

Aperture coding schemes, with the exception of the special case of ternary delta modulation without Huffman coding, generate variable-rate outputs. For example, Fig. 10 shows a histogram of sample values

of N_{TRANS} , the number of transmissions per 256-sample block, in a scheme with $L = 3$. Note that the nonredundant sample probability p varies in the range $0.31 \leq p \leq 0.66$.

If a variable-rate procedure is used to decrease the average bit rate, one needs the additional provision of a bit buffer to be able to deliver bits at a constant rate into a channel (that accepts them in that format). The necessary length of such a buffer can be equated to the peak-to-peak variation of the quantity

$$\sum_{u=1}^N B_u - \bar{B}.N, \quad (5)$$

where B_u is the number of bits used to encode speech sample u ($B \geq 0$), N is the total number of speech samples in a (statistically long enough) test input, and \bar{B} is the average bit rate (bits/sample) needed to transmit that test input.

Using the sentence-length utterance mentioned earlier, we evaluated the peak-to-peak excursion of (5) for three cases: (i) $L = 1$ plus Huffman coding, $G_2 = 35$; (ii) $L = 2$ without Huffman coding, $G_2 = 35$; and (iii) $L = 3$ with Huffman coding, $G_2 = 45$. Cases (i) and (ii) correspond to $\bar{B} = 1.4$, and case (iii) is $\bar{B} = 1.2$. Respective buffer requirements were approximately 600, 400, and 800 bits. Respective encoding delays (for 8-kHz sampled speech and appropriate \bar{B} values) are approximately 50, 35, and 80 ms.

For speech transmission applications, the above delays are significant if not prohibitive. Furthermore, in practical designs of aperture coding, one should specify a maximum buffer length, and they should include an automatic procedure⁸ for increasing or decreasing the local average rate \bar{B} , depending on current buffer status as given by (5). Clearly, the parameter G_2 would be a natural means for controlling local values of \bar{B} .

In multiplex-speech situations, active (high output bit-rate) and inactive (low output bit-rate) segments get more intermixed in time than with a single speech channel. Consequently, buffering problems are expected to be less severe with multiplex-speech inputs. In fact, there is at least one "digital TASI" application, SPEC (Speech Predictive Encoded Communications), which indeed employs a simple form of aperture coding.⁹

The most straightforward application of aperture coding will perhaps be in the context of speech storage. In storage applications, encoding delays are less objectionable than in transmission, and buffer overflow problems, if any, need not be combatted in real time.

REFERENCES

1. C. M. Kortman, "Redundancy Reduction—A Practical Method of Data Compression," Proc. IEEE, 55, No. 3 (March 1967), pp. 253-263.

2. L. Ehrman, "Analysis of Some Redundancy Removal Bandwidth Compression Techniques," *Proc. IEEE*, 55, No. 3 (March 1967), pp. 278-287.
3. L. D. Davisson and R. M. Gray, ed., *Data Compression*, Halsted Press, 1976.
4. K. Murakami, K. Tachibana, H. Fujishita, and K. Omura, "Variable Sampling Rate Coder," *Technol. Report*, Vol. 26, Univ. of Osaka, Japan, October 1976, pp. 499-505.
5. N. S. Jayant, "Pitch-Adaptive DPCM Coding of Speech with Two-Bit Quantization and Fixed Prediction," *B.S.T.J.*, 56, No. 3 (March 1977), pp. 439-454.
6. N. S. Jayant and S. A. Christensen, "Adaptive Aperture Coding of Speech Waveforms—II," unpublished work.
7. T. A. Hawkes and P. A. Simonpieri, "Signal Coding Using Asynchronous Delta Modulation," *IEEE Trans. on Comm.*, COM-22, No. 5 (May 1974), pp. 729-731.
8. J. J. Dubnowski and R. E. Crochiere, "Variable Rate Coding of Speech," *B.S.T.J.*, 58, No. 2 (February 1979), pp. 577-600.
9. S. J. Campanella and J. A. Sciulli, "Speech Predictive Encoded Communications," Second International Conference on Digital Satellite Communications, paper E4, Paris, November 1972.

Statistical Block Protection Coding for DPCM-Encoded Speech

By R. STEELE, N. S. JAYANT, and C. E. SCHMIDT

(Manuscript received January 3, 1979)

Blocks of speech-carrying DPCM bits are protected from transmission errors by means of explicit communication of two block statistics—the maximum and the root-mean-square (rms) values of the adjacent-sample differences in the DPCM-quantized speech. At the receiver, the maximum value is used as a cue for error-detection, while the rms value is used for a partial waveform correction procedure that provides intelligible speech at bit error rates as high as 10 percent.

I. INTRODUCTION

Block protection coding, whereby a block of data words is protected by the addition of special code words or letters, is a common feature in communication systems for noisy channels. In algebraic error detection and correction, for example, the protection is derived from parity check bits. The number of parity checks, and hence the redundancy, increases with the number of data bits protected, but the resulting error-coding procedures are quite general, being applicable to any type of data, irrespective of its source. Nevertheless, with sources such as speech, where it is not crucial to recover every speech-carrying bit without error, it is meaningful to look for certain special, compact forms of *non-algebraic* block protection. The idea is to transmit a protection word that identifies some perceptually significant parameter of a speech-waveform segment; knowing the (correct) value of this parameter, the receiver can perform error-detecting and error-correcting operations, which may be only partial in an algebraic sense (due to the compactness of the protecting procedure) but nevertheless quite adequate from a speech-perception viewpoint.

In one recent investigation¹ along these lines, each block of differential PCM words was protected by a reference PCM word that signified the speech amplitude at the end of the block. Error detection was

based on comparing the DPCM decoder amplitude at the end of the block with PCM reference. Procedures for locating (and correcting) errors within the block were simple for single errors, but fairly involved for multiple errors in a block. A very successful error-location procedure had however been noted in an earlier investigation;² this depended on the detection of a statistically unlikely change between adjacent samples in the corrupted speech signal, relative to the root-mean-square (rms) value of these differences measured over a suitably long block containing these samples. The rms parameter in Ref. 2 was obtained from the corrupted speech, and this affected the success of the procedure at high error rates (say, 5 percent or higher).

The scheme to be described in this paper recognizes and extends the statistical notions of Ref. 2 and incorporates them in a block protection system that is effective even at error rates as high as 10 percent. This statistical block protection coding (SBPC) system is discussed for the specific case of non-adaptive DPCM, but extension to an adaptive system should be possible, at least in cases where the (step size) adaptation is slow or syllabic.*

II. STATISTICAL BLOCK PROTECTION CODING (SBPC)

The SBPC system employs a simple protection code consisting of two words which represent:

(i) The maximum difference between adjacent locally decoded speech samples within the block of W samples.

(ii) The rms value of the differences between adjacent locally decoded speech samples within the block of W samples. Notice that the extremal statistic (i), together with the central statistic (ii), constitute a partial description of the PDF (probability density function) of first differences.

2.1 Transmitter

The arrangement of the DPCM encoder and the system for generating the protection code are shown in Fig. 1. Suppose that the m th block of speech samples is being processed. The input speech sample x_{mW+r} , corresponding to the r th instant in the m th block, is encoded into a quantized sample q_{mW+r} by a DPCM encoder using a uniform quantizer. The predictor is of first order, with a coefficient value of $LK < 1$. Z^{-1} represents a delay of one sample period.

Denoting the locally decoded speech sample by y_{mW+r} , the protection code words are defined in the form

* Recent studies have shown that our technique works quite well in conjunction with an adaptive procedure where the quantizer step size is constant within a block (several milliseconds or tens of milliseconds long) of samples, but is modified once at the beginning of each block in response to changing speech level.

$$d_{\max} = \text{Max}_{2 \leq r < W} |y_{mW+r} - y_{mW+r-1}|_{Q2} \quad (1)$$

$$d_{\text{rms}} = \left[\frac{1}{W-1} \sum_{r=2}^W (y_{mW+r} - y_{mW+r-1})^2_{Q2} \right]^{1/2} \quad (2)$$

The quantizer Q2 has the same number of levels as the DPCM quantizer Q1, but is arranged to quantize only positive samples. Thus, after multiplexing d_{\max} , d_{rms} and W DPCM words, the frame consists of $(W + 2)$ n -bit words. It is important, or at least very desirable, to protect the "protecting words," d_{\max} and d_{rms} , by transmitting them in a redundant format. For example, one might transmit three versions

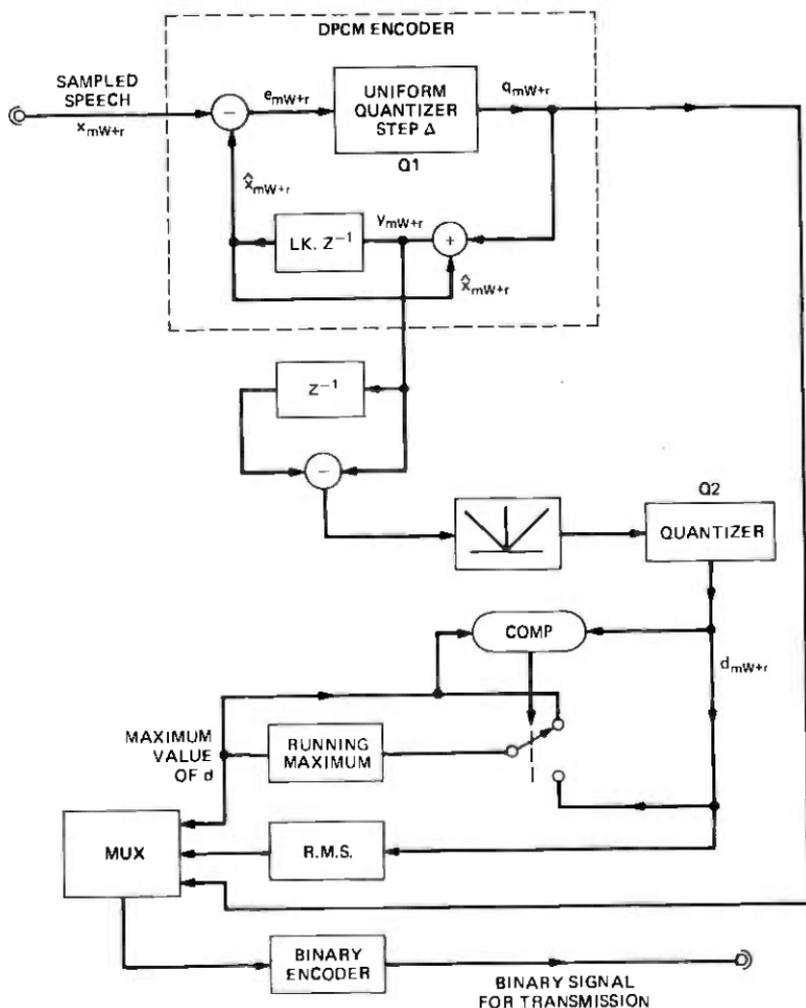


Fig. 1—SBPC encoder.

of each bit in the protecting word and decode the bit on the basis of a majority count. The overhead constituted by this protection arrangement would be $3 \times 2 = 6$ words, or 2.3 percent if $W = 256$; this overhead is much smaller than the redundancy required of an algebraic code that would correct some patterns we will discuss later in this paper.

The DPCM-encoded speech together with its simple protection code is transmitted through a channel which may cause some bits to be inverted. The probability of bit inversion is called the error rate ER .

2.2 Receiver

The receiver demultiplexes each frame into its data block and protection code. The DPCM sequence is decoded into Y_{mW+k} ; $k = 1, 2, \dots, W$. (Note that cap letters Y and D will be used to signify variables at the receiver.) Figure 2 shows the essential features of the SBPC

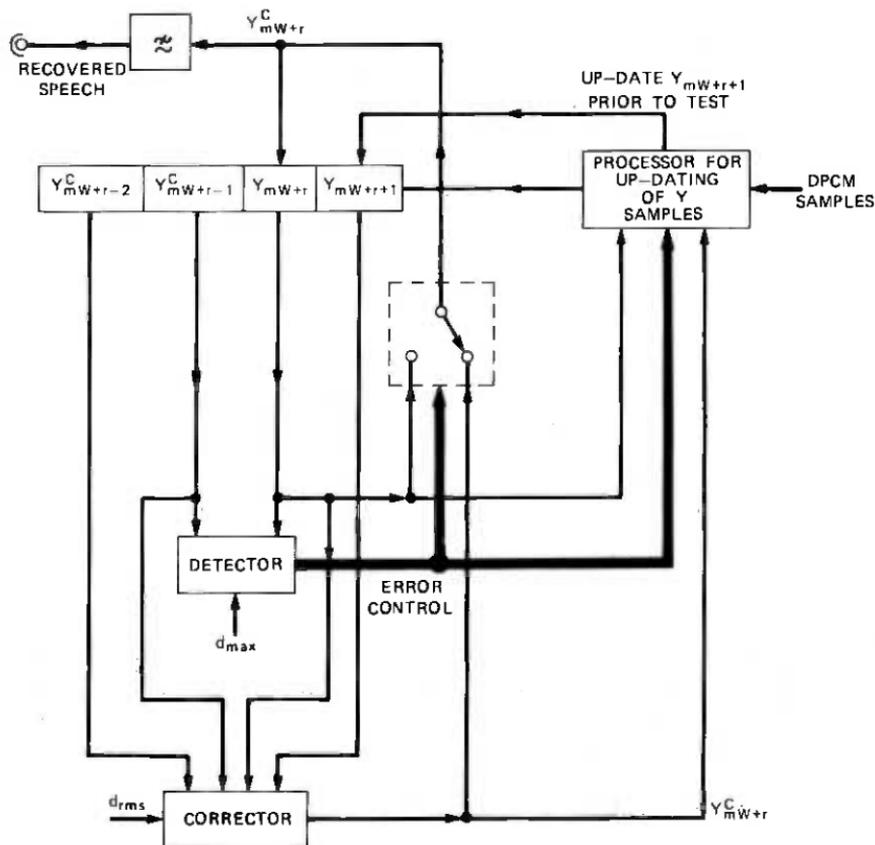
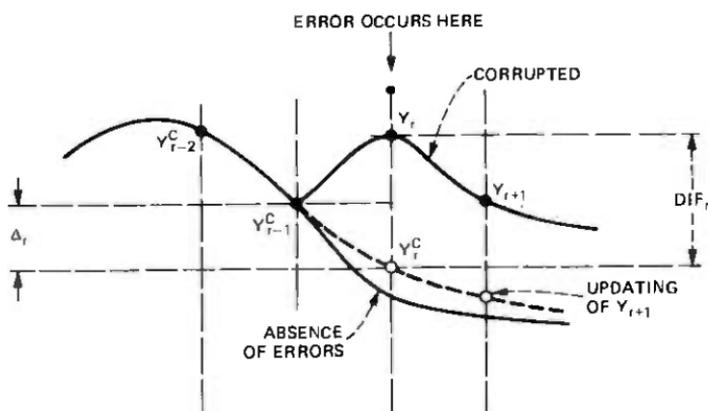


Fig. 2—SBPC decoder.



$$\begin{aligned}
 \text{DETECTION OF ERROR} & \left\{ \begin{array}{l} \text{IF } |Y_r - Y_{r-1}^C| > d_{\max} \end{array} \right. \\
 \text{CORRECTION OF ERROR} & \left\{ \begin{array}{l} Y_r^C = Y_{r-1}^C + \Delta_r \\ \Delta_r = d_{\text{rms}} \cdot \text{sgn}(Y_{r-1}^C - Y_{r-2}^C) \text{ IF} \\ \quad \text{sgn}(Y_{r-1}^C - Y_{r-2}^C) = \text{sgn}(Y_{r+1} - Y_r) \\ \quad = 0 \text{ OTHERWISE} \end{array} \right. \\
 \text{UPDATING OF SAMPLE } r+n & \left\{ \begin{array}{l} Y_{r+n} \rightarrow Y_{r+n} + (DIF)_r \cdot LK^n \\ \text{WHERE } (DIF)_r \triangleq Y_r^C - Y_r \end{array} \right. \\
 \text{(n > 0) AFTER CORRECTION OF } r &
 \end{aligned}$$

Fig. 3—Waveform correction at time r .

correction procedure. For simplicity, the demultiplexer decoders and control facilities have been omitted.

We suppose that samples $\dots, Y_{mW+r-3}^C, Y_{mW+r-2}^C, Y_{mW+r-1}^C$, have been either considered correct and passed to the output, or deemed to be in error and partially corrected; hence the superscripts C . We now test sample Y_{mW+r} . We find the quantized magnitude difference D_{mW+r} between Y_{mW+r} and Y_{mW+r-1}^C , and compare the difference with the maximum transmitted difference d_{\max} . Y_{mW+r} must be erroneous if

$$D_{mW+r} > d_{\max} \quad (3)$$

If inequality (3) is satisfied, the correction must be switched into the circuit and the erroneous Y_{mW+r} replaced by a corrected value Y_{mW+r}^C .

The corrections are described by the algorithm (Fig. 3):

$$Y_{mW+r}^C = Y_{mW+r-1}^C + \Delta_{mW+r}, \quad (4)$$

where

$$\Delta_{mW+r} = d_{\text{rms}} \text{sgn}(Y_{mW+r-1}^C - Y_{mW+r-2}^C) \quad (5)$$

if $[\text{sgn}(Y_{mW+r-1}^C - Y_{mW+r-2}^C) = \text{sgn}(Y_{mW+r+1} - Y_{mW+r})]$

$$= 0 \text{ otherwise.} \quad (6)$$

Clearly, this correction is based on a "smooth" output waveform model where the sign of the slope at time r equals that at time $r - 1$ if the latter equals that at time $r + 1$ [eq. (5)], while, if the slopes at times $r - 1$ and $r + 1$ are opposite in sign, that at time r is given the average value of zero [eq. (6)]. Furthermore, in the former case, the magnitude of the slope at time r is set to the block-specific rms value d_{rms} . Strictly speaking, the optimum setting of this magnitude would take the form $J \cdot d_{rms}$, where J would be a constant depending on the shape of the first-difference PDF.

The correction algorithm has also been deployed³ with d_{rms} being derived from the corrupted speech. With large values of the error rate ER , this would give rise to poor corrections. By explicitly transmitting the value of d_{rms} , the corrections are significantly improved.

2.3 Updating samples following a correction

Having made the correction to sample Y_{mW+r} , we remove the error from the subsequent samples before we continue testing the next sample Y_{mW+r+1} (Figs. 3 and 4). This is done as follows. Let

$$DIF_{mW+r} = Y_{mW+r} - Y_{mW+r}^C. \quad (7)$$

As the propagation of the error is due to the integrator, and the integrator leakage factor is LK , the subsequent decoded samples are reduced to

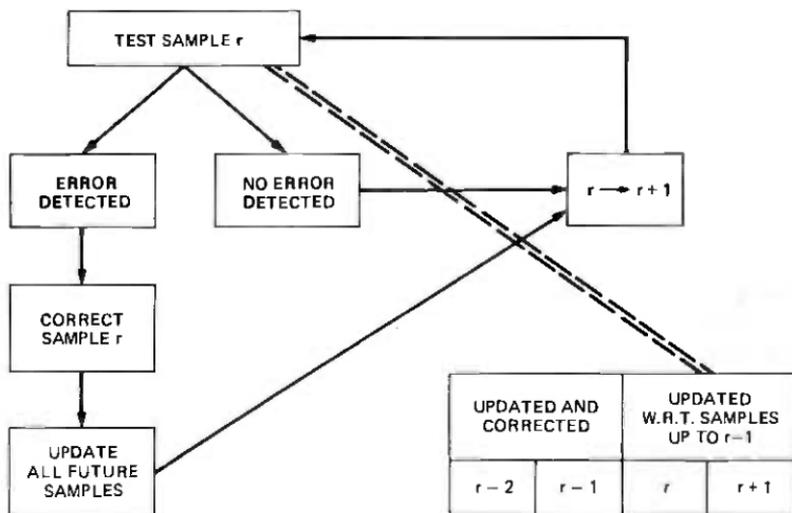


Fig. 4—Error detection, correction and sample updating.

$$Y_{mW+r+n} = Y_{mW+r+n} - (LK)^n DIF_{mW+r}$$

$$n = 1, 2, \dots, W - r. \quad (8)$$

The value of the error at the end of the block is

$$E_{(m+1)W} = (LK)^{W-r} DIF_{mW+r}, \quad (9)$$

and this is stored.

For each correction, the propagation of the error is removed leaving a residue at the end of the block whose size depends on the position of the sample corrected, as shown by eq. (9). These residuals are summed to give the total residual propagation error $E_{(m+1)W}^{(T)}$. When the next block of DPCM samples are processed, each one is modified to remove the propagation effects from errors in the previous block:

$$Y_{W(m+1)+r} = Y_{W(m+1)+r} - (LK)^r E_{(m+1)W}^{(T)}$$

$$r = 1, 2, \dots, W. \quad (10)$$

The detection and correction method epitomized by eqs. (3) and (8) are again used to process the DPCM samples in eq. (10).

2.4 Organization of the transmission block

The first part of the transmission frame contains the protection code. It is placed there in order for the detection and correction process to begin immediately. When testing Y_{mW+1} , samples Y_{mW} and Y_{mW-1} from the previous block must be available. When testing the last sample, $Y_{(m+1)W}$, the first sample $Y_{(m+1)W+1}$ from the next data block is required. Consequently, the total delay of the decoded speech is $(W + 1)$ sampling intervals.

The larger the value of W , the smaller the fractional increase in required channel capacity (due to the protection-word overhead), but the longer the decoding delay at the receiver output.

III. RESULTS AND DISCUSSION

The block protection scheme was simulated on a Data General Eclipse computer. The band-limited input signal, a single sentence spoken by a male, was sampled at 8 kHz prior to encoding by a uniform 7-bit DPCM encoder with predictor coefficient $LK = 0.9$. The coding of the quantizer output levels was such that an error in the most significant bit caused an error in the received sample equal to half the range of the quantizer.

The DPCM code words were assembled into blocks of W words with the protection code previously described. The DPCM code words were subjected to random errors, but the protection code words were left uncorrupted.

As a supplement to listening tests, the segmental signal-to-noise ratio,⁴

$$\text{SNRSEG} = \text{Average} [\text{short-time SNR}(\text{in dB})], \quad (11)$$

was used, an objective performance criterion. The short-time SNR in (11) is a statistic computed over an interval typically 16 to 32 ms long. By performing the decibel operation prior to long-time averaging, the SNRSEG measure preserves information about how well the low-level segments of speech are reproduced.

Figure 5 shows the variation of SNR as a function of amplitude scaling AS of the input speech signal. From the zero error rate curve, it can be seen that optimum loading occurs for $AS = 0.04$. When $ER = 4.2$ percent, the decoded signal is very corrupted and SNRSEG is reduced by 40 dB in the underloaded condition. However, the SBPC system dramatically improves the performance of the DPCM system, increasing the SNRSEG by 11 dB for $AS = 0.04$, and by 19 dB for $AS = 0.01$.

The unusual characteristic of the SBPC system is that, with large values of ER , the variation of SNRSEG is substantially independent of AS . This is a property found in adaptive DPCM. The reason for the nearly flat SNR characteristic is: In the presence of low level speech, d_{\max} is a low number, and if many errors occur there will be numerous occasions when the differences between adjacent samples in the corrupted decoded signal exceed d_{\max} . These erroneous differences are identified and will be partially corrected. Only those errors which

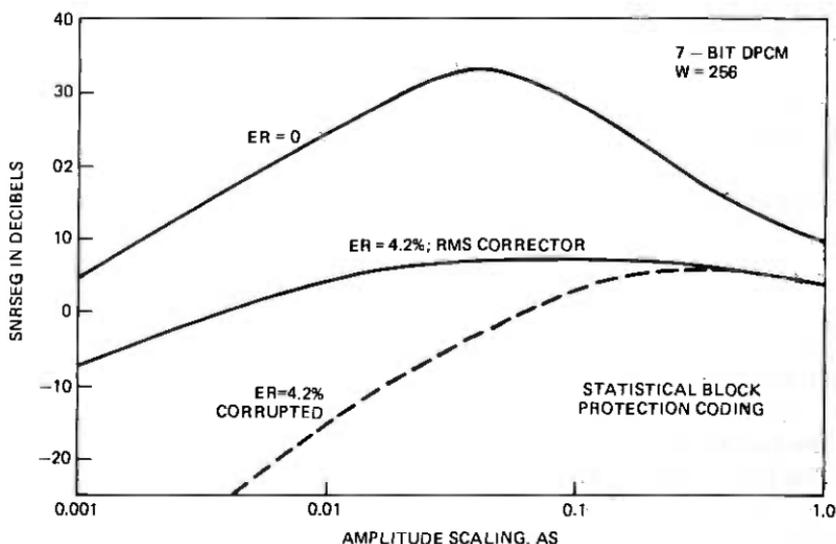


Fig. 5—SBPC gain as a function of input speech level AS .

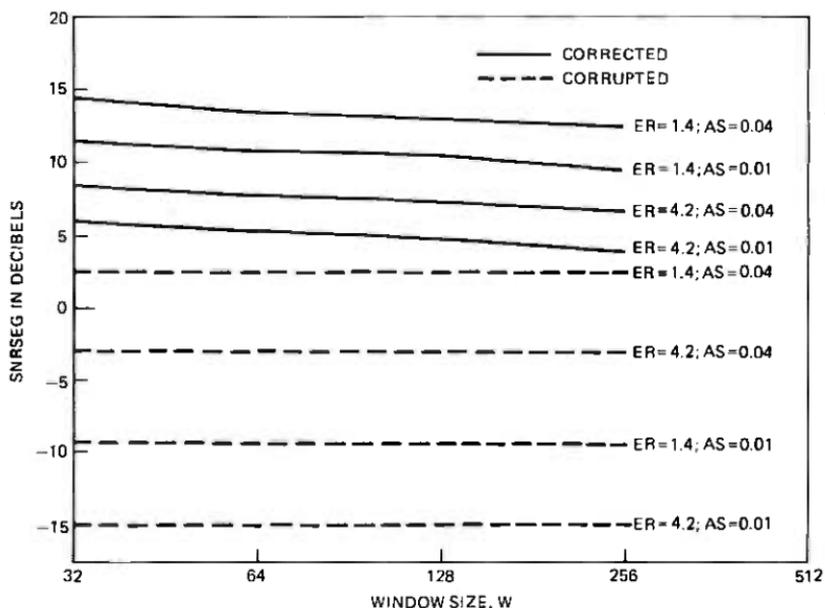


Fig. 6—SBPC gain as a function of block length W .

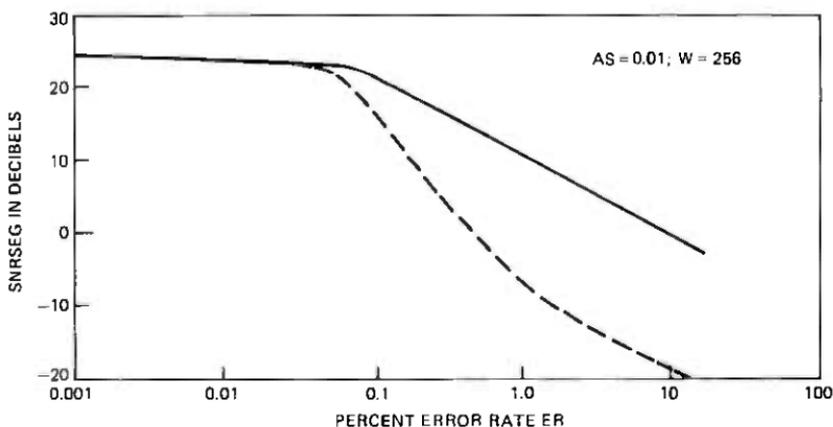


Fig. 7—SBPC gain as a function of error rate ER .

result in differences less than d_{\max} will be missed. However, when the coder is occasionally experiencing some overloading ($AS > 0.04$, say), the maximum value d_{\max} in some blocks will merely reflect quantizer saturation, rather than providing a cue for detecting transmission errors, and improvements are now gained only in purely unvoiced or silent intervals in speech.

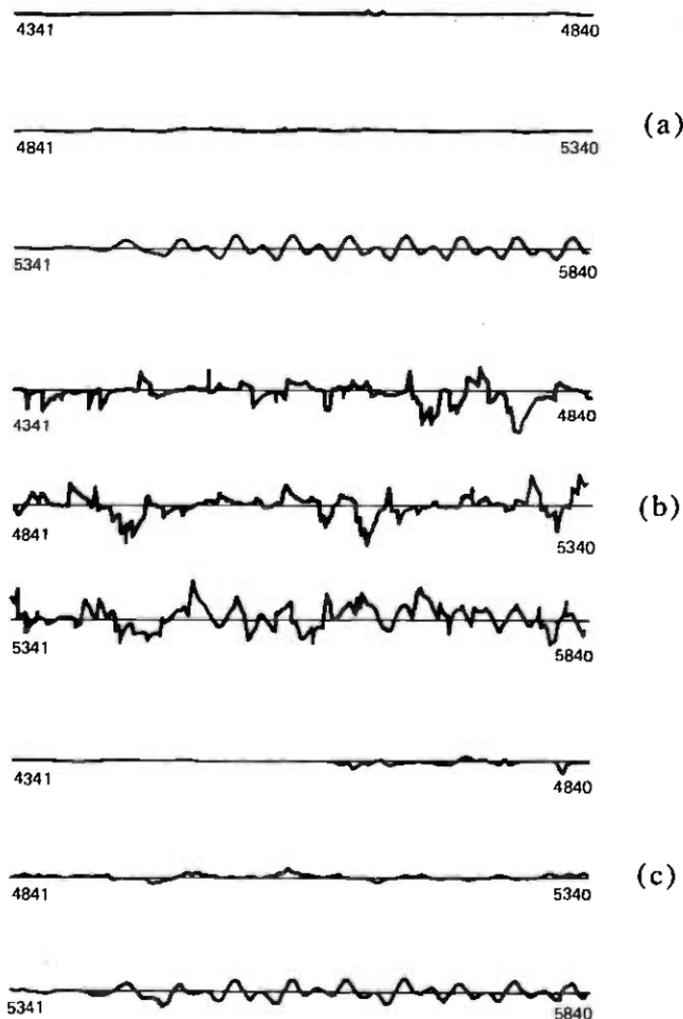


Fig. 8—Waveforms of (a) original, (b) corrupted, and (c) corrected speech ($ER = 10\%$).

At $AS = 0.01$ and $ER = 4.2$ percent, the corrupted speech is perceptually very poor, and sounds almost like bandlimited white noise. By using the SBPC system, the speech is rendered intelligible, although of poor quality. The overall perceptual improvement is dramatic.

The variation of SNR_{SEG} as a function of block size W is shown in Fig. 6. Increasing W from 32 to 256 results in a decrease in SNR_{SEG} of less than 2 dB. The near-independence of SNR from W is perhaps related to the fact that none of the W values used is large enough to encompass a significantly nonstationary segment of speech.

The gain is SNRSEG as a function of ER is shown in Fig. 7 for $AS = 0.01$, $W = 256$. The objective gain is very slight for low error rates (say, $ER < 0.1$ percent), but significant for high error rates (say, $ER > 0.5$ percent). In particular, the gains are quite dramatic with $ER = 10$ percent. These objective gains are well reflected by the perceptual gains noticed in informal listening tests, and by the illustrative speech waveforms in Fig. 8.

REFERENCES

1. R. Steele, D. J. Goodman, and C. A. McGonegal, "DPCM with Forced Updating and Partial Correction of Transmission Errors," *B.S.T.J.*, 58, No. 3 (March 1979), pp. 721-728.
2. R. Steele and D. J. Goodman, "Detection and Selective Smoothing of Transmission Errors in Linear PCM," *B.S.T.J.*, 56, No. 3 (March 1977), p. 399.
3. R. Steele, D. J. Esdale, and D. J. Goodman, "Partial Correction of Transmission Errors in DPCM Without Recourse to Error Correction Coding," *Elec. Lett.*, 13, No. 12 (June 9, 1977), p. 351.
4. P. Noll, "Adaptive Quantizing in Speech Coding Systems," International Zurich Seminar, Zurich, Switzerland, April 1974.

An ADPCM Approach to Reduce the Bit Rate of μ -Law Encoded Speech

By H. W. ADELMANN, Y. C. CHING, and B. GOTZ

(Manuscript received March 14, 1979)

A new ROM-oriented ADPCM architecture for processing μ -law encoded speech is presented. The architecture is shown to offer: (i) simple hardware realization, (ii) flexibility of algorithm modification, (iii) fixed hardware complexity with respect to numerous nonlinear processings, and (iv) excellent time-sharing capability. Performance measurements of a hardware implementation are also included.

I. INTRODUCTION

Bit rate reduction from the present 64 kb/s per trunk is of interest both to reduce per-trunk cost of existing and planned digital transmission facilities and to prove-in digital transmission.

Adaptive differential PCM (ADPCM) coding of the type reported in Refs. 1 through 4 appears promising for bit rate reduction of *speech* signals. Our specific interest is the bit rate reduction of the conventional μ -255, 64-kb/s signal. The purpose of this paper is to present an ADPCM architecture that offers: (i) simple hardware realization, (ii) flexible algorithm modification, (iii) fixed hardware complexity with respect to numerous nonlinear processings, and (iv) time-sharing capability over many trunks. The performance of a hardware realization of such a coder, which is shared by 24 voice trunks, is also reported.

II. THE PROPOSED ADPCM ARCHITECTURE

2.1 Basic structure

DPCM with adaptive backward quantization and fixed one-tap predictor, as proposed and analyzed in Refs. 1, 2, and 3, is our point of departure. Figure 1 is the conventional block diagram of an ADPCM codec and its interface to a μ -law environment. The quantizer with scale (e.g., uniform step size) Δ is denoted by Q_{Δ} , while the "inverse

quantizer" which maps a quantization interval code IQ into a quantized signal level is denoted by Q_{Δ}^{-1} . The quantization scale $\Delta(i)$ is determined by a scale adaptation algorithm from the quantization index sequence $IQ(i)^*$, which is sent over the channel. We add a prime to the decoder variables.

Prior research on ADPCM assumes a fixed linear signal representation. In the case of the μ -law signal environment, this requires μ -to-linear (μ/L) and L/μ conversions, as shown in Fig. 1. The 14 bits per sample required by the linear representation is an indication of the complexity of the straightforward approach.

Our basic approach is to put as much of the ADPCM algorithm as is presently practical into read-only memory (ROM). This allows us to confine the 14-bit linear representation to the ROM table computation (i.e., to the firmware). We motivate our architecture by relating it to the more familiar architecture of Fig. 1. For this purpose, assume a uniform B -bit quantizer which can be considered to be a cascade of an infinite amplitude quantizer and a B -bit limiter. The argument to follow could be generalized to a nonuniform quantizer. (Note that the limiting is on the quantizer output code as opposed to input amplitude.) Retaining the limiter in its place but shifting the quantizer and inverse quantizer past the difference and summing nodes† produces Fig. 2 from Fig. 1. (For convenience, only a local decoder is shown together with the new notation.) To complete the transition to the proposed architecture as shown in Fig. 3, we first combine μ/L and Q_{Δ} into one block for $\mu/APCM$ conversion via a ROM realization. We next note that, instead of \hat{X}_L , we could store any other absolute representation of the local decoder estimate; in particular, we use the μ -law representation. Combining Q_{Δ}^{-1} and L/μ into an $APCM/\mu$ block and combining the μ/L , predictor multiplication and Q_{Δ} into a $(\mu/APCM)_A$ block, we obtain the proposed architecture shown in Fig. 3 where $\mu/APCM$, $APCM/\mu$, $(\mu/APCM)_A$, and SCALE ADAPTATION are to be implemented with ROMs.

The effects of these actions are discussed below. By moving the adaptive quantizers through the difference node, we can combine many nonlinear functions, which are difficult to realize in combinatorial logic, into ROM tables. The generation of these tables can be achieved with an off-line computer, where the 14-bit precision inherent in the μ -law code is used. In addition, the outputs of these ROMs can be restricted to 8-bit APCM words that can be easily manipulated. We next examine in more detail our proposed quantization and scale adaptation scheme.

* $IQ \in \{\pm 0, \pm 1, \dots, \pm 2^{B-1} - 1\}$, where B is the number of bits/sample sent over the channel.

† Note that $Q(x + y) \approx Q(x) + Q(y)$ to within one quantizer interval.

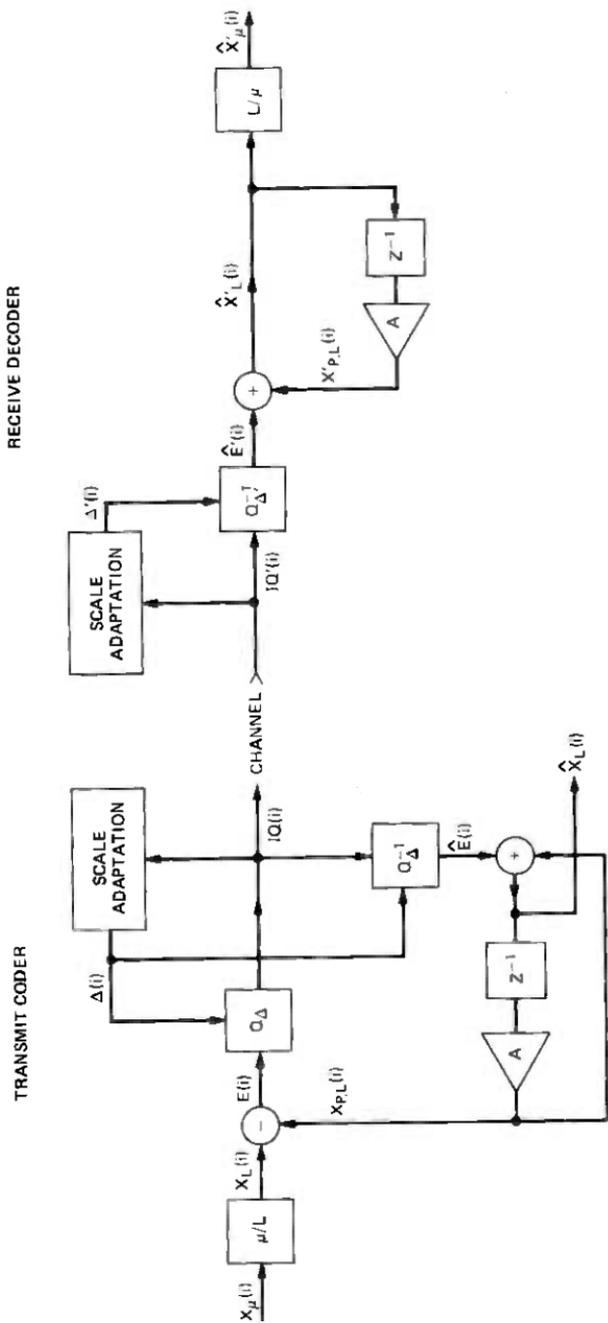


Fig. 1—Conventional ADPCM structure and interface to μ -law environment.

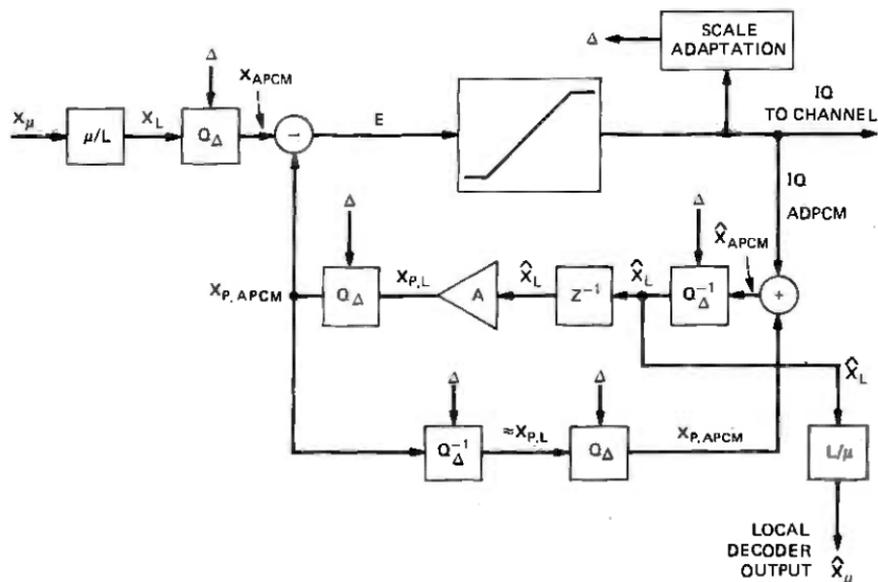


Fig. 2—Relating a conventional to a proposed ADPCM structure.

2.2 Quantization and scale adaptation details

The ROM architecture of Fig. 3 permits a very general interpretation of scale as scale index d and of Q_d and Q_d^{-1} as corresponding arbitrary quantizers. In this general framework, the adaptation rule of Refs. 1, 2, and 3 evolves to the following heuristic scale index adaptation and corresponding quantization constraint: For large (small) magnitude prediction error, the index d is increased (decreased) and the corresponding quantization Q_d and Q_d^{-1} is made coarser (finer). The quantization we propose is a marriage of uniform and μ -255 quantization; we call it semiuniform quantization. The semiuniform quantization and corresponding adaptation strategy is best explained by starting with uniform quantization where scale is synonymous to step size of the uniform quantizer.

Denoting the uniform quantizer step by Δ , the adaptation algorithm is

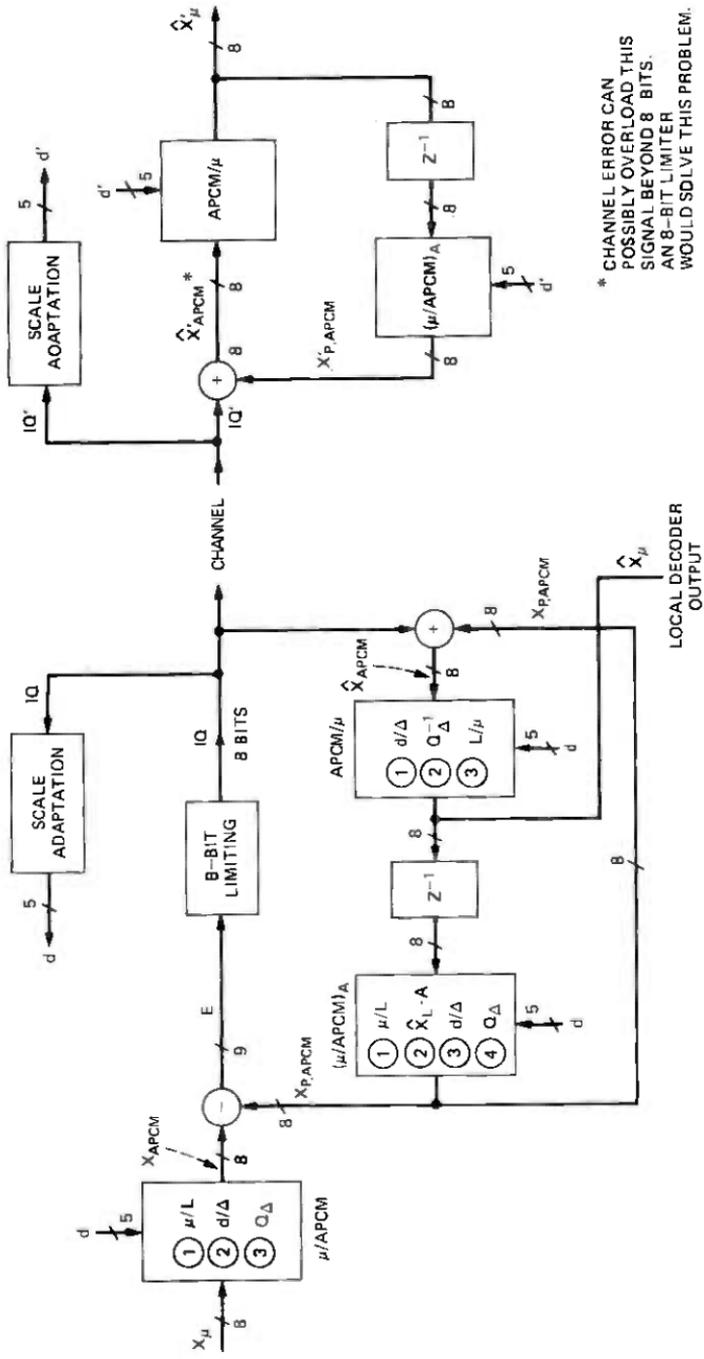
$$\Delta(i+1) = \Delta(i)M(i), \quad (1)$$

where $M(i) \triangleq M(|IQ(i)|) \triangleq M_{|IQ(i)|}$ are positive and have the ordering

$$M_0 \leq M_1 \leq \dots \leq M_{2^B-1},$$

where

$$M_0 < 1 \text{ and } M_{2^B-1} > 1. \quad (2)$$



* CHANNEL ERROR CAN POSSIBLY OVERLOAD THIS SIGNAL BEYOND 8 BITS. AN 8-BIT LIMITER WOULD SOLVE THIS PROBLEM.

Fig. 3—The proposed ADPCM architecture.

In Refs. 2 and 5, optimized adaptation vectors

$$\vec{M} = (M_0, M_1, \dots, M_{2^B-1})$$

are proposed and related to quantizer input variance estimation. It will be convenient to work with the log transformed version of (1)

$$d(i+1) = d(i) + m(i), \quad (3)$$

where $\Delta = \Delta_0 q^d$ and $M = q^m$ for some convenient fixed Δ_0 and base q . The scale adaptation can thus be done with the simpler iteration (3), while the translation from d to Δ (d/Δ) is done in the ROM generating software, as indicated in Fig. 3. We next discuss the finite precision aspects of a digital realization and the associated ROM size implications.

Consider first the scale adaptation algorithm. We assume d to be a K -bit nonnegative integer, $0 \leq d \leq 2^K - 1$, representing 2^K permissible scales. Thus, $\Delta_{\min} = \Delta_0$ and $\Delta_{\max} = \Delta_0 q^{(2^K-1)}$. To permit the ADPCM coder to approximate the idle channel performance of the μ -law coder, we let the minimum step size Δ_0 equal the first chord μ -law step size which, for the 14-bit linear representation we use, corresponds to $\Delta_0 = 2$. To accommodate the 40- to 50-dB dynamic range of a speech signal, it is desired that

$$\frac{\Delta_{\max}}{\Delta_{\min}} = q^{(2^K-1)} \quad (4)$$

be between 100 and 300. The step-size resolution factor q should not be greater than 2, which is the μ -law resolution; nor is much to be gained by making q too close to 1. A useful design range is $q \in [2^{1/4}, 2]$ and $K \in [3, 5]$ with a smaller q requiring the larger K . Figure 3 assumes $K = 5$, although a smaller value might be sufficient. Figure 4 is a block diagram for the scale adaptation. We show a scale adaptation preprocessor where the adaptation increment index IA is based on the quantizer output past history such as min-max, averaging, or other.* For concreteness, we assume IA to be the three most significant magnitude bits of IQ . We have introduced F additional fractional bits for the internal scale index which we designate by d_I . The scale adaptation iteration is

$$d_I(i+1) = \begin{cases} 0 & \text{if } d_I(i) + m(i) < 0 \\ d_I(i) + m(i) & \text{otherwise} \\ d_{\max} & \text{if } d_I(i) + m(i) > d_{\max}, \end{cases} \quad (5)$$

* In the hardware realization we have found useful a two-word memory preprocessor where the most significant previous $|IQ|$ bit in addition to the two current most significant $|IQ|$ bits determines the current IA .

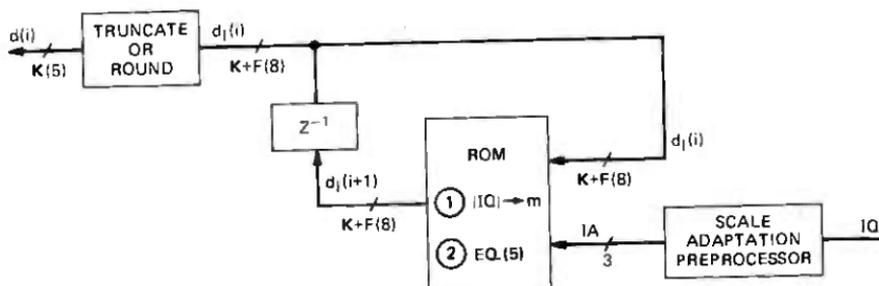


Fig. 4—Scale adaptation block diagram.

where d_l and $m(i)$ have K -integer plus F -fractional-bit representation and $d_{\max} \leq \log_q(16,000/2^{B-1})$ (where the upper bound on d_{\max} corresponds to the step size necessary to accommodate the largest possible prediction error (16,000) without overload).* The F fractional bits are useful for two reasons. First, they permit adaptation design flexibility even with $q = 2$; for example, with $q = 2$ and $F = 0$, the largest step-reducing multiplier (closest to 1) is $2^{-1} = 0.5$ while with $F = 3$ the largest step-reducing multiplier is $2^{-0.125} \approx 0.9$. Second, they are useful in the control of adaptation mistrack due to channel errors.^{6,7} Figure 4 assumes $K = 5$, $F = 3$, and up to 8 distinct adaptation multipliers implying a $2^{11} \times 8$ ROM.

Consider next the quantization. For strictly uniform quantization, large amplitude, slowly varying ADPCM input signals would generate small step sizes and large magnitudes for X_{APCM} , \hat{X}_{APCM} , and $X_{P,APCM}$. These large magnitudes might require a large number of bits to avoid amplitude overload. The amplitude overload problem as a function of the number of bits/word (≤ 14) must be assessed. We propose to avoid the amplitude overload problem with the following semiuniform quantization scheme. For each scale index d , there is a threshold value $T(d)$ such that for $X_L < T(d)$ the quantization is uniform, while for $X_L > T(d)$ the quantization is a copy of the μ -law quantizer as illustrated in Fig. 5. The threshold $T(d)$ is located at approximately a μ -law segment boundary such that, for the segments below $T(d)$, the segment step sizes are $\leq \Delta = \Delta_0 q^d$ while, for the segments above $T(d)$, the segment step sizes are $> \Delta$. Thus wasteful quantizer level assignments (and therefore bits) are avoided when the final μ -law quantization would ignore them anyway. Figure 6 shows the details of uniform and semiuniform quantization assuming $\Delta = 3$. The important feature of the semiuniform quantization is that the number of quantizing inter-

* For example, for $q = \sqrt{2}$ and $B = 4$, $d_{\max} \approx 22$ and therefore $K = 4$ with $d_{\max} = 15$ or $K = 5$ with $d_{\max} \in [16, 22]$ should be satisfactory.

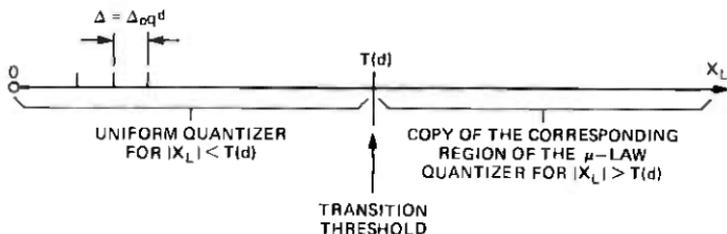


Fig. 5—Basic structure of semiuniform quantizer (positive half).

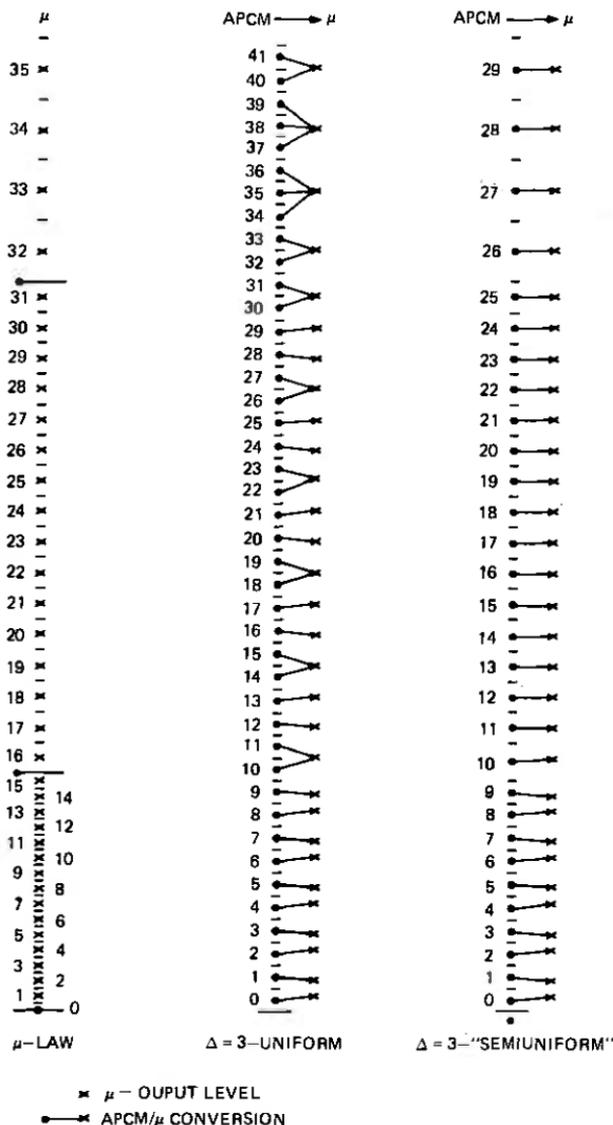


Fig. 6—Illustrating uniform and semiuniform quantization.

vals is ≤ 256 and therefore 8 bits is sufficient for the representation of X_{APCM} , \hat{X}_{APCM} , and $X_{P,APCM}$. The blocks $\mu/APCM$, $APCM/\mu$, and $(\mu/APCM)_A$ including a sign magnitude/2's-complement or inverse-conversion can each be realized with a $2^{13} \times 8$ ROM, or alternatively one can use $2^{12} \times 7$ ROMs with the sign bit feeding around the ROMs to separate sign magnitude/2's-complement converters.

Although the optimizations reported in Refs. 2, 3, and 5 do not quantitatively apply to our ADPCM coder, the qualitative concepts developed there with respect to loading, adaptation speed, and others are very useful in tuning our coder, and they could be utilized in our hardware realization.

The next section briefly discusses two approaches to dealing with transmitter-receiver scale mistrack due to channel errors. The discussion serves to illustrate the versatility of our ROM architecture.

III. CHANNEL ERROR AND SCALE MISTRACK

In a backward adapting quantizer, the possibility of scale mistrack due to channel errors must be considered. A robust quantizer is obtained by modifying (3) [and similarly modifying (5)] to

$$d(i+1) = \beta d(i) + m(i), \quad (6)$$

where β is some number less than but close to 1. The modification in (6) is incorporated into our architecture by a change of the scale adaptation ROM. References 6 and 7 discuss in detail performance aspects and design guidelines related to the finite arithmetic implementation of (6). Also, the improved quantizer reconstruction strategy recommended for the robust quantizer in Ref. 8 can be implemented by a change in the $APCM/\mu$ reconstruction table, i.e., the $APCM/\mu$ ROM. These improvements require no additional hardware.

An alternate approach to scale mistrack control is to modify (3) to

$$d_I(i+1) = d_I(i) + m(i, d_I(i)), \quad (7a)$$

where

$$m(i, d_I(i)) = \begin{cases} m(i) - 2^{-F} & D_2 < d_I(i) \\ m(i) & D_1 \leq d_I(i) < D_2 \\ m(i) + 2^F & d_I(i) < D_1, \end{cases} \quad (7b)$$

$\vec{m} = \log_q \vec{M}$ is an optimized adaptation vector in the absence of channel errors, and D_1 and D_2 (e.g., $D_1 = (1/3)d_{\max}$ and $D_2 = (2/3)d_{\max}$) are design constants which, for the reason to follow, we call mistrack correction levels. Each iteration that $d_I(i)$ and $d'_I(i)$ are on opposite sides of n ($n = 1$ or 2) of the correction levels, the mistrack $|d_I(i) -$

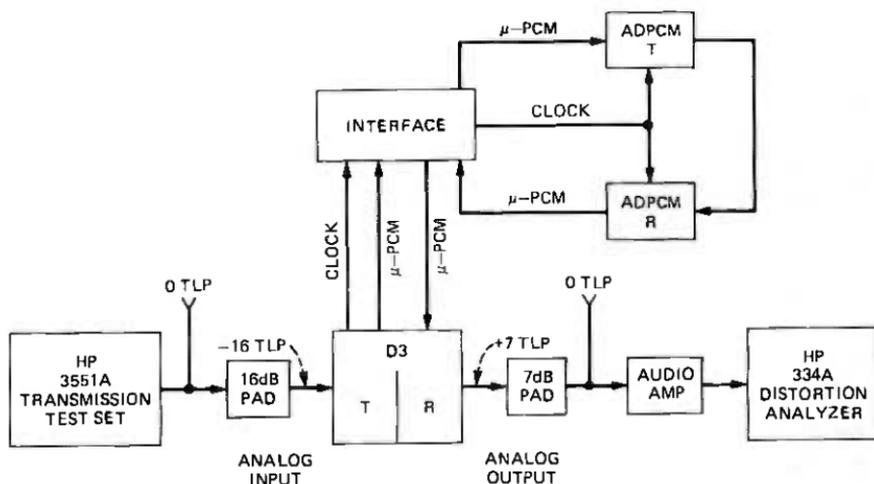


Fig. 7—ADPCM coder performance measurements.

$d_i(i)$ is reduced by $n2^{-F}$. For* $2^{-F} \ll |m_{\min}|$ the approach in (7) should avoid the error-free dynamic range penalty^{6,7} associated with (6). Again, the important point is that (7) (or a generalization of it to an arbitrary number of correcting levels) can be incorporated into the codec by a simple ROM table modification without hardware or complexity penalty.

IV. HARDWARE REALIZATION AND PERFORMANCE

We have constructed an ADPCM coder which interfaces directly with a D3 bank. Because our coder uses parallel processing and the ROMs are relatively fast, we can share the coder among all 24 voice trunks of the channel bank. In fact, over 120 trunks could be accommodated if we so desire. The coder is implemented with three wire-wrapped circuit packs: one each for the interface, the transmitter, and the receiver.

For the signal-to-noise measurements, we chose a prediction coefficient $A = 0.8$ and a scale adaptation vector of $(-1, 0, 0, 0, 1, 2, 3, 5)$ with a base $q = \sqrt{2}$. Because of the flexibility of the coder, we were able to vary the number of bits per sample feeding the transmission channel as well as the local decoder. The measurement arrangement is shown in Fig. 7. The signal-to-noise versus frequency for the coder is shown by the graph of Fig. 8. The measurements were made with the coder operating in the 5-bit ADPCM mode and with 5 bits in the

* If it is desired to have a small m such as $m = \pm 2^{-F}$ or $m = 0$, then (7b) can be simply modified to exclude the perturbation 2^{-F} for those m . The ROM modification is again straightforward. Also, interleaving as in Ref. 6 should reduce the required F .

feedback circuit. The analog input was -15 dBm0. As expected, the coder performance drops with increased frequency. Above 1500 Hz when the adjacent sample correlation is below 0.4, the differential mode of coding has no advantage and in fact has penalty. The remainder of the performance tests were made by measuring the signal-to-noise ratio versus amplitude for 1005 Hz. These measurements were made for the following coder options:

(i) ADPCM with the number of transmit bits equal to 5 and the number of feedback bits equal to 5, 4, and 3.

(ii) ADPCM with the number of transmit and feedback bits both equal to 4 and 3.

Figure 9 demonstrates the performance insensitivity as the number of feedback bits is dropped. Figure 10 shows the performance for 3 and

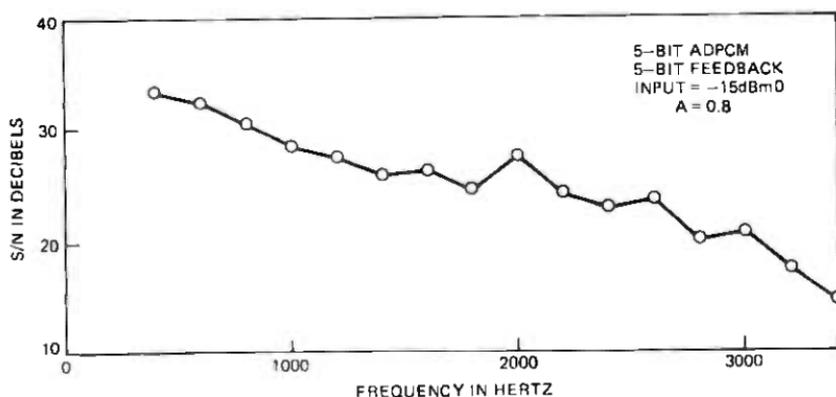


Fig. 8—ADPCM coder performance, frequency response.

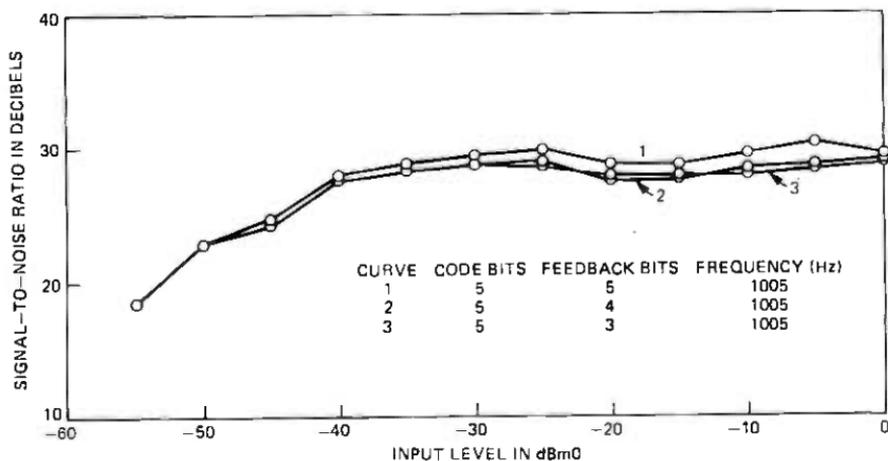


Fig. 9—ADPCM coder performance, 5-bit.

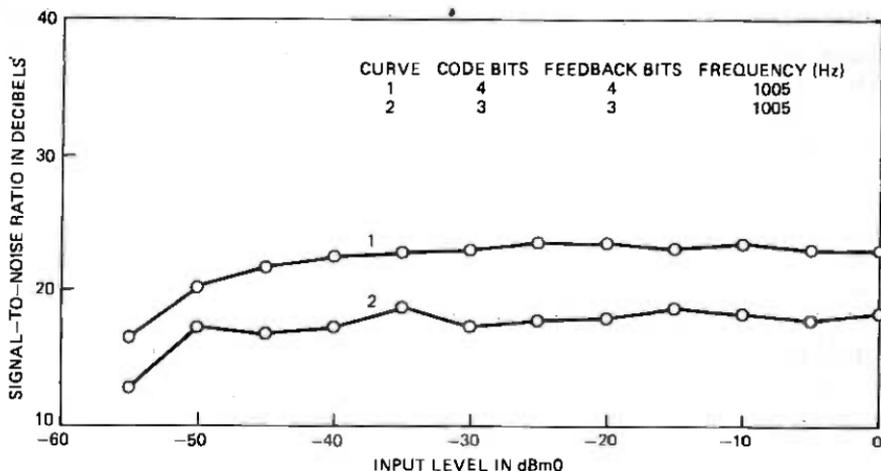


Fig. 10—ADPCM coder performance, 4- and 3-bit.

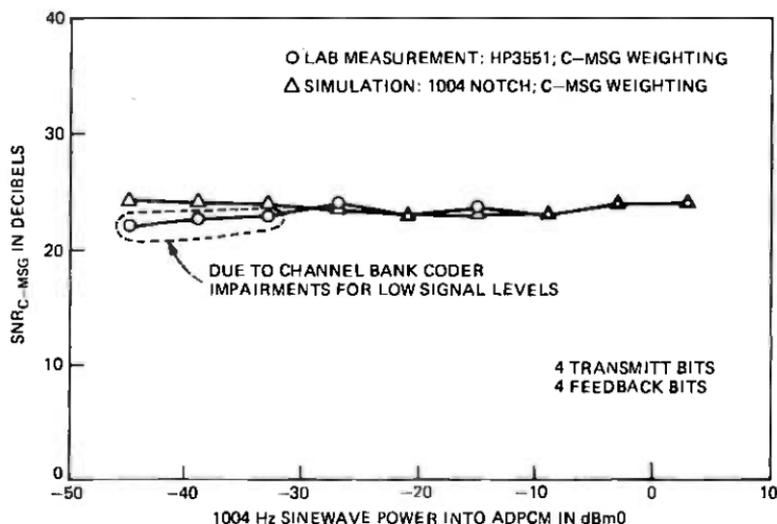


Fig. 11—Laboratory measurement vs simulation.

4 bits. Experimental and simulation results for 1004 Hz coding with 4-bit ADPCM are compared in Fig. 11.

V. ACKNOWLEDGMENT

We thank M. R. Aaron for his advice and encouragement.

REFERENCES

1. P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech," *B.S.T.J.*, 52, No. 7 (September 1973), pp. 1105-1118.

2. N. S. Jayant, "Adaptive Quantization With a One Word Memory," *B.S.T.J.*, 52, No. 7 (September 1973), pp. 1119-1144.
3. D. J. Goodman and A. Gersho, "Theory of an Adaptive Quantizer," *IEEE Trans. Comm.*, *COM-22*, No. 8 (August 1974), pp. 1037-1045.
4. P. Noll, "A Comparative Study of Various Quantization Schemes for Speech Encoding," *B.S.T.J.*, 54, No. 9 (November 1975), pp. 1597-1614.
5. B. McDermott, C. Scagliola, and D. Goodman, "Perceptual and Objective Evaluation of Speech Processed by Adaptive Differential PCM," *B.S.T.J.*, 57, No. 5 (May-June 1978), pp. 1597-1618.
6. D. Mitra and B. Gotz, "An Adaptive PCM System Designed for Noisy Channels and Digital Implementations," *B.S.T.J.*, 57, No. 7 (September 1978), pp. 2727-2763.
7. D. J. Goodman and C. Scagliola, "Channel Error Recovery in Digital Realizations of the Robust Adaptive Quantizer," to be published at *IEEE Trans. on Acoustics, Speech and Signal Processing*.
8. D. Mitra, "A Generalized Adaptive Quantization System With a New Reconstruction Method for Noisy Transmission," *International Conference on Acoustics, Speech and Signal Processing*, Washington, April 1979.



Transform Domain Motion Estimation

By J. A. STULLER and A. N. NETRAVALI

(Manuscript received March 21, 1979)

This paper introduces an algorithm for estimating the displacement of moving objects in a television scene from spatial transform coefficients of successive frames. The algorithm works recursively in such a way that the displacement estimates are updated from coefficient to coefficient. A promising application of this algorithm is in motion-compensated interframe hybrid transform-DPCM image coding. We give a statistical analysis of the transform domain displacement estimation algorithm and prove its convergence under certain realistic conditions. An analytical derivation is presented that gives sufficient conditions for the rate of convergence of the algorithm to be independent of the transform type. This result is supported by a number of simulation examples using Hadamard, Haar, and Slant transforms. We also describe an extension of the algorithm that adaptively updates displacement estimation according to the local features of the moving objects. Simulation results demonstrate that the adaptive displacement estimation algorithm has good convergence properties in estimating displacement even for very noisy images.

1. INTRODUCTION

The coefficient-recursive algorithm described in this paper estimates the displacement of objects in a television scene. It is a generalization of a pel-recursive displacement estimation algorithm recently introduced by Netravali and Robbins.^{1,2} Coefficient-recursive displacement estimation has potential application in hybrid transform-DPCM^{3,4} interframe image coders of the type discussed by Reader,⁵ Roese,⁶ and Jones.⁷ The performance of a hybrid transform-DPCM interframe coder using coefficient recursive motion compensation is described in a companion paper.⁸

Before defining the coefficient-recursive displacement estimation algorithm, it is useful to first describe pel-recursive displacement

estimation. Let $I(\mathbf{x}_k, t)$ denote the intensity of a scene at the k th sample point \mathbf{x}_k of a scan line, and let $I(\mathbf{x}_k, t - \tau)$ denote the intensity at the same spatial location in the previous frame. If the scene consists of an object that is undergoing pure translation, then, neglecting background,

$$I(\mathbf{x}_k, t) = I(\mathbf{x}_k - \mathbf{D}, t - \tau), \quad (1)$$

where \mathbf{D} is the displacement of the object in one frame interval τ . Pel-recursive displacement estimation attempts to estimate \mathbf{D} by minimizing the squared value of the displaced frame difference,

$$DFD(\mathbf{x}_k, \hat{\mathbf{D}}) = I(\mathbf{x}_k, t) - I(\mathbf{x}_k - \hat{\mathbf{D}}, t - \tau), \quad (2)$$

recursively with k using a steepest descent algorithm of the form:

$$\hat{\mathbf{D}}_{k+1} = \hat{\mathbf{D}}_k - \frac{1}{2} \epsilon \nabla_{\hat{\mathbf{D}}_k} [DFD(\mathbf{x}_k, \hat{\mathbf{D}}_k)]^2, \quad (3a)$$

where $\nabla_{\hat{\mathbf{D}}_k}$ is the two-dimensional gradient operator with respect to $\hat{\mathbf{D}}_k$. Carrying out this operation in (3a) and using (2) yields

$$\hat{\mathbf{D}}_{k+1} = \hat{\mathbf{D}}_k - \epsilon DFD(\mathbf{x}_k, \hat{\mathbf{D}}_k) \nabla I(\mathbf{x}_k - \hat{\mathbf{D}}_k, t - \tau), \quad (3b)$$

where $\nabla = \nabla_{\mathbf{x}}$ is the two-dimensional spatial gradient operator with respect to horizontal and vertical coordinates x_1 and x_2 in $\mathbf{x} = (x_1, x_2)^T$:

$$\nabla I(\mathbf{x}_k - \hat{\mathbf{D}}_k, t - \tau) = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \end{bmatrix} I(\mathbf{x}, t - \tau) \Big|_{\mathbf{x} = \mathbf{x}_k - \hat{\mathbf{D}}_k}. \quad (4)$$

Superscript T denotes transpose of a vector or matrix. The pel-domain interframe coder of Netravali and Robbins predicts intensity $I(\mathbf{x}_k, t)$ by the displaced previous frame intensity $I(\mathbf{x}_k - \hat{\mathbf{D}}_k, t - \tau)$ using interpolation for nonintegral $\hat{\mathbf{D}}_k$. If the magnitude of the prediction error exceeds a predetermined threshold, the coder transmits a quantized version of $DFD(\mathbf{x}_k, \hat{\mathbf{D}}_k)$ and address information to the receiver. Both receiver and transmitter then update $\hat{\mathbf{D}}_k$ according to (3b) using this quantized version. Netravali and Robbins² found that a coder using this algorithm consistently obtained bit rates that were 30 to 60 percent lower than those obtained by "frame-difference" prediction, which is commonly used in interframe coders.

In an interframe hybrid transform-DPCM coder,⁵⁻⁷ individual frames of video are partitioned into blocks having dimension N_r rows by N_c columns, and a two-dimensional transform is performed on each block to produce a set of coefficients. The transform coefficients of the q th block of the present frame are predicted by the corresponding coeffi-

coefficients of the q th block of the previous (reference) frame and, if the prediction error is sufficiently high, the quantized prediction errors are transmitted to the receiver. These quantized errors add as correction terms to the coefficients predicted by the decoder, which inverse-transforms the result to obtain the decoded image. This process repeats with both coder and decoder predicting the transform coefficients of the next frame by the coefficients of the decoded frame, as illustrated in Fig. 1. In this type of codec, data compression is achieved both by the redundancy reduction implicit in the prediction process and by the fact that some coefficients can be reproduced with low precision (or totally omitted) without visibly degrading the reconstructed image. An advantage of interframe hybrid transform-DPCM coding over conventional (3-dimensional block) interframe transform coding⁹ is the fact that the hybrid coder requires only a single frame of storage while the conventional transform coder requires many.

In a motion-compensated hybrid transform-DPCM coder of the type envisioned (Fig. 2), the n th coefficient of the q th present-frame block would be predicted by the n th coefficient of the *displaced* q th block of the previous frame where the displacement is a recursively updated estimate of frame-to-frame translation of the moving object. In this paper, we introduce and analyze a displacement estimation technique that operates recursively on coefficients in a manner analogous to the way (3) operates on pels.

Section II of this paper defines the coefficient-recursive displacement estimation algorithm for any real linear transform and gives illustrative simulation results using a separable 2-row by 8-column (2×8) transform block. A statistical analysis of the algorithm is given in Section III. In the analysis of Section III, a single frame is modeled as an image drawn at random from a stationary and ergodic ensemble of images. This random sample is assumed to be undergoing pure translation from frame to frame. An important result of this analysis is stated in Assertion 3 of Section 3.2, which says that, under certain conditions, the convergence properties of the coefficient-recursive displacement estimation are independent of the transform used. Section III presents simulation results that support this claim using Hadamard, Haar, and Slant transforms. Section IV describes an adaptive version of the coefficient-recursive algorithm and presents simulation results that indicate that this version can be used to some advantage in displacement estimation for noisy images. Illustrative simulation results are shown here using a 2×4 cosine transform block.

The algorithms discussed in this paper are local in nature and, as such, can estimate the individual frame-to-frame displacements of several objects that may be present in the television scene. However, we emphasize that all results presented here apply to objects undergoing pure translation; other types of motion are applicable to this study

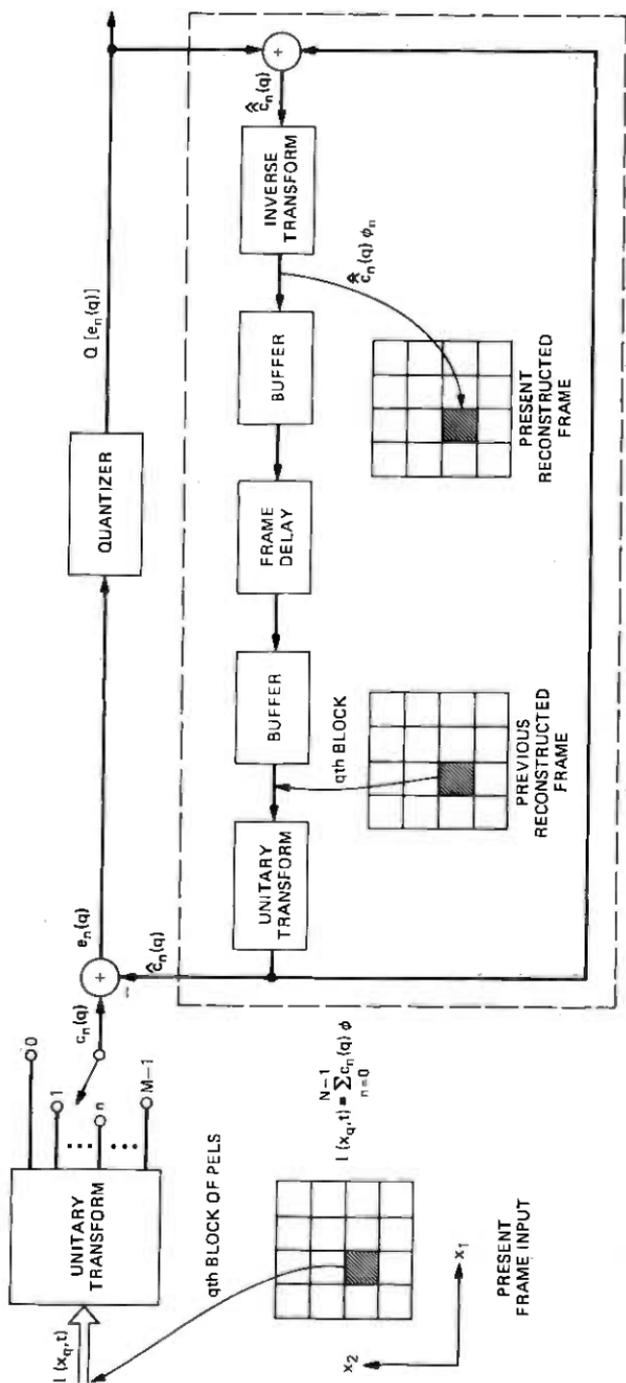


Fig. 1—Simplified diagram of hybrid transform-DPCM codec. The transform/inverse transform operations inside the dashed box are included for conceptual purposes.

to the extent that these can be approximated by pure translation over the spatial dimensions of a transform block. Background uncovered by moving objects is also ignored throughout this paper. In spite of the approximations involved, simulation results to be described in Ref. 8 show that the coefficient-recursive displacement algorithm studied here can be substantially beneficial when used in an interframe hybrid-DPCM codec operating on real-life scenes.

II. COEFFICIENT-RECURSIVE DISPLACEMENT ESTIMATION

Let a field of video be partitioned into rectangular blocks of pels, each having dimension N_r rows and N_c columns ($N_r \times N_c$). Let $\mathbf{x}_q = (x_{1q}, x_{2q})^T$ denote the coordinates of the upper left-hand pel of the q th block, where the blocks in each row of blocks are numbered from left to right with $q = 0, 1, 2, \dots$. We number the $N = N_r N_c$ pel intensities of block q in a column-scanning fashion and denote them by a column vector $I(\mathbf{x}_q, t)$. Let the N component vector ϕ_n be the n th basis vector of a nonzero but otherwise arbitrary real linear transform, and denote the n th coefficient of the q th block of this transform in the present frame by $c_n(q)$, where

$$c_n(q) = \mathbf{I}^T(\mathbf{x}_q, t)\phi_n \quad (5)$$

and n is numbered from 0 to $N - 1$. The displaced previous frame value of this coefficient is

$$\hat{c}_n(q, \hat{\mathbf{D}}) = \mathbf{I}^T(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)\phi_n, \quad (6)$$

where $\mathbf{I}(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)$ is the column vector of intensities of the displaced q th block of the previous frame and $\hat{\mathbf{D}}$ is the estimated displacement of the moving object. Computation of the elements in $\mathbf{I}(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)$ generally requires an interpolation among the given previous-frame pel intensities. Prediction of $c_n(q)$ of (5) by $\hat{c}_n(q, \hat{\mathbf{D}})$ of (6) results in coefficient prediction error

$$e_n(q, \hat{\mathbf{D}}) = [\mathbf{I}(\mathbf{x}_q, t) - \mathbf{I}(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)]^T \phi_n. \quad (7)$$

The algorithm defined in this section attempts to decrease the squared-prediction errors $e(q, \hat{\mathbf{D}})^2$ in a coefficient-recursive manner by steepest descent iteration of the form

$$\begin{aligned} \hat{\mathbf{D}}_{n+1}(q) &= \hat{\mathbf{D}}_n(q) - \frac{\epsilon}{2} \nabla_{\hat{\mathbf{D}}_n(q)} e_n^2(q, \hat{\mathbf{D}}_n(q)) \\ &= \hat{\mathbf{D}}_n(q) - \epsilon e_n(q, \hat{\mathbf{D}}_n(q)) \mathbf{G}_n(q) \end{aligned} \quad (8a)$$

for $n = 0, 1, \dots, M-2$, $M \leq N$, and $q = 0, 1, 2, \dots$, with

$$\begin{aligned} \hat{\mathbf{D}}_0(q) &= \hat{\mathbf{D}}_{M-1}(q-1) \\ &- \epsilon e_{M-1}(q-1, \hat{\mathbf{D}}_{M-1}(q-1)) \mathbf{G}_{M-1}(q-1). \end{aligned} \quad (8b)$$

In (8), $G_n(q)$ is the *coefficient gradient vector*

$$G_n(q) = \nabla \mathbf{I}^T(\mathbf{x}_q - \hat{D}_n(q), t - \tau)\phi_n. \quad (9)$$

Note that (8) operates upon coefficients 0 through $M - 1$, where for generality we assume $M \leq N$. Iteration in (8) progresses as follows. The initial displacement estimate of the q th block ($q > 1$), $\hat{D}_0(q)$ is formed by updating the final displacement estimate $\hat{D}_{M-1}(q - 1)$ of the previous block as in (8b). The next displacement estimate of the q th block $\hat{D}_1(q)$ is formed from (8a) with $n = 0$. Iteration progresses in the q th block by (8a) with $n = 1, 2, \dots, M - 2$, resulting finally in displacement estimate $\hat{D}_{M-1}(q)$ which, when updated in (8b) (with $q \rightarrow q + 1$), forms the initial displacement estimate $\hat{D}_0(q + 1)$ of block $q + 1$. This iteration procedure continues along all horizontal blocks of raster. The procedure is started in the $q = 0$ block with an arbitrarily chosen initial displacement estimate $\hat{D}_0(0)$ followed by iterations of (8a) for $n = 0, 1, \dots, M - 2$ and $q = 0$. In the sequel we assume that $\hat{D}_0(0)$ is zero.

The envisioned motion-compensated interframe hybrid transform-DPCM coder transmits a quantized version of coefficient prediction error $e_n(q, \hat{D}(q))$ to the receiver whenever the magnitude $|e_n(q, \hat{D}(q))|$ exceeds a threshold, thereby enabling the decoder to update its displacement estimate $\hat{D}_n(q)$ as in (8) as well as correcting its prediction $\hat{c}_n(q, \hat{D}_n(q))$ of coefficient $c_n(q)$. Both encoder and decoder use the updated displacement estimate in predicting the next coefficient, and the process continues. A simplified block diagram of the system that omits the thresholding operations is given in Fig. 2.

In the sequel, it is convenient to rewrite (8) in a form that explicitly describes the iteration convention. This can be done by defining a single index i ,

$$i = qM + n; \quad i = 0, 1, 2, \dots \quad (10a)$$

that equals the total number of iterations of (8) that have occurred in iterating from $\hat{D}_0(0)$ to $\hat{D}_n(q)$. Quantities q and n are related to i by

$$n = ((i)) \quad (10b)$$

$$q = [[i]], \quad (10c)$$

where we use the notation $((i))$ to denote i modulo M and $[[i]]$ to denote the integer part of i/M .

Using (10), we set $\hat{D}_i \triangleq \hat{D}_n(q)$ and rewrite (8) as

$$\hat{D}_{i+1} = \hat{D}_i - \epsilon e_{((i))}([[i]]), \hat{D}_i) G_{((i))}([[i]]) \quad (11)$$

with $i = 0, 1, 2, \dots$. Note that the Netravali-Robbins pel-recursive

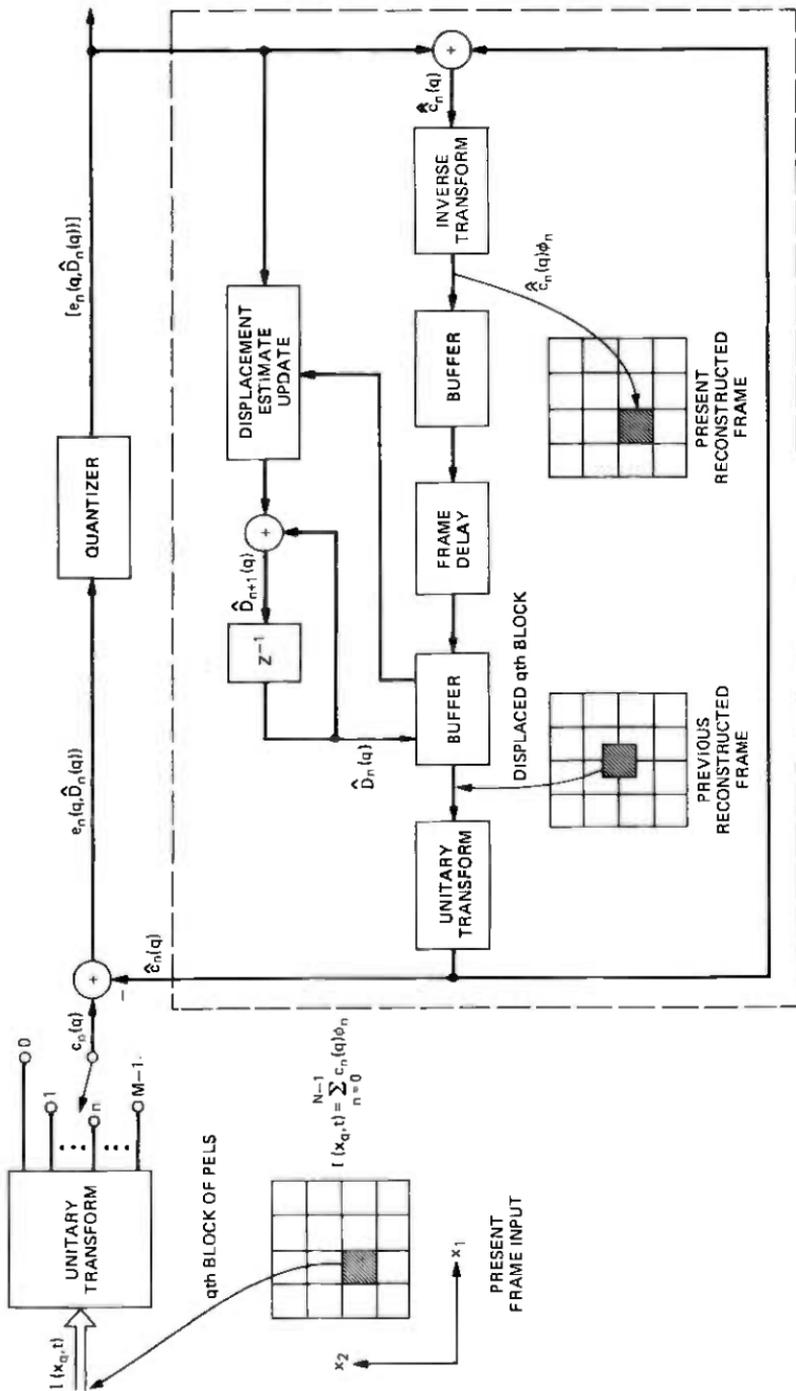


Fig. 2—Simplified diagram of motion-compensated hybrid transform-DPCM codec.

displacement estimation algorithm (3b) is a special case of (10) that results for a transform "block" having dimension 1 pel by 1 pel and single "basis vector" $\phi_0 = 1$.

We emphasize that recursions (8) and (11) were derived with the objective of decreasing the squared coefficient prediction errors of a hybrid transform-DPCM codec. As shown in Section III, coefficient prediction error is related to displacement estimation error (approximately) by a dot product between the displacement estimation error vector and vector $G_n(q)$ that describes the spatial rates of change of the coefficient estimate $\hat{c}_n(q, \hat{D}_n(q))$ with respect to small displacements of the block. Therefore, only the component of displacement estimation error in the direction of $G_n(q)$ contributes to coefficient prediction error, and it is this component that is relevant in evaluating the performance of (8) or (11). For this reason, experimental results given in this paper refer to the component of displacement estimation error measured in the direction of its corresponding coefficient gradient $G_n(q)$.

Experimental illustrations of the behavior of (11) are given in Figs. 3 through 5 where the moving object was the synthetically generated pattern of Fig. 6, displaced 2 pels in the horizontal direction each frame interval τ . This is a radial cosine function having a radius of 60 pels, and peak-to-peak amplitude 220 (out of an intensity range 0 to 255) at its center, decreasing to 130 at the circumference. The period P decreases with radial distance R starting with a period of 20 pels at center to 10 pels at the circumference. The pattern is described mathematically by the intensity function

$$f(R) = 100 \exp(-0.01 R) \cos(2\pi R/P) + 128; \quad 0 \leq R \leq 60, \quad (12a)$$

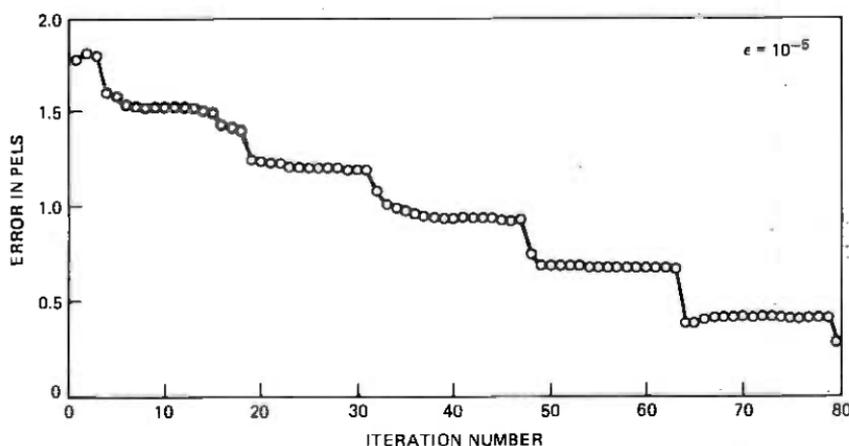


Fig. 3—Single-line convergence results using 2×8 separable Hadamard transform with $\epsilon = 10^{-5}$.

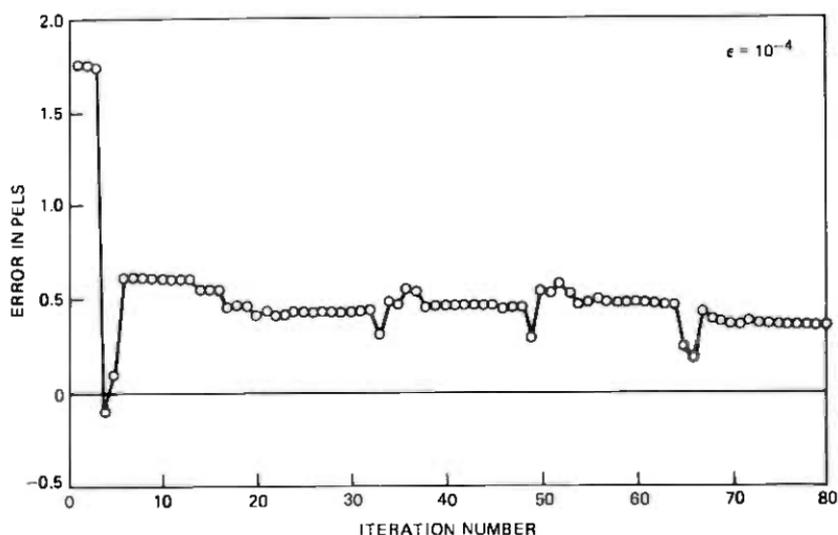


Fig. 4—Single-line convergence results using 2×8 separable Hadamard transform with $\epsilon = 10^{-4}$.

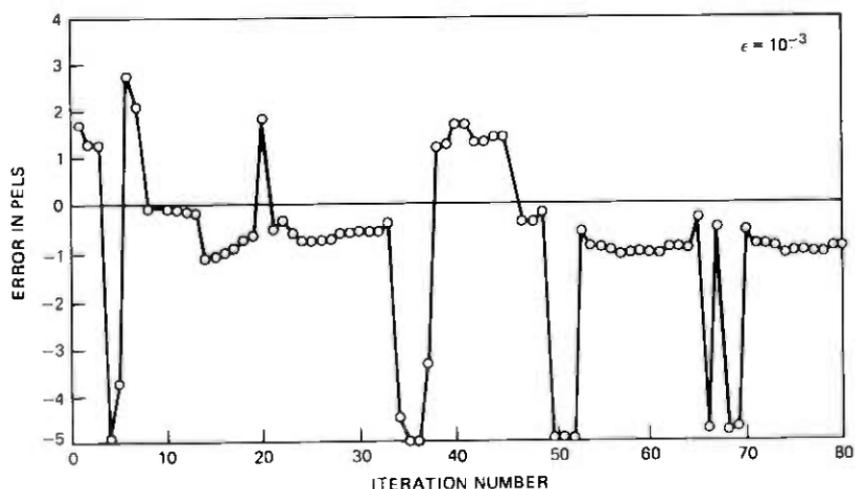


Fig. 5—Single-line convergence results using 2×8 separable Hadamard transform with $\epsilon = 10^{-3}$.

where

$$P = (1 - R/60.) 10 + 10. \quad (12b)$$

This function is displayed on a 256×256 element raster in two interlaced fields of 128 lines each. In applying (11), the spatial transforms were taken over a single field with coefficient prediction performed from the corresponding field separated in time by a frame

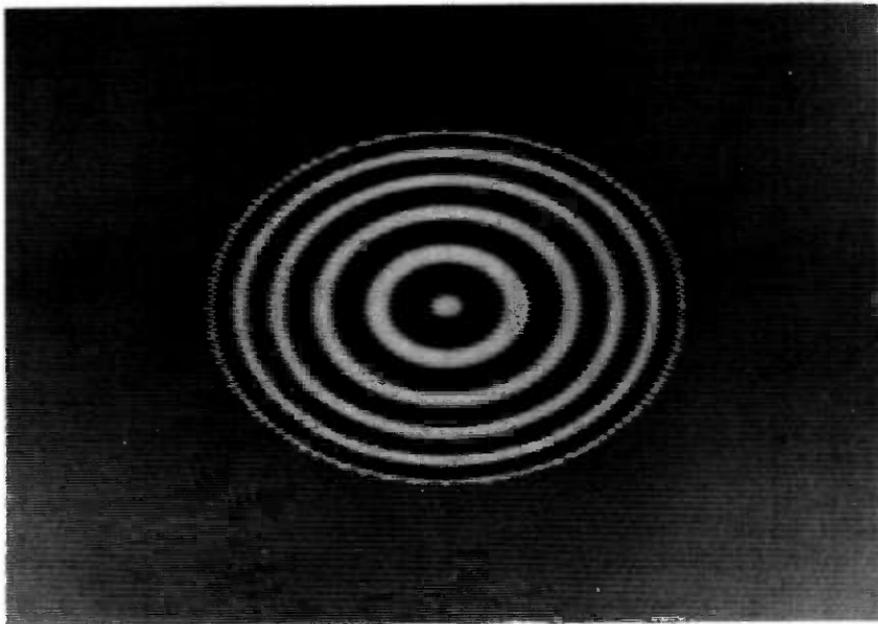


Fig. 6—Synthetic image used in simulations. This image is described by eq. (11) of Section II.

interval τ . Figures 3 through 5 show displacement estimation error in the direction of the spatial gradient of the corresponding coefficient versus iteration for a sequence of 2×8 blocks located 10 field lines above the center of the figure. Iteration was initiated with a zero displacement estimate approximately 7 pels within the pattern for each horizontal sequence. In these examples and all others presented in this paper, the two-dimensional transforms concerned were separable transforms of the form¹⁰

$$C = V[I]H, \quad (13)$$

where C is the $N_r \times N_c$ coefficient matrix, $[I]$ is the $N_r \times N_c$ matrix of pels, and V and H are unitary matrices having dimensions $N_r \times N_r$ and $N_c \times N_c$, respectively. Coefficient $c_n(q)$ of (5) is the n th column scanned coefficient of matrix C , with $[I]$ taken to be the q th pixel block of the present frame. In Figs. 3 through 5, V and H were the normalized sequency-ordered 2×2 and 8×8 Hadamard matrices of Fig. 7, and iteration of (11) progressed through all $M = 16$ coefficients in a block (i.e., $M = N$). Figure 3 illustrates the behavior of (11) for $\epsilon = 10^{-5}$. It can be seen by inspection of this figure that displacement estimation error tends to decrease roughly in a series of steps of 16 iterations. Iterations 1 to 6, 17 to 22, etc., corresponding to the first few (low sequency) vectors in $\{\phi_n\}$ tend to affect error significantly, while the

iterations corresponding to the higher sequency basis vectors do not. This type of behavior is scene-dependent and is investigated in the analysis of Section III. Figures 4 and 5 illustrate convergence for increased values of ϵ . In general, convergence rate increases gradually with increasing ϵ up to a point after which oscillations occur. Convergence to within 0.5-pel error was achieved for this particular example in two or three iterations for $\epsilon = 10^{-4}$ while the recursion became oscillatory (and eventually unstable as in Fig. 5) at values exceeding this. All these results are scene-dependent and, to some extent, dependent upon the row position of the sequence of blocks. Because of this, the behavior of (11) will be examined from a statistical viewpoint in Section III.

III. PROPERTIES OF COEFFICIENT-RECURSIVE DISPLACEMENT ESTIMATION

The displacement estimation procedure defined by (8) is a nonlinear recursion relation whose dynamic behavior is complicated by the fact that the error $e_n(q, \hat{D}_n(q))^2$ is generally a multimodal function of $\hat{D}_n(q)$, having global minimum at $\hat{D}_n(q) = \mathbf{D}$ and local minima elsewhere. Convergent solutions to eq. (8) can, therefore, exist at displacement estimates other than the true displacement \mathbf{D} . Subsequent analysis will restrict consideration to the case in which the displacement estimate is sufficiently close to the true displacement that $e^2(q, \hat{D})$ can be approximated as a quadratic function of $\hat{D} - \mathbf{D}$. Under this restriction, the anomalous solutions can be ignored, and (8) reduces to an approximately linear stochastic recurrence relation.

Section 3.1 derives the linearized approximation to (8) and the associated quadratic error expression. The dynamic behavior of the coefficient-recursive displacement estimator is analyzed in Section 3.2. An important result of this section is that, for ϵ sufficiently small, the block-to-block convergence rate of mean displacement estimation error resulting from (8) is independent of the transform used for any unitary transform.

$$\begin{array}{c}
 \left(\begin{array}{cc} 0.707 & 0.707 \\ 0.707 & -0.707 \end{array} \right) \\
 \text{(a)}
 \end{array}
 \qquad
 \begin{array}{c}
 \left(\begin{array}{cccccccc} 0.353 & 0.353 & 0.353 & 0.353 & 0.353 & 0.353 & 0.353 & 0.353 \\ 0.353 & 0.353 & 0.353 & 0.353 & -0.353 & -0.353 & -0.353 & -0.353 \\ 0.353 & 0.353 & -0.353 & -0.353 & -0.353 & -0.353 & 0.353 & 0.353 \\ 0.353 & 0.353 & -0.353 & -0.353 & 0.353 & 0.353 & -0.353 & -0.353 \\ 0.353 & -0.353 & -0.353 & 0.353 & 0.353 & -0.353 & -0.353 & 0.353 \\ 0.353 & -0.353 & -0.353 & 0.353 & -0.353 & 0.353 & 0.353 & -0.353 \\ 0.353 & -0.353 & 0.353 & -0.353 & -0.353 & 0.353 & -0.353 & 0.353 \end{array} \right) \\
 \text{(b)}
 \end{array}$$

Fig. 7—Sequency-ordered Hadamard matrices. (a) 2×2 . (b) 8×8 .

3.1 Linear analysis

Assume that the pel intensities are samples of an object that is undergoing pure translation \mathbf{D} from frame to frame as in (1) so that, neglecting background,

$$\mathbf{I}(\mathbf{x}_q, t) = \mathbf{I}(\mathbf{x}_q - \mathbf{D}, t - \tau). \quad (14)$$

For Euclidean norm $\|\hat{\mathbf{D}}_n(q) - \mathbf{D}\|$ sufficiently small, eq. (7) becomes, by Taylor's expansion about $\mathbf{D} - \hat{\mathbf{D}}(q)$, to a linear approximation

$$\begin{aligned} e_n(q, \hat{\mathbf{D}}_n(q)) &= [\mathbf{I}(\mathbf{x}_q - \mathbf{D}, t - \tau) - \mathbf{I}(\mathbf{x}_q - \hat{\mathbf{D}}_n(q), t - \tau)]^T \phi_n \\ &\approx \mathbf{G}_n^T(q) \Delta_n(q), \end{aligned} \quad (15)$$

where $\mathbf{G}_n(q)$ is given by (9) and $\Delta_n(q)$ is the displacement estimation error $\Delta_n(q) = \hat{\mathbf{D}}_n(q) - \mathbf{D}$. Using the approximation in (15), we can approximate the squared coefficient estimate error by

$$e_n^2(q, \hat{\mathbf{D}}_n(q)) \approx \Delta_n^T(q) [\mathbf{G}_n(q) \mathbf{G}_n^T(q)] \Delta_n(q), \quad (16)$$

which is a quadratic function of the horizontal and vertical components of $\Delta_n(q)$.

In terms of approximation (15), (8) assumes the form

$$\Delta_{n+1}(q) = [U - \epsilon \mathbf{G}_n(q) \mathbf{G}_n^T(q)] \Delta_n(q) \quad (17a)$$

with

$$\Delta_0(q) = [U - \epsilon \mathbf{G}_{M-1}(q-1) \mathbf{G}_{M-1}^T(q-1)] \Delta_{M-1}(q-1), \quad (17b)$$

where U is the 2×2 identity matrix. Similarly, (11) becomes

$$\Delta_{i+1} = [U - \epsilon \mathbf{G}_{(i)} \mathbf{G}_{(i)}^T] \Delta_i, \quad i = 0, 1, \dots, \quad (18)$$

where $\Delta_i \triangleq \Delta_n(q)$.

Considering the image as a random process, (18) is a stochastic recurrence relation. Equations similar to, but somewhat simpler than, (18) have appeared in the problem of adaptive tap gain adjustment of automatic channel equalizers. Unfortunately, a complete statistical description of the behavior of these simpler equations has not yet been obtained. The difficulty in analyzing these equations is that their solution depends upon products of matrices that are statistically dependent. It has been found, however, that useful approximate results can be obtained by treating the dependent matrices as if they were actually independent.¹¹ We use this method in Section 3.2 to analyze (18). As shown in Section 3.2 and Appendix A, there is some analytical justification for this approach because of certain properties of the transforms conventionally used in image coding. Further justification is given in the asymptotic analysis of Appendix B.

3.2 Statistical analysis

This section studies the behavior of mean displacement estimation error under the assumption that the sequence of gradient vectors $\overline{\mathbf{G}_{(i)}}(\{[i]\})$ entering (18) are statistically independent ("independence assumption"). Note that, if the sequence of $\overline{\mathbf{G}_{(i)}}(\{[i]\})$ in (18) are independent, then Δ_i will be independent of the matrix premultiplying it. This follows from the fact that Δ_i is determined by the $\overline{\mathbf{G}_{(j)}}(\{[j]\})$, $j < i$, which, by assumption, are independent of $\overline{\mathbf{G}_{(i)}}(\{[i]\})$. The ensemble mean of (18) (denoted by the overhead bar) then becomes

$$\overline{\Delta}_{i+1} = [U - \overline{\epsilon \mathbf{G}_{(i)} \mathbf{G}_{(i)}^T}] \overline{\Delta}_i, \quad i = 0, 1, 2, \dots \quad (19)$$

In writing (19), we have used (9) and assumed a stationary ergodic image ensemble. The matrix $\overline{\mathbf{G}_n(q) \mathbf{G}_n^T(q)}$ in this case will not depend upon $\mathbf{x}_q - \mathbf{D}_n(q)$ and is written simply as $\overline{\mathbf{G}_{(i)} \mathbf{G}_{(i)}^T}$. The matrix $\overline{\mathbf{G}_{(i)} \mathbf{G}_{(i)}^T}$ is periodic in i with period M , having values specified by $\overline{\mathbf{G}_n \mathbf{G}_n^T}$, $n = 0, 1, \dots, M-1$. Alternative expressions for $\overline{\mathbf{G}_n \mathbf{G}_n^T}$ are

$$\overline{\mathbf{G}_n \mathbf{G}_n^T} = \overline{\nabla \mathbf{I}^T(\mathbf{x}) \phi_n \phi_n^T \nabla^T \mathbf{I}(\mathbf{x})} \quad (20a)$$

and

$$\overline{\mathbf{G}_n \mathbf{G}_n^T} = \begin{bmatrix} \phi_n^T R_1 \phi_n & \phi_n^T R_{12} \phi_n \\ \phi_n^T R_{12} \phi_n & \phi_n^T R_2 \phi_n \end{bmatrix}, \quad (20b)$$

where R_1 , R_2 and R_{12} are auto- and cross-correlation matrices of $\mathbf{I}/\partial x_1$ and $\mathbf{I}/\partial x_2$, and

$$\nabla^T = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2} \right].$$

Equation (19) can be interpreted in terms of an optimum (Wiener) displacement estimator. Consider the mean square n th coefficient prediction error resulting from a given displacement error $\Delta_n(q) = \delta$. To the linear approximation of (15), this is

$$F_n(\delta) = \delta^T \overline{\mathbf{G}_n \mathbf{G}_n^T} \delta, \quad (21)$$

which is a quadratic function of the components of δ . A steepest descent algorithm for arriving at the minimum of the $F_n(\delta)$ for $n = 0, 1, \dots, M-1$ is

$$\delta_{i+1} = [U - \overline{\epsilon \mathbf{G}_{(i)} \mathbf{G}_{(i)}^T}] \delta_i, \quad i = 0, 1, 2, \dots \quad (22)$$

Comparing (19) and (22), we see that, under the assumption of mutually independent $\overline{\mathbf{G}_{(i)}}(\{[i]\})$, the mean displacement estimation error $\overline{\Delta}_i$ satisfies a recursion (22) that minimizes mean-square prediction

error of coefficients $n = 0, 1, \dots, M - 1$. The convergence of (19) and (22) is established below in Assertion 1.

We emphasize that recursion (19) describes the progression of mean displacement estimation error under the assumption of independent $\mathbf{G}_{(i)}([[i]])$. The conventional approach to transform image coding is to choose basis vectors $\{\phi_n\}$ so that the transmitted coefficients (or the prediction errors of these coefficients) are as mutually "independent as possible."¹² This is best achieved by choosing the basis vectors $\{\phi_n\}$ of the transform to be the eigenvectors of the covariance matrix (i.e., the Karhunen-Loeve basis) of the block of pels. This basis results in transform coefficients that are *linearly* independent (uncorrelated) within the transform block, although dependency among coefficients from block to block may persist. Other transforms, such as the cosine or Hadamard transforms can be viewed as practical approximations to the Karhunen-Loeve transform. Assuming that the Karhunen-Loeve-basis vectors result in coefficient *gradients* that are linearly independent as well, this would help justify the application of independence theory in describing the behavior of (18). In Appendix A we show that this assumption is indeed correct for the stochastic image model most widely applied in image processing analyses. Some insight into the behavior of (18) for dependent $\mathbf{G}_{(i)}([[i]])$ is given by the analysis in Appendix B.

Assertion 1: Under the independence assumption, mean displacement estimation error in (19) converges to zero if and only if the eigenvalues of Ψ are inside the unit circle, where

$$\Psi = \prod_{n=0}^{M-1} [U - \overline{\epsilon \mathbf{G}_n \mathbf{G}_n^T}]. \quad (23)$$

In our product notation (23), matrix $U - \overline{\epsilon \mathbf{G}_0 \mathbf{G}_0^T}$ is premultiplied by $U - \overline{\epsilon \mathbf{G}_1 \mathbf{G}_1^T}$, etc.

Proof: This can easily be shown by iterating (19) from $i = 0$ to $i = qM + n - 1$:

$$\begin{aligned} \bar{\Delta}_{qM+n} &= \left[\prod_{i=0}^{qM+n-1} [U - \overline{\epsilon \mathbf{G}_{(i)} \mathbf{G}_{(i)}^T}] \right] \bar{\Delta}_0 \\ &= [U - \overline{\epsilon \mathbf{G}_{n-1} \mathbf{G}_{n-1}^T}] \cdots [U - \overline{\epsilon \mathbf{G}_0 \mathbf{G}_0^T}] \Psi^q \bar{\Delta}_0. \end{aligned} \quad (24)$$

Therefore, the behavior of $\bar{\Delta}_{qM+n}$ as q increases depends upon the matrix Ψ as Ψ^q , and Assertion 1 follows. QED

A useful sufficient condition for convergence of (19) is given in the following.

Assertion 2: Under the independence assumption and for any normalized set of basis vectors, mean displacement estimation error in (19) is bounded by

$$\|\bar{\Delta}_{qM+n}\| \leq \kappa^q \|\bar{\Delta}_0\| \quad (25)$$

where κ , ($0 < \kappa < 1$) is the maximum eigenvalue of matrices $U - \epsilon \mathbf{G}_n \mathbf{G}_n^T$, $n = 0, 1, \dots, M-1$, and

$$0 < \epsilon < \frac{2}{N[(\partial I / \partial x_1)^2 + (\partial I / \partial x_2)^2]} \quad (26)$$

Proof: Using the Schwartz inequality on (24):

$$\|\bar{\Delta}_{qM+n}\| \leq \|U - \epsilon \overline{\mathbf{G}_{n-1} \mathbf{G}_{n-1}^T}\| \cdots \|U - \epsilon \overline{\mathbf{G}_0 \mathbf{G}_0^T}\| \|\Psi\|^q \|\Delta_0\|, \quad (27)$$

where $\|(\cdot)\|$ for a symmetric matrix is the magnitude of the maximum magnitude eigenvalue. Similarly,

$$\|\Psi\| \leq \prod_{n=0}^{M-1} \|U - \epsilon \overline{\mathbf{G}_n \mathbf{G}_n^T}\|. \quad (28)$$

The eigenvalues of $U - \epsilon \overline{\mathbf{G}_n \mathbf{G}_n^T}$ have the form $1 - \epsilon \lambda_n^{(i)}$ where the $\lambda_n^{(i)}$, $i = 1, 2$, are the (nonnegative) eigenvalues of $\overline{\mathbf{G}_n \mathbf{G}_n^T}$. Let κ be the largest of the norms in (28) and let λ_{\max} be the largest of the $\lambda_n^{(i)}$ for $0 \leq n \leq M-1$. Then excluding trivial cases, each matrix norm in (27) and (28) will be less than unity for $0 < \epsilon < 2/\lambda_{\max}$, and at least q of the norms in (27) is κ . Therefore,

$$\|\bar{\Delta}_{qM+n}\| \leq \kappa^q \|\bar{\Delta}_0\| \quad (29)$$

for $0 < \epsilon < 2/\lambda_{\max}$.

Assume that λ_{\max} corresponds to $\overline{\mathbf{G}_n \mathbf{G}_n^T}$ for $n = k$. Then using (20) and the fact that the eigenvalues of a nonnegative-definite matrix are bounded by the trace of the matrix gives the chain of inequalities

$$\begin{aligned} \lambda_{\max} &\leq \text{Tr}[\overline{\mathbf{G}_k \mathbf{G}_k^T}] \\ &\leq \phi_k^T \mathbf{R}_1 \phi_k + \phi_k^T \mathbf{R}_2 \phi_k \\ &\leq N[(\partial I / \partial x_1)^2 + (\partial I / \partial x_2)^2]. \end{aligned} \quad (30)$$

Therefore ϵ in the range (26) guarantees $0 < \epsilon < 2/\lambda_{\max}$ which, in turn, guarantees (29). QED

Allowing for variations in scene statistics, a conservative choice of ϵ would be somewhat less than $2/\lambda_{\max}$. In this event, the following assertion applies.

Assertion 3: If iteration is taken over all coefficients (i.e., $M = N$) of any complete orthonormal basis $\{\phi_n\}$, and if ϵ is small compared to $2/\lambda_{\max}$ where λ_{\max} is the maximum eigenvalue of matrices $\{\overline{\mathbf{G}_n \mathbf{G}_n^T}; n = 0, \dots, N-1\}$, then the block-to-block convergence of $\bar{\Delta}$, is independent of the particular basis set used. Furthermore, the convergence rate is independent of block dimensions N_r and N_c .

Proof: The block-to-block dynamics of $\bar{\Delta}_i$ is determined by the matrix Ψ of (23). For $0 < \epsilon \ll 2/\lambda_{\max}$, and for $M = N$, Ψ can be approximated by

$$\Psi = U - \epsilon \sum_{n=0}^{N-1} \overline{\mathbf{G}_n \mathbf{G}_n^T}. \quad (31)$$

From (20a),

$$\sum_{n=0}^{N-1} \overline{\mathbf{G}_n \mathbf{G}_n^T} = \overline{\nabla \mathbf{I}^T(\mathbf{x}) \left[\sum_{n=0}^{N-1} \phi_n \phi_n^T \right] \nabla^T \mathbf{I}(\mathbf{x})}. \quad (32)$$

But since $\{\phi_n\}$ is a complete orthonormal set, $\sum_{n=0}^{N-1} \phi_n \phi_n^T = U$, the N -by- N identity matrix, and

$$\Psi = U - \epsilon \overline{\nabla \mathbf{I}^T(\mathbf{x}) \nabla^T \mathbf{I}(\mathbf{x})}, \quad (33)$$

which is independent of $\{\phi_n\}$.

The expectation in (33) is given by

$$\overline{\nabla \mathbf{I}^T(\mathbf{x}) \nabla^T \mathbf{I}(\mathbf{x})} = \left\{ \begin{array}{cc} \sum \frac{(\partial I / \partial x_1)^2}{(\partial I / \partial x_1)(\partial I / \partial x_2)} & \sum \frac{(\partial I / \partial x_1)(\partial I / \partial x_2)}{\sum (\partial I / \partial x_2)^2} \end{array} \right\}, \quad (34)$$

where the summations are over the N pels of a block. By stationarity, the expectation of each term in (34) is a constant and (33) becomes

$$\begin{aligned} \Psi &= U - \epsilon N \Gamma \\ &\approx (U - \epsilon \Gamma)^N, \end{aligned} \quad (35)$$

where

$$\Gamma = \begin{bmatrix} \frac{(\partial I / \partial x_1)^2}{(\partial I / \partial x_1)(\partial I / \partial x_2)} & \frac{(\partial I / \partial x_1)(\partial I / \partial x_2)}{(\partial I / \partial x_2)^2} \end{bmatrix}. \quad (36)$$

Therefore q block-to-block iterations of (19) premultiplies Δ_0 by $(U - \epsilon \Gamma)^{qN}$, which is a function only of the total number of iterations qN and is independent of basis and block dimension. QED

Experimental evidence of Assertion 3 is shown in Figs. 8a to 8d. Figure 8a shows the relevant component of displacement estimation error versus iteration number averaged over the interior of the moving cosine pattern of Fig. 6 for a pel-recursive (unity) 1×1 transform and a 1×8 Hadamard transform of the type in Fig. 7b using $\epsilon = 5 \times 10^{-5}$. For each scan line entering the average, iteration of (11) was initiated with displacement estimate $\hat{\mathbf{D}} = 0$ just inside the circumference of the pattern. In spite of the disparity between block size and transform type, the block-to-block convergence rate (measured over spans of eight iterations) of the Hadamard estimator closely matches that of the pel-recursive estimator. Although Assertion 2 concerns average displacement error, we found that it often applied as well to individual scan lines as shown in Figs. 8b to 8d. These figures show the relevant

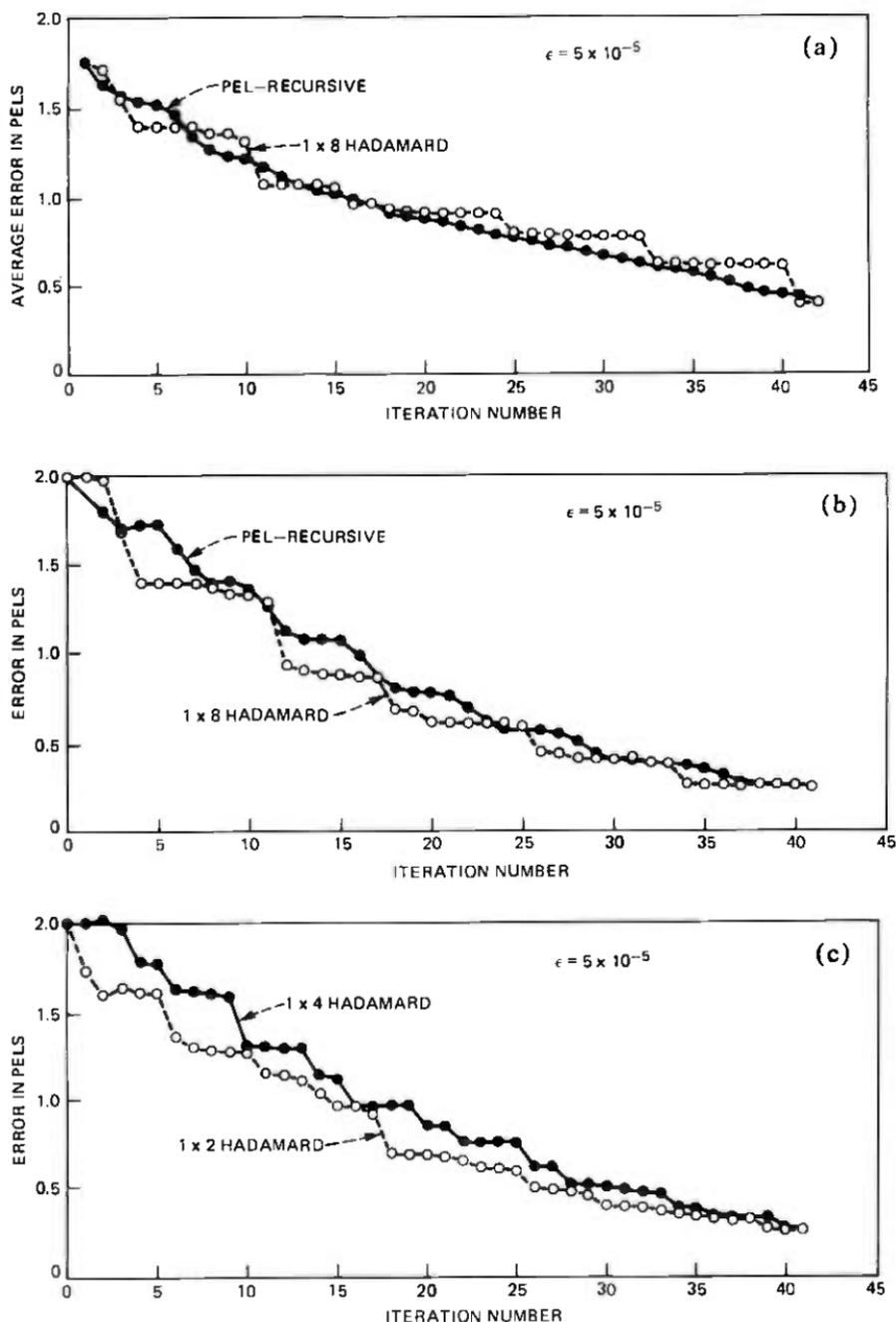


Fig. 8—Convergence results for various transforms with $\epsilon = 5 \times 10^{-5}$. (a) Pel-recursive and Hadamard 1×8 transform average relevant displacement estimation errors vs iteration number. (b) Pel-recursive and Hadamard 1×8 transform relevant displacement estimation errors for scan line through middle of moving cosine pattern. (c) Hadamard 1×4 and 1×2 transform relevant displacement estimation errors for scan line through middle of moving cosine pattern (*continued*).

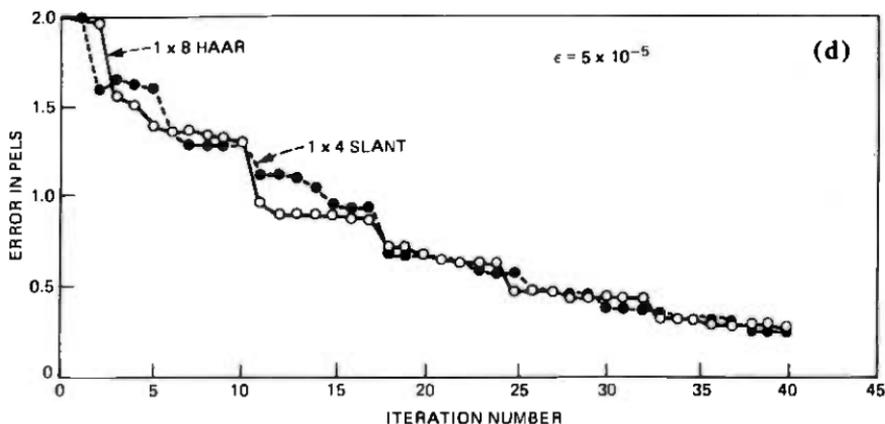


Fig. 8—(continued) (d) Haar 1×8 and Slant 1×4 transform relevant displacement estimation errors for scan line through middle of moving cosine pattern.

component of displacement estimation error versus iteration number for the scan line running through the middle of the moving cosine pattern. Hadamard 1×2 , 1×4 , 1×8 , Slant 1×4 , Haar 1×8 , and pel-recursive estimators are all seen to have similar convergence rates with $\epsilon = 5 \times 10^{-5}$ when measured over the appropriate span of iterations. This also applies to the cosine transform (not shown), which was found to behave similarly to the Hadamard transform.

Figures 9a to 9c compare convergence of the 1×8 Hadamard block and pel-recursive displacement estimators as ϵ increases. The image data in this case was also the middle scan line of the moving cosine pattern. It can be seen that the convergence rates of the Hadamard and pel-recursive estimators are in rough agreement for increasing ϵ up to the point where oscillations occur (Fig. 9c). Note that oscillations occur in the Hadamard estimator before occurring in the pel-recursive estimator. This behavior was found to be the case for other transform types as well.

Although the block-to-block convergence rate of transform domain displacement estimators is substantially independent of the transform type, this is clearly not the case for *within* block convergence rate, as evidenced by Figs. 8 and 9. An explanation of this is given from the form of Ψ in (23), in which particular basis vector ϕ_n contributes a matrix factor of the form $[U - \epsilon \mathbf{G}_n \mathbf{G}_n^T]$. This contribution of ϕ_n to reducing average displacement estimation error depends upon the eigenvalues of $\mathbf{G}_n \mathbf{G}_n^T$ which are a measure of the statistical "match" between ϕ_n and the spatial rates of change of the scene.

It is possible to vary ϵ with n ($\epsilon = \epsilon_n$) to partially compensate for differences among the eigenvalues of matrices $\mathbf{G}_n \mathbf{G}_n^T$. However, this technique can also make the algorithm move sensitive to noise that

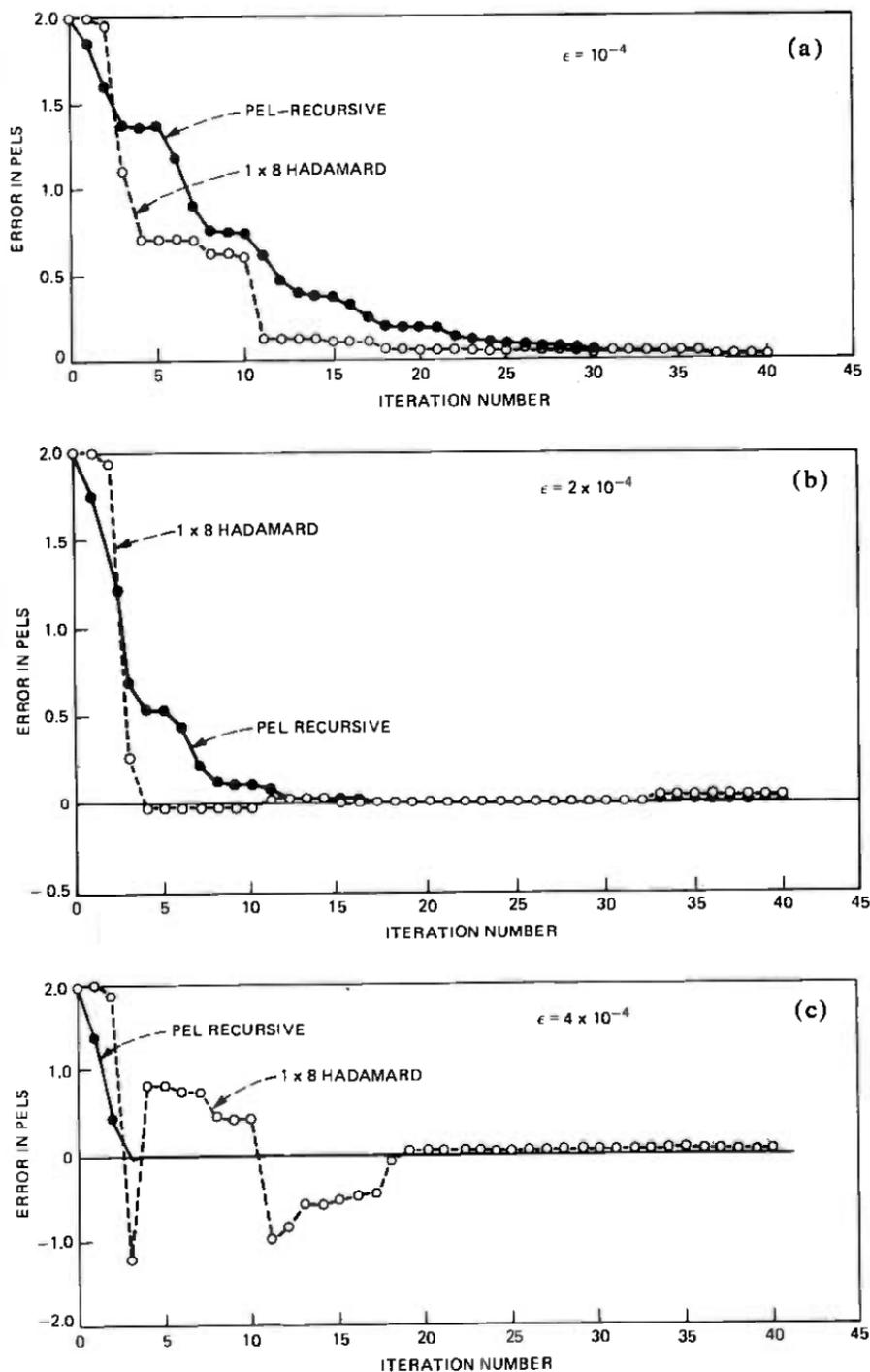


Fig. 9—Convergence results for pel-recursive and 1×8 Hadamard displacement estimators. (a) $\epsilon = 1 \times 10^{-4}$. (b) $\epsilon = 2 \times 10^{-4}$. (c) $\epsilon = 4 \times 10^{-4}$.

may be present in the image data. Section IV gives another approach that appears to have particularly good noise rejection properties.

IV. ADAPTATION

This section shows how the coefficient displacement estimation algorithm of Sections II and III can be improved by adaptively updating the displacement estimate according to the local features of the image. This is a technique that is not possible for the pel-recursive estimation. Adaptation in displacement estimation is motivated by the recognition that single frames of video are neither noiseless nor best described as stationary processes. Simulation results using noise-corrupted versions of the radial cosine object of Fig. 6 demonstrate that an adaptive algorithm of the type described here can have better convergence properties than either pel-recursive or nonadaptive coefficient recursive displacement estimation.

4.1 Preliminaries

The potential advantage of adaptation in (8) can be seen by considering the simple example of the moving edge scene of Fig. 10. This edge has constant slope $g = 3.8$ intensity increments per pel-to-pel distance over a width of 50 horizontal pel intervals, and velocity 2.7

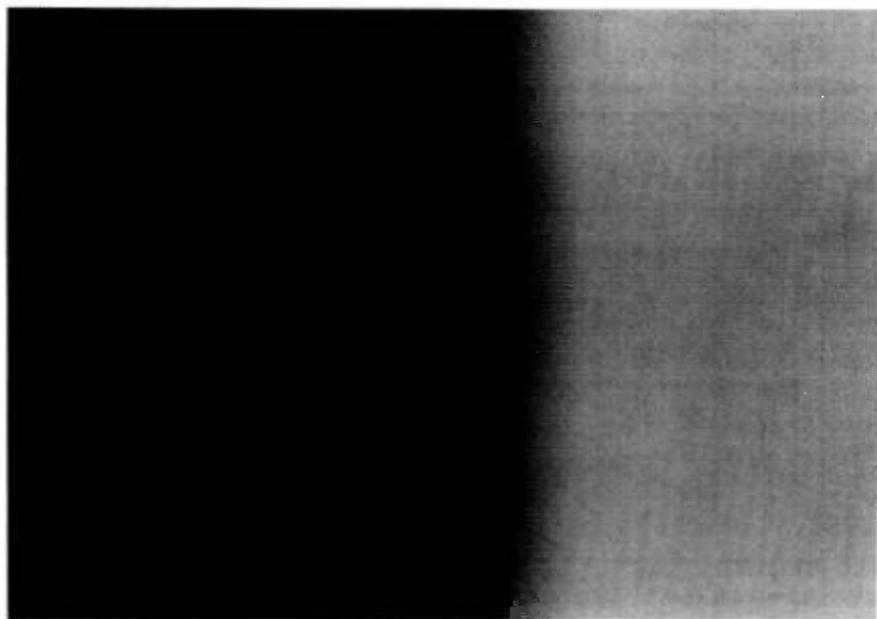


Fig. 10—Synthetic moving edge pattern.

pels horizontal per frame. Within the width of the edge, $\mathbf{G}_n(q)$ of (9) is given by

$$\mathbf{G}_n(q) = \begin{bmatrix} g & g & g & \cdots & g \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \phi_n. \quad (37)$$

For a sequency-ordered basis set $\{\phi_n\}$, $\mathbf{G}_n(q)$ will be zero for all ac-basis vectors with the result that the corresponding displacement estimate updates will be determined solely by noise that may be present on the edge. Updating displacement estimate by iterating over these basis vectors can only increase estimation error. The dc-basis vector $(1/\sqrt{N})$, however, results in

$$\mathbf{G}_n(q) = \begin{pmatrix} g\sqrt{N} \\ 0 \end{pmatrix}, \quad (38)$$

which provides signal-dependent terms in (18) proportional to \sqrt{N} . This suggests that, for this example, it may be better to iterate repeatedly (say M times) over the dc-basis vector than to sequence through the M -basis vectors. We have not been able to analyze rigorously the performance of such dc-basis iteration on the moving edge in the presence of additive noise. In Appendix C we assume that the additive noise is white with power σ_w^2 , and invoke certain assumptions regarding the independence of noise terms entering the recurrence relation. The result is the following approximate expressions for the horizontal component of displacement estimation error mean $\eta_\Delta(i)$ and steady-state variance σ_Δ^2 .

$$\eta_\Delta(i) \sim \Delta(0)(1 - \beta)^i; \quad i = 0, 1, 2, \dots \quad (39a)$$

$$\sigma_\Delta^2 \sim 2\alpha(1 + \alpha)\beta^2 \left\{ \frac{1}{2\beta - \beta^2} + \frac{\beta^2\alpha}{1 - \beta^2\alpha} \right\}, \quad (39b)$$

where $\beta = \epsilon Ng^2$ and $\alpha = \sigma_w^2/Ng^2$.

Not too surprisingly, expressions (39) indicate that, for constant rate of convergence (i.e., constant β), the steady-state displacement estimate error variance decreases inversely with block size N . This points out a possible advantage of transform domain displacement estimation compared to pel-recursive displacement estimation where the dc-basis block size is constrained to have dimension $N = 1$. Figures 11 through 13 show experimental and theoretical behavior of the horizontal component of displacement estimation error for pel-recursive and 2×8 dc-basis iteration on the moving edge. In obtaining the sample mean $\hat{\eta}_\Delta(i)$ and sample variance $\hat{\sigma}_\Delta^2$, averages were taken over 128 field lines.

Experimental and theoretical results are seen to compare favorably in the pel-recursive estimator case (Fig. 11) at a signal-to-noise ratio (SNR) = 45 dB. In Fig. 12, which describes displacement errors in the

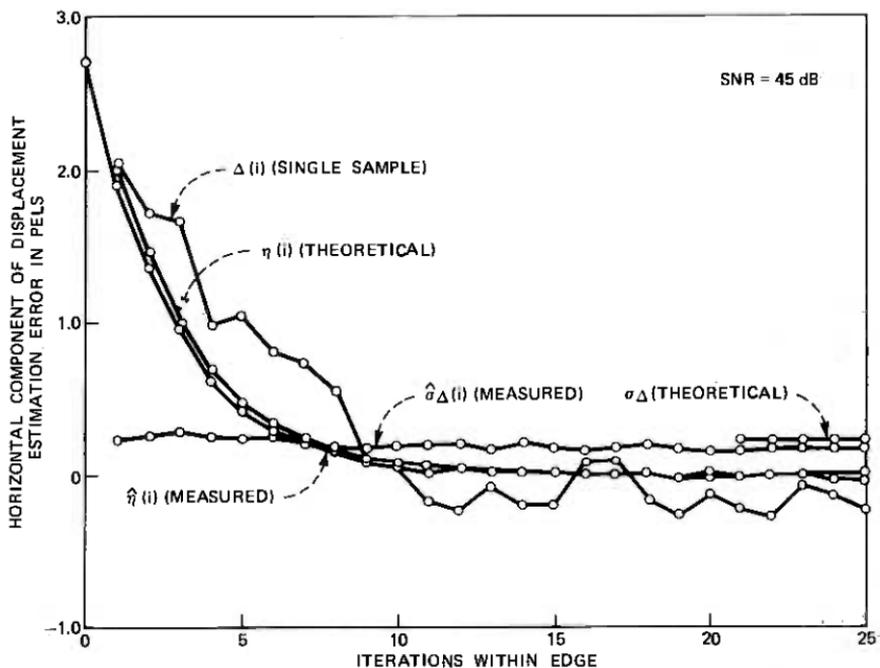


Fig. 11—Horizontal component of displacement estimation error for noisy moving edge (SNR = 45 dB) using pel-recursive displacement estimation. $\epsilon = 0.02$.

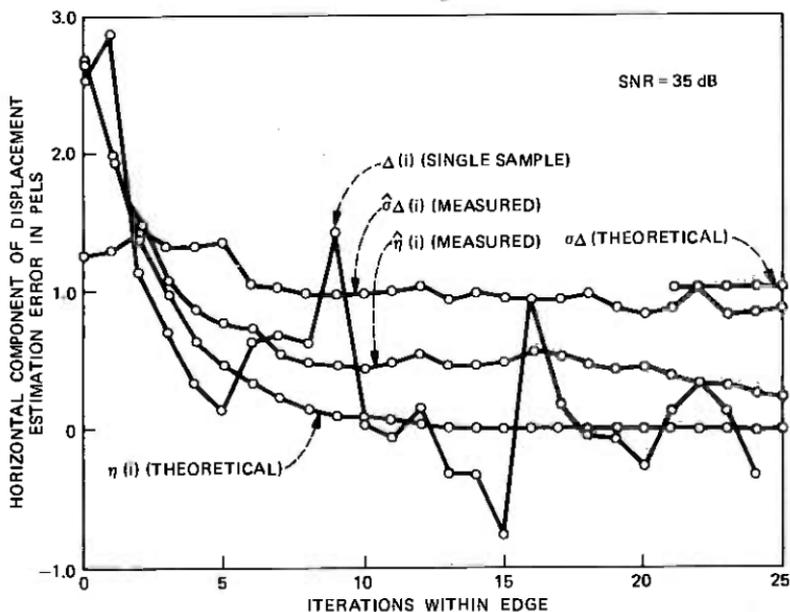


Fig. 12—Horizontal component of displacement estimation error for noisy moving edge (SNR = 35 dB) using pel-recursive displacement estimation. $\epsilon = 0.02$.

pel-recursive estimator at 35 dB, there is approximate agreement between theory and experimental data. This is also the case for the 2×8 dc-basis results of Fig. 13. From the experimental data in Figs. 12 and 13, it can be seen that the "dc-iteration" is more effective than pel-recursive estimation in combating the effects of noise in the displacement estimation of a moving edge.

The theoretical and experimental behavior of steady-state displacement error variance $\sigma_{\Delta}^2(\infty)$ versus β is plotted in Fig. 14 for pel-recursive displacement estimate for the moving edge pattern with SNR = 35 dB. The range $1 < \beta < 2$ represents oscillatory convergence with $\sigma_{\Delta}^2(\infty)$ increasing rapidly as β approaches 2. Note the trade-off between convergence rate and accuracy evidenced by Fig. 14 and eq. (39a).

4.2 Adaptation algorithm

The adaptive displacement estimation algorithm proposed here updates the displacement estimate (8) using that basis vector ϕ_{m0} whose projection onto the computed coefficient gradient of the reference frame has maximum amplitude. At each iteration step (n, q), the magnitude of the coefficient gradient vector $\|G_n(q)\|$ is computed from

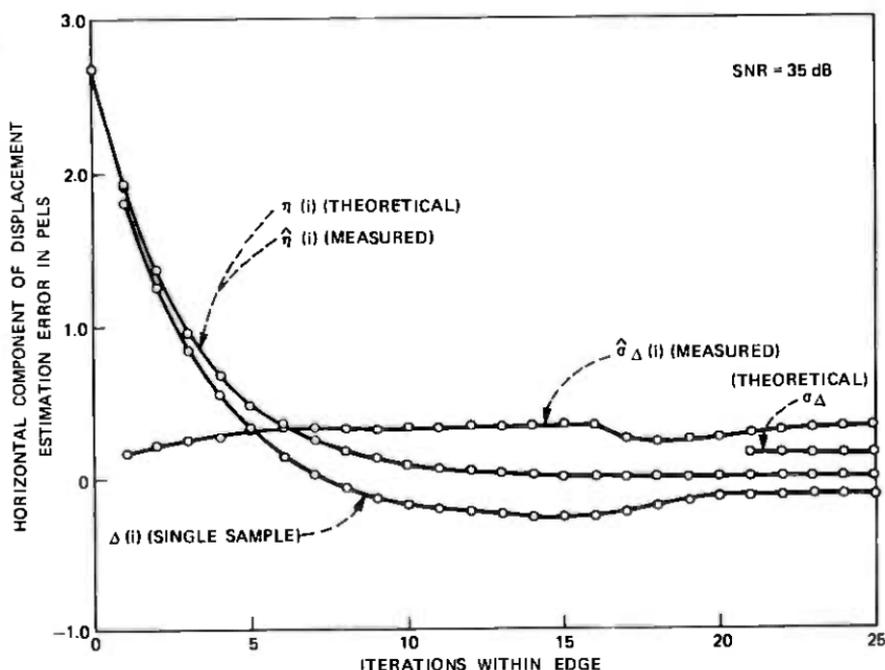


Fig. 13—Horizontal component of displacement estimation error for noisy moving edge (SNR = 35 dB) using 2×8 dc basis iteration. $\epsilon = 0.00125$. The factor ϵ was adjusted in this experiment to result in an identical average convergence rate as that of Figs. 11 and 12.

the *noisy* previous frame data for each vector in the given basis set $\{\phi_m; m = 0, 1, \dots, M - 1\}$. The particular basis ϕ_{m_0} maximizing this quantity is then used for the displacement update.

Figure 15 compares the performance of adaptive and nonadaptive displacement estimation using a separable 2×4 cosine transform and $\epsilon = 10^{-5}$ for the test image of Fig. 16. This image consists of the moving

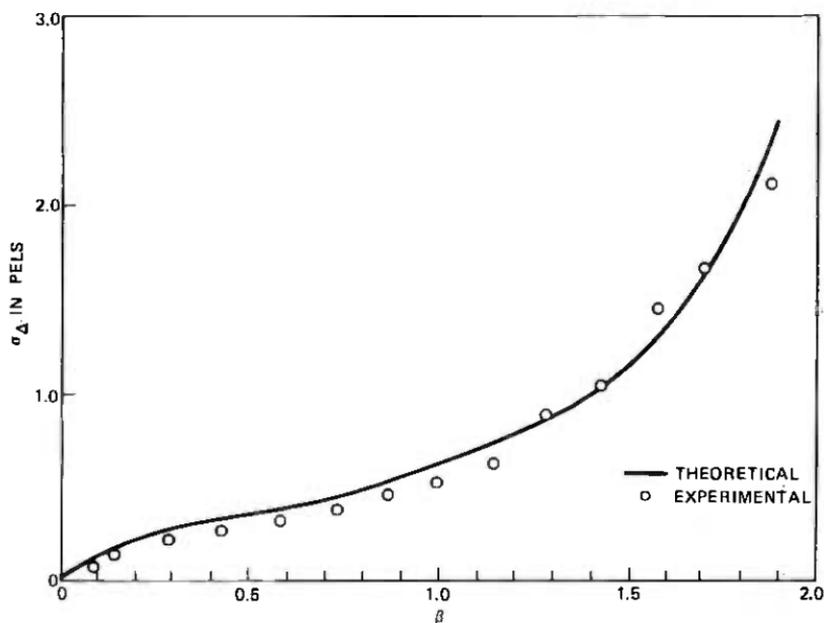


Fig. 14—Estimation error standard deviation vs rate parameter β : pel-recursive at SNR = 45.

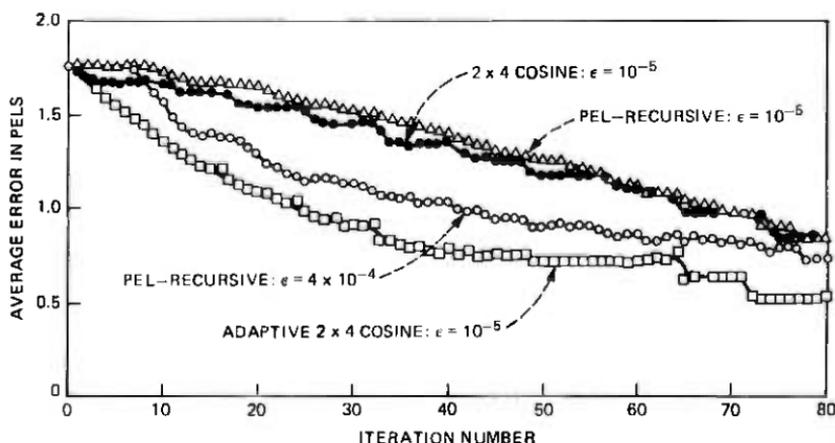


Fig. 15—Displacement estimation errors for noisy synthetic image of Fig. 16.

object of Fig. 6 corrupted by additive white Gaussian noise at 20-dB SNR. Also shown are results for pel-recursive displacement estimation at $\epsilon = 10^{-5}$ and a convergence-rate-optimized $\epsilon = 4 \times 10^{-5}$. The pel-recursive scanning pattern was chosen according to Fig. 17 to match the rate of progression of the pel-recursive algorithm along individual scan lines to that of the 2×4 transform algorithm. (For example, after

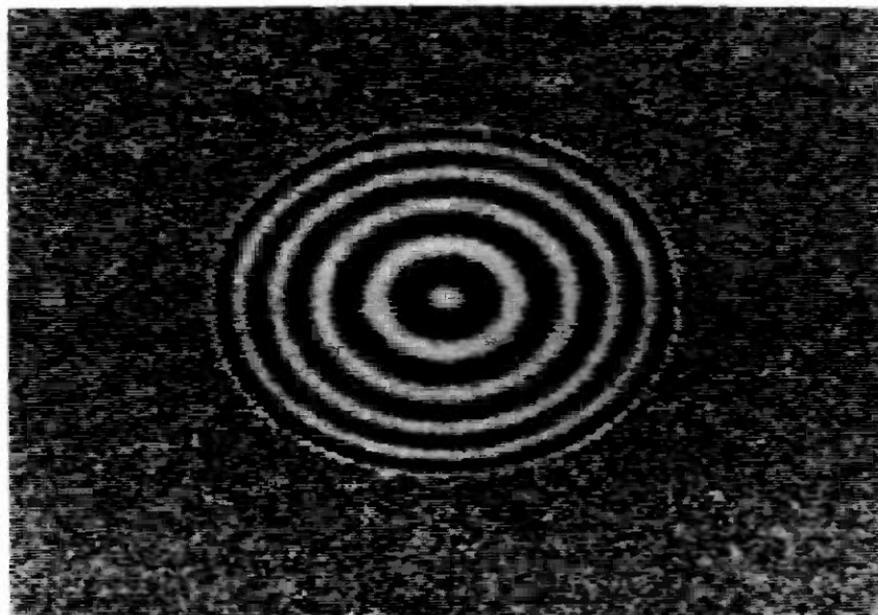


Fig. 16—Noisy test image. SNR = 20 dB.

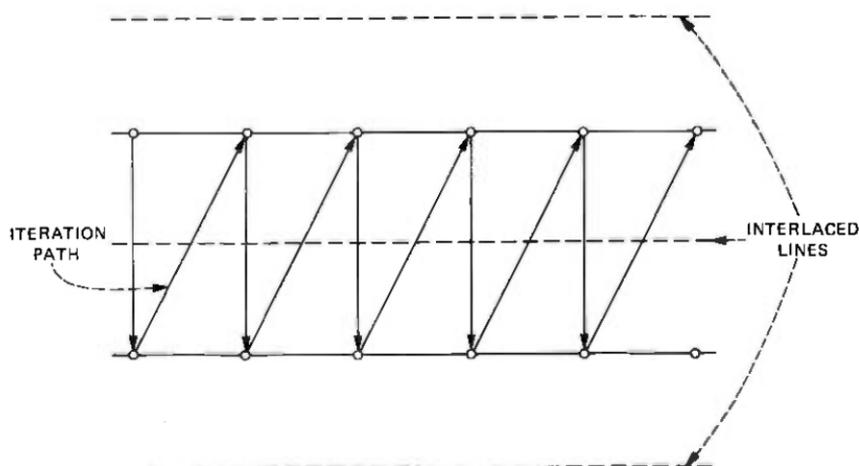


Fig. 17—Pel-recursive scanning pattern applicable to test results of Fig. 15.

40 iterations, both pel-recursive and coefficient-recursive displacement estimators will have traversed 20 columns of a given scan line.) All results of Fig. 15 are averages of relevant displacement error over the interior of the moving object. It can be seen from Fig. 15 that the adaptive 2×4 cosine displacement estimator has clearly superior average error convergence rate than either the pel-recursive or nonadaptive cosine displacement estimators. No choice of ϵ was found that could improve the convergence rate of the pel-recursive estimator beyond that shown for $\epsilon = 4 \times 10^{-5}$. We also computed the experimental standard deviations of relevant displacement estimation error. Nonadaptive cosine and pel-recursive estimators had experimental error standard deviations of 0.50 and 0.51, respectively, at $\epsilon = 10^{-5}$. The error standard deviation of the adaptive coefficient-recursive displacement estimator was 0.54, while that for the rate-optimized pel-recursive estimator was an inferior 0.67.

The above results demonstrate a potential advantage of adaptive coefficient-recursive displacement estimation over both pel-recursive and nonadaptive coefficient-recursive displacement estimation. It remains to be established, however, whether the adaptive scheme of this section will improve the performance of a motion-compensated hybrid transform-DPCM coder.

V. SUMMARY

This paper has introduced a coefficient-recursive displacement estimator having potential application in motion-compensated inter-frame hybrid transform-DPCM image coders. The convergence of the mean displacement estimate to the true displacement was established in Assertions 1 and 2 using assumptions that are supported by the analyses of Appendices A and B. Assertion C described conditions under which the rate of convergence of mean displacement estimation error is independent of the transform block size and type. An extension of the coefficient recursive algorithm was given in Section IV and shown by simulation to have improved convergence properties in the displacement estimation of noisy objects.

APPENDIX A

This appendix verifies a statement in Section III concerning the orthogonality of coefficient vectors $G_n(q)$ $n = 0, 1, \dots, M - 1$ for a separable Markov image model. This model is described as follows. Let $I_{m,n}$ denote the intensity of the pel located in the m th row and n th column of the raster. Then for the Markov image model treated here:

$$\overline{I_{m,n}I_{i,j}} = \sigma^2 \rho_r^{|m-i|} \rho_c^{|n-j|}, \quad (40)$$

with $0 < |\rho_r| < 1$, $0 < |\rho_c| < 1$. In (40), σ^2 is the intensity variance, and ρ_r and ρ_c are the correlation coefficients of adjacent pels between rows and columns, respectively. If the pel intensities in a N_r row by N_c column block are indexed in column scan fashion and denoted by a column vector \mathbf{I} , then it can be shown that

$$\overline{\mathbf{I} \cdot \mathbf{I}^T} = \sigma^2 \mathbf{M}_r \times \mathbf{M}_c, \quad (41)$$

where \times denotes the Kronecker product,¹³ and \mathbf{M}_r and \mathbf{M}_c are, respectively, $N_r \times N_r$ and $N_c \times N_c$ Toeplitz matrices having ij th entries

$$[\mathbf{M}_r]_{ij} = \rho_r^{|i-j|}$$

and

$$[\mathbf{M}_c]_{ij} = \rho_c^{|i-j|}. \quad (42)$$

Covariance models of this form have been widely applied in image processing studies.¹⁴⁻¹⁸ Expressions for the eigenvalues and eigenvectors of \mathbf{M}_r (or \mathbf{M}_c) are given in Ref. 18.

We now apply this model to a study of the covariance properties of $G_n(q)$, $n = 0, 1, \dots, M-1$. From (9) we have

$$\overline{\mathbf{G}_n(q) \mathbf{G}_m^T(q)} = \begin{bmatrix} \phi_n^T \mathbf{R}_1 \phi_m & \phi_n^T \mathbf{R}_{12} \phi_m \\ \phi_n^T \mathbf{R}_{12} \phi_m & \phi_n^T \mathbf{R}_2 \phi_m \end{bmatrix}. \quad (43)$$

As described in Section III, matrices \mathbf{R}_1 , \mathbf{R}_2 , and \mathbf{R}_{12} are auto- and cross-correlation matrices of the spatial derivatives of I . For the discrete image model specified by (40), we compute these spatial derivatives as the corresponding spatial differences. For example, derivatives in the row direction of the raster are given by $\overline{(I_{m,n} - I_{m,n-1})(I_{i,j} - I_{i,j-1})}$. Expanding this product and using (40) gives

$$\overline{(I_{m,n} - I_{m,n-1})(I_{i,j} - I_{i,j-1})} = \sigma^2 \rho_r^{|m-i|} [\alpha_c \rho_c^{|n-j|} + \beta_c \delta_{nj}], \quad (44)$$

where $\alpha_c = 2 - \rho_c^{-1} - \rho_c$, $\beta_c = \rho_c^{-1} - \rho_c$, and δ_{nj} is the Kronecker delta function. By comparing (40) and (41) with (44), we have

$$\mathbf{R}_1 = \sigma^2 \mathbf{M}_r \times [\alpha_c \mathbf{M}_c + \beta_c \mathbf{U}], \quad (45)$$

where \mathbf{U} is an identity matrix.

Similarly, with $\alpha_r = 2 - \rho_r^{-1} - \rho_r$ and $\beta_r = \rho_r^{-1} - \rho_r$:

$$\mathbf{R}_2 = \sigma^2 [\alpha_r \mathbf{M}_r + \beta_r \mathbf{U}] \times \mathbf{M}_c \quad (46)$$

and

$$\mathbf{R}_{12} = \sigma^2 [\alpha_r \mathbf{M}_r + \beta_r \mathbf{U}] \times [\alpha_c \mathbf{M}_c + \beta_c \mathbf{U}]. \quad (47)$$

As shown in Ref. 13, the eigenvectors of a Kronecker product $A \times B$ have the form

$$\begin{bmatrix} x_1^i y^j \\ x_2^i y^j \\ \vdots \\ x_N^i y^j \end{bmatrix},$$

where x_k^i , $k = 1, 2, \dots, N$, denotes the components of the eigenvector \mathbf{x}^i of A and y^j denotes an eigenvector of B . Since any vector is an eigenvector of an identity matrix, it follows that the eigenvectors of R_1 , R_2 , and R_{12} above are in fact identical to the eigenvectors of the image covariance matrix (41). The normalized eigenvectors form a complete *orthonormal* set and are considered to be the optimum bases for transform coding image blocks modeled by (41). Selecting $\{\phi_n\}$ to be this set of eigenvectors, it follows that all terms in the matrix of (43) will be identically zero for $n \neq m$, which establishes the statistical orthogonality of the $\mathbf{G}_n(q)$.

APPENDIX B

This appendix shows that mean displacement estimation error for dependent $\mathbf{G}_{((i))}([[i]])$ is approximately given by recursion (19) for small ϵ .

Iterating (18) yields

$$\Delta_{i+1} = \left(\prod_{j=0}^i [U - \epsilon \mathbf{G}_{((j))}([[j]]) \mathbf{G}_{((j))}^T([[j]])] \right) \Delta_0. \quad (48)$$

The matrix product premultiplying Δ_0 in (48) is a function of ϵ . Taylor's expansion of this function about $\epsilon = 0$ yields (for fixed i):

$$\Delta_{i+1} = \left(U - \epsilon \prod_{j=0}^i \mathbf{G}_{((j))}([[j]]) \mathbf{G}_{((j))}^T([[j]]) \right) \Delta_0 + O(\epsilon^2) \quad (49)$$

so that

$$\bar{\Delta}_{i+1} = \left(U - \epsilon \sum_{j=0}^i \overline{\mathbf{G}_{((j))} \mathbf{G}_{((j))}^T} \right) \bar{\Delta}_0 + \bar{O}(\epsilon^2). \quad (50)$$

On the other hand, repeating the steps of (48) and (49) on (19) gives (for the independent $\mathbf{G}_{((j))}([[j]])$ case):

$$\bar{\Delta}_{i+1} = \left(I - \epsilon \sum_{j=0}^i \overline{\mathbf{G}_{((j))} \mathbf{G}_{((j))}^T} \right) \bar{\Delta}_0 + O'(\epsilon^2). \quad (51)$$

Comparing (50) and (51) we have the result that mean estimation errors for the dependent and independent cases are equal to an

approximation $\bar{O}(\epsilon^2) - O'(\epsilon^2)$, which can be neglected for sufficiently small ϵ .

APPENDIX C

This appendix briefly sketches the steps leading to (39).

We consider the successive frames of video to consist of a moving edge contaminated by independent noise:

$$\begin{aligned} I(\mathbf{x}, \tau) &= g(x_1 - D) + w(\mathbf{x}, t) \\ I(\mathbf{x}, t - \tau) &= g x_1 + w(\mathbf{x}, t - \tau), \end{aligned} \quad (52)$$

where the noise $w(\cdot, \cdot)$ is white with respect to both pel-to-pel and frame-to-frame dimensions. An analysis similar to that leading to (18) then yields

$$\begin{aligned} \Delta_{i+1} &= \Delta_i - \epsilon[\mathbf{G}_n(q) + \nabla[\mathbf{W}^T(\mathbf{x}_{[i]}) - \hat{\mathbf{D}}_{i,t} - \tau)\phi_n]] \\ &\cdot [\mathbf{G}_n^T(q)\Delta_i + \mathbf{W}^T(\mathbf{x}_{[i]})\phi_n - \mathbf{W}^T(\mathbf{x}_{[i]}) - \hat{\mathbf{D}}_{i,t} - \tau)\phi_n]. \end{aligned} \quad (53)$$

We consider iterating (53) repeatedly, using dc-basis vector $\phi_0 = 1/\sqrt{N}$ and $\mathbf{G}_n(q)$ given by (38). Let $\Delta(i)$ denote the displacement error in the horizontal direction at iteration i ; $Z_1(i)$ denote the difference of the two noise terms in the final term of (53); and $Z_2(i)$ denote the horizontal component of the noise gradient term in (53). Then the horizontal component of (53) becomes

$$\Delta(i+1) = \Delta(i) - \epsilon N g^2 [\Delta(i) + Z_1(i)/\sqrt{N}g][1 + Z_2(i)/\sqrt{N}g]. \quad (54)$$

Defining $\beta = \epsilon N g^2$, $\gamma = 1/\sqrt{N}g$ and rearranging terms yield

$$\Delta(i+1) = [1 - \beta(1 + \gamma Z_2(i))]\Delta(i) - \beta\gamma(1 + \gamma Z_2(i))Z_1(i). \quad (55)$$

Note that each noise term in (55) is multiplied by a factor γ which, for given β , decreases as $1/\sqrt{N}$. Neglecting dependencies between $\{Z_2(i)\}$, and $\{Z_1(i)\}$, this equation is linear with respect to input Z_1 and output Δ . The solution has the form

$$\Delta(i) = \Delta(0) \prod_{j=1}^{i-1} [1 - \beta(1 + \gamma Z_2(j))] + \sum_{k=0}^i h(i, k) Z_1(k), \quad (56)$$

where $h(i, k)$ is the response of (55) for $\Delta(0) = 0$ and input $Z_1(i) = \delta_{ik}$. By assuming that $\{Z_1(i)\}$ and $\{Z_2(i)\}$ are mutually independent white sequences, the mean and variance of $\Delta(i)$ can now be derived by a tedious but conventional analysis, resulting in (39).

REFERENCES

1. A. N. Netravali and J. D. Robbins, "Motion-Compensated Television Coding: Part I," B.S.T.J., 58, No. 3 (March 1979), pp. 631-670.

2. J. D. Robbins and A. N. Netravali, "Interframe Coding Using Movement Compensation," Int. Conf. on Commun., June 1979.
3. D. O. Reudink, unpublished work.
4. A. Habibi, "Hybrid Coding of Pictorial Data," IEEE Trans. Commun., COM-22, No. 5 (May 1974), pp. 614-624.
5. C. Reader, "Orthogonal Transform Coding of Still and Moving Pictures," Ph.D. dissertation, The University of Sussex, United Kingdom, 1974.
6. J. A. Roese, W. K. Pratt, and G. S. Robinson, "Interframe Cosine Transform Image Coding," IEEE Trans. Commun., COM-25, No. 11 (November 1977), pp. 1329-1339.
7. H. W. Jones, "A Conditional Replenishment Hadamard Video Compressor," SPIE, 119, Applications of Digital Image Processing, 1977, pp. 91-98.
8. A. N. Netravali and J. A. Stuller, "Motion-Compensated Transform Coding," B.S.T.J., this issue, pp. 1703-1718.
9. S. C. Knauer, "Real-Time Video Compression Algorithm for Hadamard Transform Processing," Proceedings of SPIE, 66 (August 1975), pp. 58-69.
10. H. C. Andrews and C. L. Patterson, "Outer Product Expansions and Their Uses in Digital Image Processing," American Math. Monthly, 82, No. 1 (January 1974), pp. 1-13.
11. B. Widrow, "Adaptive Filters," in *Aspects of Network and System Theory*, R. Kalman and N. DeClaris, eds., New York: Holt, Rinehart, and Winston, 1971, pp. 563-587.
12. P. A. Wintz, "Transform Picture Coding," Proc. IEEE, 60 (July 1972), pp. 809-820.
13. R. Bellman, *Introduction to Matrix Analysis*, 2nd ed., New York: McGraw-Hill, 1970, pp. 235.
14. A. Arcese, P. H. Mengert, and E. W. Trombini, "Image Detection through Bipolar Correlation," IEEE Trans. Inform. Theory, IT-16, No. 5 (September, 1970), pp. 534-541.
15. A. K. Jain and E. Angel, "Image Restoration, Modelling, and Reduction of Dimensionality," IEEE Trans. Computers, C-21 (July 1972), pp. 636-641.
16. J. A. Stuller and B. Kurz, "Two-Dimensional Markov Representations of Sampled Images," IEEE Trans. Commun., COM-24 (October 1976), pp. 1148-1152.
17. J. A. Stuller and B. Kurz, "Interframe Sequential Picture Coding," IEEE Trans. Commun., COM-25, No. 5 (May 1977), pp. 485-495.
18. W. K. Pratt, "Generalized Wiener Filtering Computation Techniques," IEEE Trans. Computers, C-21, No. 7 (July 1972), pp. 636-641.

Motion-Compensated Transform Coding

By A. N. NETRAVALI and J. A. STULLER

(Manuscript received April 10, 1979)

Interframe hybrid transform/DPCM coders encode television signals by taking a spatial transform of a block of picture elements in a frame and predictively coding the resulting coefficients using the corresponding coefficients of the spatial block at the same location in the previous frame. These coders can be made more efficient for scenes containing objects in translational motion by first estimating the translational displacement of objects and then using coefficients of a spatially displaced block in the previous frame for prediction. This paper presents simulation results for such motion-compensated transform coders using two algorithms for estimating displacements. The first algorithm, which is developed in a companion paper, recursively estimates the displacements from the previously transmitted transform coefficients, thereby eliminating the need to transmit the displacement estimates. The second algorithm, due to Limb and Murphy, estimates displacements by taking ratios of accumulated frame difference and spatial difference signals in a block. In this scheme, the displacement estimates are transmitted to the receiver. Computer simulations on two typical real-life sequences of frames show that motion-compensated coefficient prediction results in coder bit rates that are 20 to 40 percent lower than conventional interframe transform coders using "frame difference of coefficients." Comparisons of bit rates for approximately the same picture quality show that the two methods of displacement estimation are quite similar in performance with a slight preference for the scheme with recursive displacement estimation.

I. INTRODUCTION

Television signals, which are generated by scanning a scene 30 times a second, contain a significant amount of frame-to-frame redundancy. A large part of this redundancy can be removed by the technique of conditional replenishment.¹⁻⁵ In conditional replenishment, each frame

is segmented into two parts: background, which consists of picture elements (pels) having intensities similar to the previous frame pels, and moving area, which consists of pels that differ significantly from the previous frame pels. Information is transmitted only about the moving area in the form of prediction errors and addresses of the moving area pels. Conditional replenishment schemes can be improved by estimating the displacement of objects in the scene and using the displacement estimate for predictive coding by taking differences of elements in the moving area with respect to appropriately displaced elements in the previous frame. Such schemes have been referred to as motion-compensated coding schemes.⁶⁻¹¹

Transform domain methods have been widely discussed for bandwidth compression of still images or single frames.¹² They can also be used for coding of sequences of television frames by taking a two-dimensional spatial transform followed by predictive coding using corresponding coefficients from the spatial transform of the previous frame.¹³⁻¹⁶ This type of hybrid coding¹⁷ relieves the storage problems associated with the use of three-dimensional transform blocks. Such a scheme can be made more efficient for scenes containing objects in motion by using, for prediction, coefficients of blocks from the previous frame that are spatially displaced from the present frame block by an amount equal to the displacement of objects. As in the pel domain, the success of motion compensation in transform coders depends upon: (i) the amount of purely translational motion of objects in the scene, (ii) the ability of the displacement estimation algorithm to estimate the translation with an accuracy necessary for good prediction of the coefficients, and (iii) the robustness of the displacement estimation algorithm when the resolution of the transmitted picture is changed to match the coder bit rate to the channel rate.

In this paper, we use two previously published displacement estimation algorithms for motion-compensated transform coding. The first algorithm is an extension of a corresponding method in the pel domain.^{10,11} It works recursively on the previously transmitted transform coefficients of the present as well as the previous frame. It therefore requires no separate transmission of the displacement estimate. This algorithm is discussed in detail in a companion paper,¹⁸ where its properties are described both analytically and experimentally in terms of certain simple synthetically generated scenes. The other method of displacement estimation that we use is due to Limb and Murphy.¹⁹ It estimates displacements in a block of pels using a ratio of accumulated frame difference and spatial difference signals from future as well as past data. These displacement estimates are nonrecursive and must be transmitted separately to the receiver. The present paper investigates the performance of the two displacement estimation algorithms

in the context of interframe coders operating on real-life scenes that contain fairly complex (nontranslational) motion. Results are given here on the effects of various coder parameters such as block size, particular transform (Hadamard, cosine, etc.), and other parameters of the displacement estimators. The primary result of this paper is that the application of either recursive or nonrecursive motion estimation provides a 20 to 40 percent decrease in bit rate, compared to conventional, uncompensated hybrid transform/DPCM coding. We have found that the use of large block sizes in motion estimation degrades the coder performance. This may be a result of spatially nonuniform displacements being averaged over the transform block by the displacement estimator. Also, since the motion in real scenes is generally not uniform in rectangular blocks, as the block size is increased, only a fraction of elements in a block are compensable with a given displacement, and therefore transmitting coefficients of a larger block containing some compensable and some uncompensable pels becomes inefficient.

III. HYBRID TRANSFORM CODING WITHOUT MOTION COMPENSATION

In an interframe hybrid transform-DPCM coder, a field of video is partitioned into blocks having dimensions N_r rows by N_c columns, and a two-dimensional transform is performed on each block to obtain a set of coefficients. Transform coefficients of the q th block of the present frame are predicted by the corresponding coefficients of the q th block of the previously encoded frame, and, if the prediction error is above a specified threshold, the quantized prediction errors are transmitted to the receiver. These quantized errors are added to the coefficients predicted by the receiver, which inversely transforms the result to obtain an image for display at the receiver. A block diagram of an interframe hybrid transform-DPCM transmitter is shown in Fig. 1. Data compression is achieved both by the redundancy removal implicit in the prediction process and because some coefficients can be reproduced with low precision (or totally omitted from transmission) without visible degradation in the reconstructed picture.

The performance of the interframe hybrid transform-DPCM coder and the other coders described in later sections of this paper is evaluated in terms of bit rate for an acceptable subjective picture quality using two scenes, one called Judy and one Mike and Nadine. The coding degradation was judged in informal tests by the authors to be just perceptible from a viewing distance of six times the picture height. These scenes consist of 64 frames (2:1 interlaced fields) of 256×256 samples each, obtained at 30 times a second and sampled at Nyquist rate from a video signal of 1-MHz bandwidth. The scene Judy contains head-and-shoulders views of a person engaged in a rather

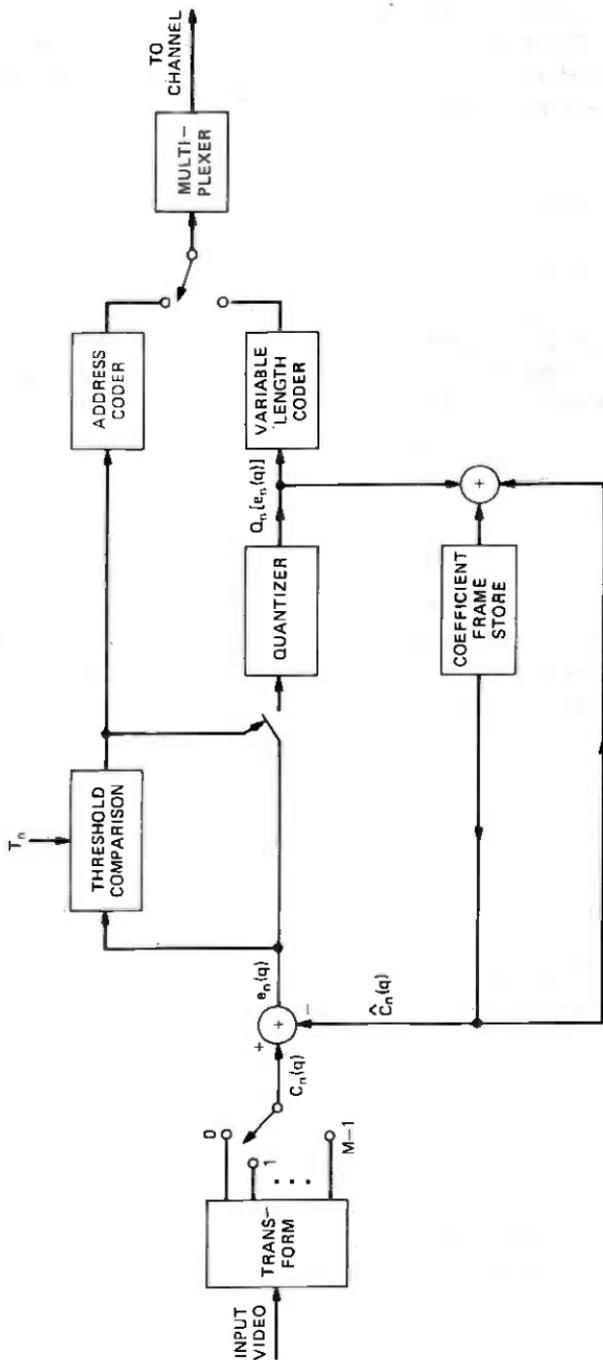


Fig. 1—Block diagram of hybrid transform/DPCM interframe image encoder.

active conversation. The portion of a frame classified as moving area varies from 15 to 51 percent. The motion is not strictly translational, and there are different parts of the scene moving differently (such as lips, eyes, and head). Four frames of this scene are shown in Fig. 4 of Ref. 10. The scene Mike and Nadine contains a panned full-body view of two people briskly walking around each other on a set with severe nonuniform and time-varying illumination. The percentage of a frame classified as moving area varied from 92 to 96 percent. Four frames from this sequence are shown in Fig. 5 of Ref. 10.

In our simulations of the interframe hybrid transform-DPCM (called conditional replenishment in the transform domain), the coefficients of two corresponding spatial blocks of the same field from two successive frames are compared, and if the difference is more than a threshold, the coefficient is transmitted. Thus, if $\{c_k\}_{k=0, \dots, M-1}$, and $\{\bar{c}_k\}_{k=0, \dots, M-1}$ are M selected coefficients (out of N coefficients in a block) of the present and coded previous frame blocks, respectively, then the quantized error, $Q_k[c_k - \bar{c}_k]$, is transmitted only if $|c_k - \bar{c}_k| \geq T_k$, where $Q_k[\cdot]$ is the quantizer for the k th coefficient, and T_k is the threshold. If c_k is not transmitted, then its value at the receiver is assumed to be \bar{c}_k . Thus the transmission consists of the quantized prediction error of the coefficients that were selected for transmission and the addresses of the coefficients that were dropped from the transmission. The information necessary to convey addresses of the coefficients selected for transmission was computed based on the run-length coding of runs of coefficients within a block and then from block to block. Parameters of the coder such as the number of coefficients that were entirely dropped from the transmission, the thresholds $\{T_k\}$ for selecting the transmitted coefficients, and the quantizer scales were adjusted* to produce pictures in which coding degradations were just perceptible. The entropies of the prediction errors and the run lengths specifying addresses of the transmitted coefficients are added to compute the total bit rate.

The results are shown in Fig. 2, in which the bit rate is plotted as a function of the frame number for 60 frames. In these simulations and those of the next section, the coder was initialized so that it used the unquantized original first frame for prediction of the second frame. For comparison, the results from Ref. 10 are reproduced for conditional replenishment in pel domain. The comparison shows that, in the transform domain, using a cosine transform on a 2×4 block, there is a reduction of about 10 percent in bit rate over that obtained in pel

* We do not claim that these adjustments resulted in an optimum set of parameters. However, a sufficiently large set of parameters was tried, giving us confidence that our results are not far from the optimum.

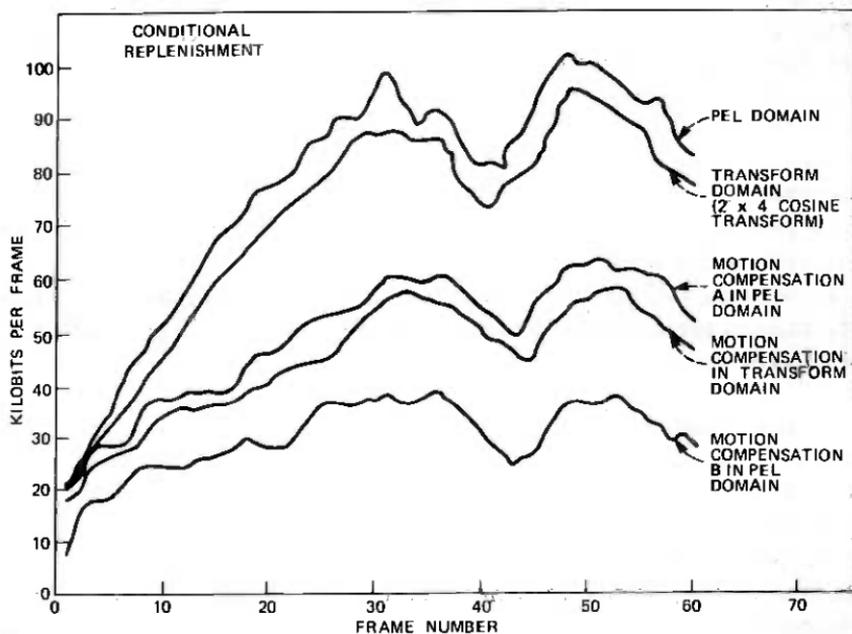


Fig. 2—Performance of conditional replenishment and motion-compensated transform coders. Kilobits/frame are plotted as a function of the frame number for a typical sequence containing active motion of a head-and-shoulders view.

domain.* For this particular case, we dropped the eighth coefficient entirely, and the prediction errors for the other seven coefficients were quantized with uniform quantization scales with step sizes of 3, 5, 5, 7, 7, 9, 9, respectively. The thresholds $\{T_k\}$ for predictability were chosen to be 1, 2, 2, 3, 3, 4, 4 (out of 255) for the seven coefficients, respectively.

We varied some parameters of the transform to evaluate the sensitivity of these results to the block size and the type of transform used. Some of these are shown in Fig. 3. It is seen from this figure that a one-dimensional cosine transform with four elements did worse than the conditional replenishment in the pel domain (between 5 to 10 percent). As the transform size was increased, the bit rate dropped; for 2×2 block and cosine transform, the results were similar to the conditional replenishment in pel domain; the 2×8 block using the cosine transform, on the other hand, did about 15 percent better than the conditional replenishment in pel domain. We also tried different transforms and found that for small block sizes they were equivalent to the cosine transform but, as the block size was increased, the cosine

* Of course, several other modifications can be made to improve the pel domain conditional replenishment. Our comparison is not meant to be a comparison between pel domain and transform domain coding in general.

transform behaved better than the other transforms. The results for the 2×8 block using the Hadamard transform basis were very similar to those of 2×4 block and cosine transform but were inferior to those of the 2×8 block and cosine transform.

Figure 4 shows the distribution of the bits required for addressing and for the transmission of the first coefficient. As is seen, the addressing bits are about 50 percent of the total bits. This is a significant increase in addressing requirement compared to the conditional replenishment in the pel domain, where the addressing accounts for only about 20 to 30 percent of the total bits. This may be a result of using only the prediction error corresponding to the coefficient being coded for deciding whether that coefficient should be transmitted. This may have made the decision to transmit a coefficient unnecessarily noisy. We did, however, try several methods of reducing the addressing bits, but none of these resulted in an overall bit rate reduction. In the fraction of the bits that are required to send the prediction errors, those for the first coefficient account for more than half, as shown in Fig. 4. Thus the addressing and the first coefficient take up around 80 percent of the total bits generated by the conditional replenishment coder.

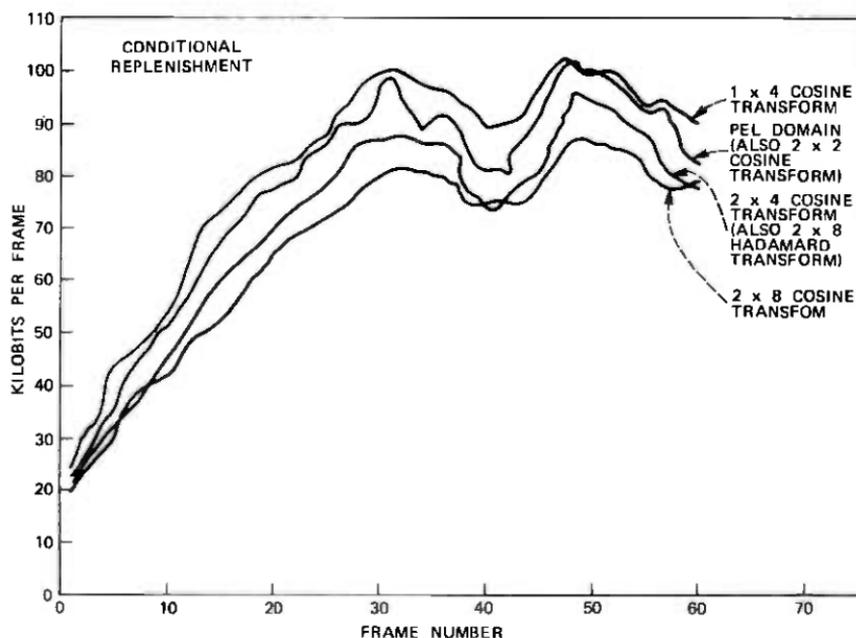


Fig. 3—Performance of conditional replenishment in the transform domain with various transforms.

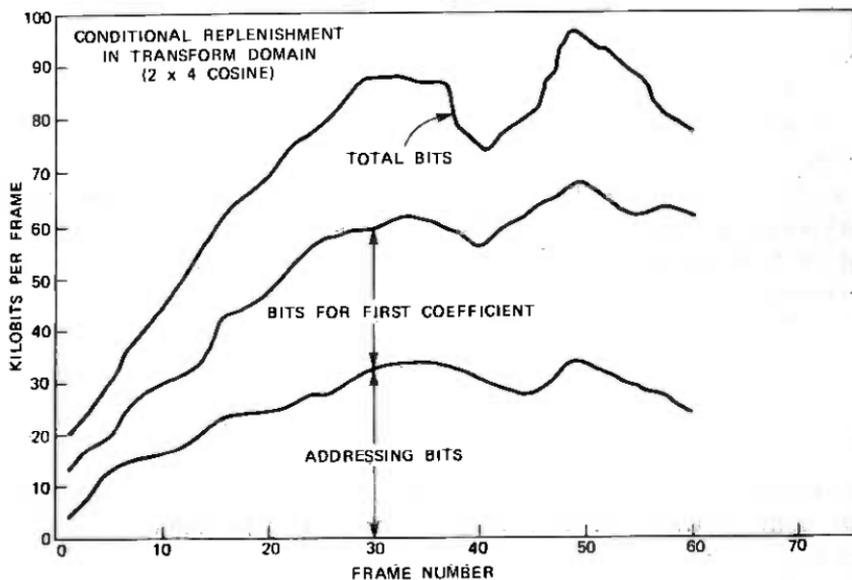


Fig. 4—Distribution of total compressed bits/frame into addressing information and that required for the transmission of the first coefficient.

III. MOTION COMPENSATION WITH RECURSIVE DISPLACEMENT ESTIMATION

In the motion-compensated hybrid transform-DPCM coder shown in Fig. 5, the n th coefficient of the q th present field block is predicted by the n th coefficient of either the displaced or the nondisplaced block of the previous frame, depending on which was better for the $(n - 1)$ th coefficient, where the displacement is an estimate of the frame-to-frame translation of a moving object. The displacement estimation technique used in this section is identical to the one given in our companion paper.¹⁸ We describe it as follows: Let $\mathbf{x}_q = (x_{1q}, x_{2q})^T$ denote the coordinate of the upper left-hand pel of the q th block, where the blocks in each row of blocks are numbered from left to right with $q = 0, 1, 2, \dots$, and superscript T denotes the transpose of a vector or matrix. The pel intensities of block q in a column-scanning fashion are denoted by a column vector $\mathbf{I}(\mathbf{x}_q, t)$. Let the n th basis vector of the transform be denoted by ϕ_n , and, therefore, the n th coefficient of the q th block of the transform of the present frame can be written as

$$c_n(q) = \mathbf{I}^T(\mathbf{x}_q, t)\phi_n. \quad (1)$$

The displaced previous frame value of this coefficient is

$$\hat{c}_n(q, \hat{\mathbf{D}}) = \mathbf{I}^T(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)\phi_n, \quad (2)$$

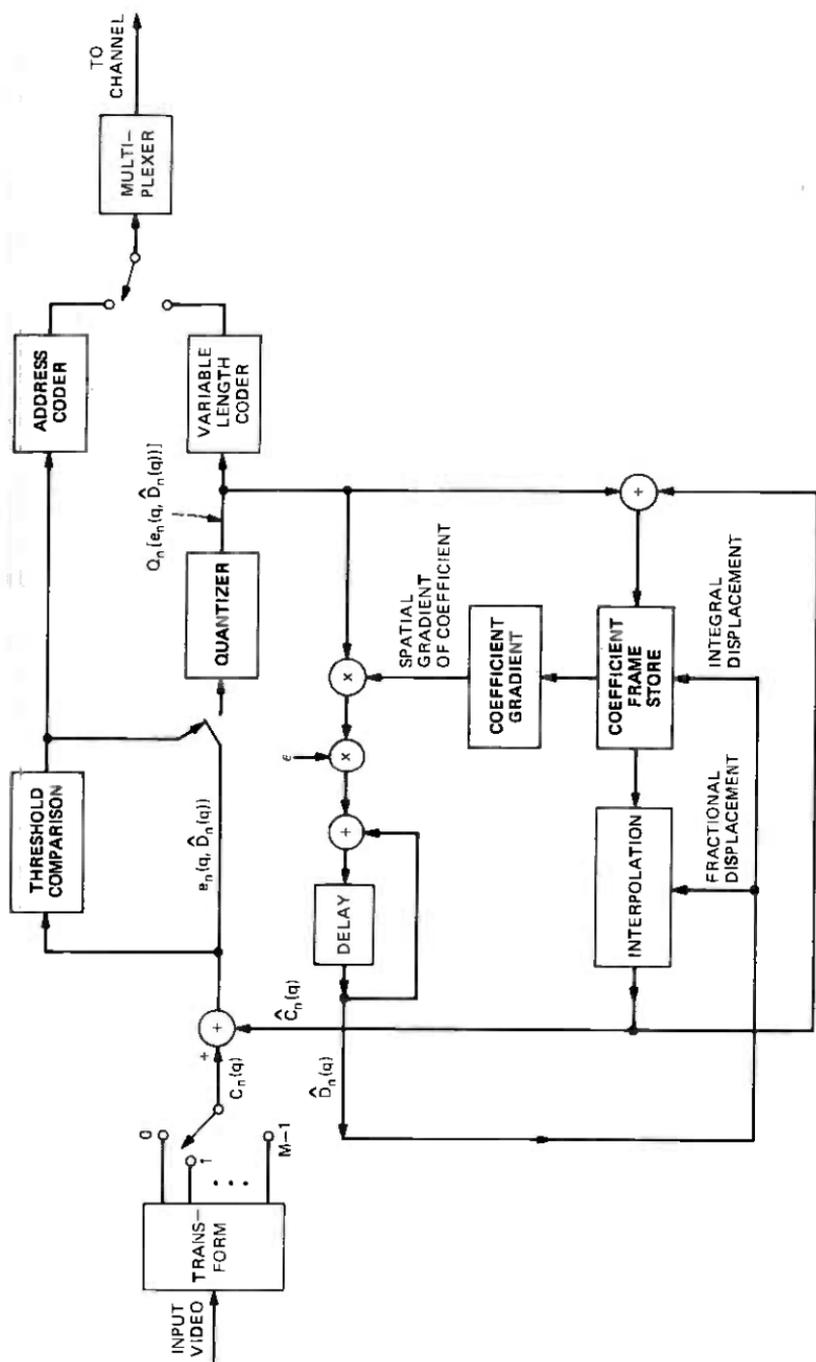


Fig. 5—Block diagram of motion-compensated transform/DPCM coder.

where $\mathbf{I}(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)$ is the column vector of intensities of the displaced q th block of the previous frame and $\hat{\mathbf{D}}$ is the estimated displacement of the moving object. Computation of the elements in $\mathbf{I}(\mathbf{x}_q - \hat{\mathbf{D}}, t - \tau)$ generally requires an interpolation among the given previous-frame pel intensities. The displacement estimation algorithm attempts to minimize the prediction error in predicting $c_n(q)$ by $\hat{c}_n(q, \hat{\mathbf{D}})$ by the steepest descent iteration of the form

$$\begin{aligned} \hat{\mathbf{D}}_{n+1}(q) &= \hat{\mathbf{D}}_n(q) - \frac{\epsilon}{2} \nabla_{\hat{\mathbf{D}}_n(q)} e_n^2(q, \hat{\mathbf{D}}_n(q)) \\ &= \hat{\mathbf{D}}_n(q) - \epsilon e_n(q, \hat{\mathbf{D}}_n(q)) \nabla I^T(\mathbf{x}_q - \hat{\mathbf{D}}_n(q), t - \tau) \boldsymbol{\phi}_n \end{aligned} \quad (3)$$

for $n = 0, 1, \dots, M-2$ and $q = 0, 1, 2, \dots$, with

$$\begin{aligned} \hat{\mathbf{D}}_0(q) &= \hat{\mathbf{D}}_{M-1}(q-1) \\ &\quad - \epsilon e_{M-1}(q-1, \hat{\mathbf{D}}_{M-1}(q-1)) \\ &\quad \cdot \nabla I^T(\mathbf{x}_{q-1} - \hat{\mathbf{D}}_{M-1}(q-1), t - \tau) \boldsymbol{\phi}_0, \end{aligned} \quad (4)$$

where $e_n(q, \hat{\mathbf{D}}_n(q))$ is the error in the prediction of $c_n(q)$ (i.e., $c_n(q) - \hat{c}_n(q, \hat{\mathbf{D}}_n(q))$) and M is the number of displacement iterations performed per block. Thus the iteration proceeds by first assuming the initial displacement estimate of the q th block, $\hat{\mathbf{D}}_0(q)$, as an update from the final displacement estimate of the $q-1$ block $\hat{\mathbf{D}}_{M-1}(q-1)$. The next displacement estimate of the q th block, $\hat{\mathbf{D}}_1(q)$, is formed from eq. (3) with $n = 0$. Iteration progresses in the q th block from coefficient to coefficient, resulting finally in displacement estimate $\hat{\mathbf{D}}_{M-1}(q)$. This iteration procedure continues along all horizontal blocks of the raster. The initial displacement of the leftmost block is assumed arbitrarily to be zero.

Such a motion-compensated transform encoder transmits a quantized version of the coefficient prediction error $e_n(q, \hat{\mathbf{D}}(q))$ to the receiver whenever the magnitude $|e_n(q, \hat{\mathbf{D}}(q))|$ exceeds a given threshold T_n , thereby enabling the decoder to update its displacement estimate $\hat{\mathbf{D}}_n(q)$ as in eqs. (3) and (4), as well as correcting its prediction of coefficient $c_n(q)$. Both the encoder and decoder use the updated displacement estimate in predicting the next coefficient, and the process continues. We note that, since only previously transmitted information is used for displacement updating, no separate transmission of displacement is necessary. A simplified block diagram of the hybrid motion-compensated coder is shown in Fig. 5.

The results of motion-compensated coding in the transform domain for the scene Judy are given in Fig. 2. In this figure, total bits per frame are plotted against the frame number. For purposes of comparison,*

* It should be noted that motion compensation in pel domain used intensities of the previous field rather than frame, whereas motion compensation in transform domain used intensities of the previous frame. It was found that, for the pel domain case (Ref. 10), previous field intensities give better results.

this figure also shows conditional replenishment in the pel domain as well as in the transform domain and the two motion compensation techniques of Refs. 10 and 11 in the pel domain. Motion compensation in the transform domain is about 20 to 40 percent better than conditional replenishment in the transform domain. Also, motion compensation in the transform domain is better than one of the pel domain motion compensation techniques by about 5 to 10 percent. This pel domain technique is described in Ref. 10. It segments a frame into three types of areas: background, compensable moving area, and uncompensable moving area, and, therefore, requires a significant amount of address transmission. Motion compensation in the transform domain results in about 20-percent higher bit rates compared to the second motion compensation technique in pel domain. In this second technique, which is described in Ref. 11, each frame is divided only in two parts, predictable and unpredictable, and thus transmission of moving area address information is eliminated.

Results of motion compensation in the transform domain which uses different types of transforms and block sizes are given in Fig. 6. This figure shows that the cosine transform with 2×4 block does the best. Increasing the block size increases the bit rate, perhaps as a result of the uncompensable area (i.e., the pels for which the prediction error is larger than threshold T_n) being in small isolated fragments. This result is in contrast with the result obtained with larger block sizes in conditional replenishment, where a larger block size, such as 2×8 , gave better results than a smaller block size, such as 2×4 . A one-dimensional transform, for example, the 1×4 block cosine transform, does worse than motion compensation A in the pel domain, whereas a 2×2 block using the cosine transform, on the other hand, is equivalent to motion compensation A in the pel domain.

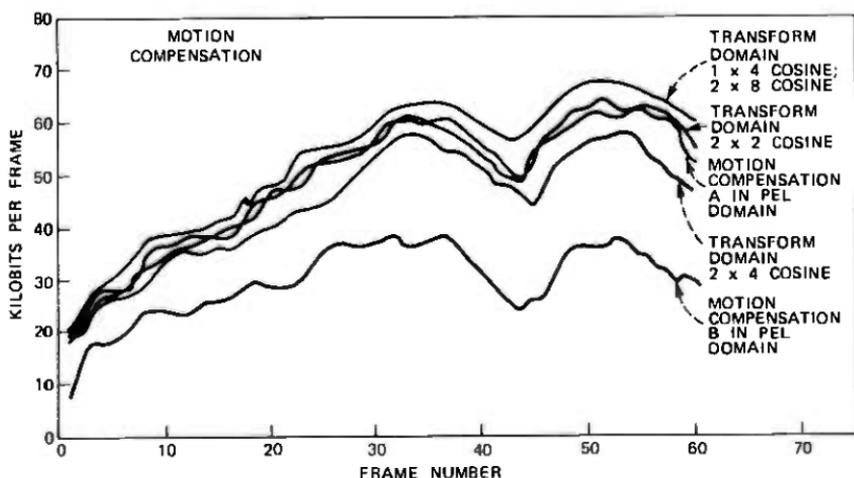


Fig. 6—Performance of motion-compensated coder with different transforms.

Figure 7 is a plot of the portion of the bits required for addressing and the error transmission for the first coefficient. It is seen that, with motion compensation in the transform domain, as in the pel domain, the addressing takes up a significant portion of the total bits. This portion varies from 40 to 60 percent of the total bits. The first coefficient transmission requires more than 50 percent of the bits required for transmission of the coefficients. Although the figure shows the results for 2×4 block and cosine transform, similar results were obtained for other transforms.

We found that more coefficients could be dropped altogether from transmission in the motion-compensated transform coder than in the conditional replenishment transform coder. For example, with a 2×4 block and the cosine transform, only five coefficients were needed in motion compensation, compared to the seven coefficients that were necessary for conditional replenishment. Unfortunately, however, the effect of dropping a larger number of coefficients did not result in a large bit-rate reduction, since the number of bits required for these coefficients was very small.

The results of Fig. 6 were obtained by adjusting the quantization scales and the predictability thresholds (T_n) in such a way that coding degradations in pictures were just perceptible in informal viewing by the authors. The quantization scales that we used were from uniform quantizers with step sizes of 3, 5, 5, 7, 7 (for the first five coefficients of the 2×4 cosine transform), and the predictability thresholds were 2, 3, 3, 4, 4 (out of 255) for the first five coefficients. Coarser quantization of the higher order coefficients was possible in motion compensation,

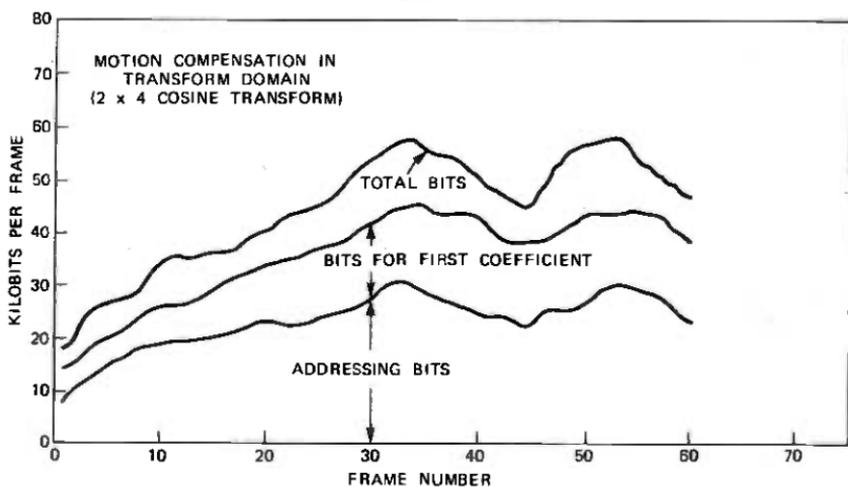


Fig. 7—Distribution of total compressed bits/frame into addressing information and that required for the transmission of the first coefficient.

compared to conditional replenishment, without significantly degrading picture quality. An increase of the predictability thresholds of the first coefficient resulted in rapid degradation of the picture quality. As the predictability threshold was increased, the block structure of the transform became clearly evident, and the frame-to-frame effects of the visibility of the block structure were found rather annoying. For higher order coefficients, however, the picture quality was not a sensitive function of the predictability thresholds. It appears that due to better prediction in motion-compensated transform coding, effects of coarser quantization and higher predictability thresholds are seen in smaller disjoint areas of the picture and, therefore, their visibility is lower.

The recursive displacement estimation algorithm of eq. (2) was also varied by changing ϵ and by changing the order of iteration of the coefficients within a block. We found that ϵ variation did not change the bit rates significantly as long as ϵ was within a decade of 0.0001. As expected,¹⁸ much larger ϵ resulted in noisy estimates of displacement, whereas smaller values of ϵ took longer to converge. The order of iteration was varied by estimating displacement starting from the first coefficient to the fifth (for a 2×4 transform block and cosine transform with transmission of only five coefficients) or starting from the fifth coefficient to the first. This corresponded to iterating from low frequency to high frequency or vice versa. We found that going from high frequency to low frequency resulted in a smaller number of bits for transmission of the prediction error by about 5 to 10 percent. However, since amplitude bits account for only about 50 percent of the total bits, the overall reduction was only about 2 to 5 percent. Another variation that was tried consisted of iterating only the first (dc) coefficient for the entire block; that is, iterating the first coefficient five times instead of iterating all the five coefficients once. This variation resulted in performance which was very similar to the case in which all the coefficients were iterated. Iterating some other higher order coefficients five times (with no iteration of the first coefficient), however, was found to be quite inferior. Although all the above conclusions are based on the scene Judy, similar conclusions are true for the scene Mike and Nadine. In general, as in pel domain,^{10, 11} the bit rate for Mike and Nadine was much higher than that for Judy. It varied between 170 and 200 kilobits per frame for conditional replenishment in the transform domain, compared to 150 to 175 kilobits per frame for motion-compensated transform coding.

IV. MOTION COMPENSATION WITH TRANSMITTED DISPLACEMENT

In this section, we give results of estimating displacement by a technique proposed by Limb and Murphy¹⁹ and then use it for motion-

compensated prediction. The displacement computation was done in "displacement blocks" with varying sizes, such that the transform block was an exact submultiple of the displacement block in both dimensions. Also, coded values of intensities of the previous frame were used to obviate the need of an additional frame store. Having computed the displacement for the displacement block by the Limb/Murphy algorithm, each transform coefficient within the transform block is predicted by using the displaced coefficient from the previous frame or the nondisplaced coefficient from the previous frame, depending on which was better for the previous coefficient of the same block. In this scheme, there is a tradeoff between the displacement block size and the total number of bits required for a given picture quality. A large displacement block size tends to average all the local variations of the displacement and, consequently, may not result in a good prediction; however, it requires less overhead for transmission of the displacement estimate. On the other hand, a small displacement block requires larger overhead but is potentially superior for displacement estimation in noiseless data. For real scenes, however, the quality of displacement estimation using small blocks might also suffer.

Our simulations used three sizes for the displacement blocks: 16×32 (i.e., 16 lines \times 32 elements in the same field), 8×16 , and 4×8 . These blocks were approximately square, considering the interlace. Only a 2×4 transform block with the cosine transform was used. All the rest of the coder parameters were adjusted to generate pictures of approximately the same quality as before. For both the scenes, without accounting for bits required for transmission of displacement information, a displacement block size of 8×16 did the best in terms of bits per frame. For the scene Judy, displacement blocks of 16×32 and 4×8 resulted in bit rates that were higher by approximately 1000 bits per frame and 2000 bits per frame, respectively. For the scene Mike and Nadine, similar comparisons resulted in about 3000 bits per frame and 5000 bits per frame. Also without accounting for those bits necessary for transmission of displacement information, the 8×16 displacement block resulted in bit rates comparable to those of previous sections with recursive displacement estimation for Judy, but about 5000 to 7000 bits per frame higher for the scene Mike and Nadine. This, however, is a small percentage of the total bits transmitted per frame. As mentioned earlier, the schemes of this section require transmission of displacement information. We did not study any schemes to optimize transmission of this information. Assuming that each D_x and D_y can be specified by 8 bits, we would need 2024, 8096, and 32,384 bits per frame for 16×32 , 8×16 , and 4×8 displacement blocks, respectively. Clearly, considering the overall bit rate, displacement blocks of 8×16 and 16×32 are similar in performance, with a

slight preference for the 16×32 block. The 4×8 displacement block is significantly worse, perhaps due to very noisy displacement estimates. Comparison of the techniques of this section with that of the previous section indicates a slight preference for recursive schemes.

V. CONCLUSIONS

We have developed two schemes for motion compensation in the transform domain. In the first scheme, displacement is estimated recursively from previously transmitted coefficients, whereas the second scheme estimates displacements in a block (generally larger than the transform block itself) using some past and future data, and, therefore, requires separate transmission of displacement information. We found that motion compensation resulted in bit rates which were 20 to 40 percent better than the conventional hybrid interframe coders, which use frame difference prediction. Motion compensation in the pel domain was superior to that in the transform domain for the particular coder structures we investigated. The two methods of displacement estimation were quite similar in performance with a slight preference for the scheme with recursive displacement estimation. None of these comparisons was based on hardware complexity, and it is possible that hardware considerations may change the preferences.

REFERENCES

1. F. W. Mounts, "A Video Encoding System Using Conditional Picture-Element Replenishment," *B.S.T.J.*, 48, No. 7 (September 1969), pp. 2545-2554.
2. B. G. Haskell, F. W. Mounts, and J. C. Candy, "Interframe Coding of Videotelephone Pictures," *Proc. IEEE*, 60, No. 7 (July 1972), pp. 792-800.
3. J. O. Limb, R. F. W. Pease, and K. A. Walsh, "Combining Intraframe and Frame-to-Frame Coding for Television," *B.S.T.J.*, 53, No. 6 (August 1974), pp. 1137-1173.
4. T. Ishiguro, K. Inuma, Y. Iijima, T. Koga, S. Azami, and T. Mune, "Composite Interframe Coding of NTSC Television Signals," 1976 Nat. Telecommun. Conf. Record, 1, Dallas, Texas, November 1976, pp. 6.4-1 to 6.4-5.
5. B. G. Haskell, P. L. Gordon, R. L. Schmidt, and J. V. Scattaglia, "Interframe Coding of 525-Line Monochrome Television at 1.5Mbs," *IEEE Trans. Commun.*, COM-25 (October 1977), pp. 1461-1466.
6. F. Rocca, "Television Bandwidth Compression Utilizing Frame-to-Frame Correlation and Movement Compensation," Symposium on Picture Bandwidth Compression, M.I.T., Cambridge, Mass., 1969; Gordon and Breach, 1972.
7. B. G. Haskell and J. O. Limb, "Predictive Video Encoding Using Measured Subject Velocity," January 1972, U.S. Patent 3, 632, 865.
8. C. Cafforio and F. Rocca, "Methods for Measuring Small Displacements of Television Images," *IEEE Trans. Inform. Theory*, IT-22, No. 5 (September 1976), pp. 573-579.
9. B. G. Haskell, "Entropy Measurements for Nonadaptive and Adaptive, Frame-to-Frame, Linear Predictive Coding of Video Telephone Signals," *B.S.T.J.*, 54, No. 6 (August 1975), pp. 1155-1174.
10. A. N. Netravali and J. D. Robbins, "Motion-Compensated Television Coding: Part I," *B.S.T.J.*, 58, No. 3 (March 1979), pp. 629-668.
11. J. D. Robbins and A. N. Netravali, "Interframe Television Coding Using Movement Compensation," Int. Conf. Commun., Boston, Mass., June 1979.
12. P. A. Wintz, "Transform Picture Coding," *Proc. IEEE*, 60, No. 7 (July 1972), pp. 809-820.

13. C. Reader, "Orthogonal Transform Coding of Still and Moving Pictures," Ph.D. dissertation, The University of Sussex, United Kingdom, 1974.
14. J. A. Roese, W. K. Pratt, and G. S. Robinson, "Interframe Cosine Transform Image Coding," *IEEE Trans. Commun.*, *COM-25*, No. 11 (November 1977), pp. 1329-1339.
15. H. W. Jones, "A Conditional Replenishment Hadamard Video Compressor," *SPIE*, *119*, Applications of Digital Image Processing, 1977, pp. 91-98.
16. S. C. Knauer, "Real-Time Video Compression Algorithm for Hadamard Transform Processing," *Proc. SPIE*, *66* (August 1975), pp. 58-69.
17. A. Habibi, "Hybrid Coding of Pictorial Data," *IEEE Trans. Commun.*, *COM-22*, No. 5 (May 1974), pp. 614-624.
18. J. A. Stuller and A. N. Netravali, "Transform Domain Motion Estimation," *B.S.T.J.*, this issue, pp. 1673-1702.
19. J. O. Limb and J. A. Murphy, "Estimating the Velocity of Moving Images from Television Signals," *Computer Graphics and Image Processing*, *4*, 1975, pp. 311-327.

Contributors to This Issue

Harry W. Adelman, A.A.S., 1970, Warren County Technical School; Bell Laboratories, 1970—. Mr. Adelman is presently involved in the area of digital signal processing which includes digital TASI and digital syllabic companding.

John C. Baumhauer, Jr., B.S. in M.E., 1968, Drexel University; M.S., 1970, and Ph.D., 1973, in Mechanics, Rensselaer Polytechnic Institute; Bell Laboratories, 1973—. Mr. Baumhauer has been engaged in the development and design of electroacoustic telephone transducers. Initially, he did analytical work in the development of an electret transmitter. Later, he had engineering responsibility for granular carbon microphones. He has taught an in-hours course on the operational principles and modeling of electroacoustic devices. More recently, he has completed the design of a piezoelectric ceramic sounder unit, and has been involved with electrodynamic loudspeakers and piezoelectric polymer transducers. Member, Acoustical Society of America, Sigma Xi, Pi Tau Sigma.

George W. Berkstresser, B.S. (Chemical Engineering), 1966, University of Southern California; Ph.D. (Chemical Engineering), 1973, University of Southern California; Tlargini Rubber Tech. Fd., 1966-68; Fractionation Research, 1968; Bell Laboratories, 1973—. From 1973 to 1978, Mr. Berkstresser was involved in the development of optoelectronic displays, optoisolators, and LED sources for optical fiber systems. Presently, he is engaged in epitaxial growth of magnetic bubble garnets. Member, American Chemical Society, American Institute of Chemical Engineers.

Alex M. Brzezinski, A.A.S., 1963, De Vry Technical Institute; B.S. in Physics, 1972, Butler University; Bell Laboratories, 1963—. Mr. Brzezinski has worked in electroacoustic transducer design and devel-

opment. Transducer technologies involved have been granular carbon, electromagnetic, piezoelectric and electret.

F. S. Chen, B.S., 1951, National Taiwan University; M.S.E.E., 1955, Purdue University; Ph.D. 1959, The Ohio State University; Bell Laboratories, 1959—. Mr. Chen has worked in the development of ferrite devices, masers, and optical modulators, and is presently engaged in development of lightwave transmitter subsystem.

Aland K. Chin, B.A., 1972, Brandeis University; M.S., 1975, Cornell University; Ph.D., 1977, Cornell University; Senior Research Engineer, Honeywell Electro-Optics Center, 1977-1978; Bell Laboratories, 1978—. Mr. Chin is involved in the design, processing, and characterization of light-emitting diodes for optical communication systems. Member, American Physical Society, American Association for the Advancement of Science, Phi Beta Kappa.

Yau-Chau Ching B.E. (E.E.), 1966, City College of New York; and Ph.D. (E.E.), 1969, New York University; Bell Laboratories, 1969—. Mr. Ching's research activities include various aspects of digital signal processing. His recent interests are digital speech interpolation and adaptive DPCM encoding of speech signals.

Susan W. Christensen, B.A. (mathematics), 1972, Vassar College; Bell Laboratories, 1973—. Ms. Christensen is a member of the Acoustics Research Department, where she has worked on problems in speech synthesis, digital speech processing, and speech recognition.

David D. Falconer, B.A.Sc., 1962, University of Toronto; S.M., 1963, and Ph.D., 1967, Massachusetts Institute of Technology; post-doctoral research, Royal Institute of Technology, Stockholm, 1966-1967; Bell Laboratories, 1967—. Mr. Falconer has worked on problems in communication theory and high-speed data communication. During 1976-77, he was a visiting professor of electrical engineering at Linköping University, Linköping, Sweden. He presently supervises a group working on digital signal processing for speech bit rate reduction. Member, Tau Beta Pi, Sigma Xi, IEEE.

Ben Gotz, B.E.E., 1966, The City College of New York; M.E.E., 1968, Ph.D., 1971, New York University; Bell Laboratories, 1966-1969,

1971—. Mr. Gotz is presently engaged in studies related to speech coding for bit rate compression.

Nuggehally S. Jayant, B.Sc. (Physics and Mathematics), 1962, Mysore University; B.E., 1965, and Ph.D. (Electrical Communication Engineering), 1970, Indian Institute of Science, Bangalore; Research Associate at Stanford University, 1967-1968; Bell Laboratories, 1968—. Mr. Jayant was a Visiting Scientist at the Indian Institute of Science from January-March 1972 and August-October 1975. He has worked on coding for burst error channels, detection of fading signals, statistical pattern discrimination, spectral analysis, and problems in adaptive quantization and prediction, with special reference to speech signals. He is also editor of an IEEE Reprint Book of *Waveform Quantization and Coding*.

Michael A. Karr, B. S. (E.E.), 1974, Fairleigh-Dickinson University; Bell Laboratories, 1962—. Mr. Karr was involved in ruby and Q-switched Nd:YAG laser development and nonlinear optics. More recently in the field of fiber optics, he was concerned with package design for the Atlanta fiber system experiment. His current work is in the development of fiber optic devices.

Vassilis G. Keramidas, Ph.D. (Solid State Science), 1973, Materials Research Laboratory, Pennsylvania State University; Bell Laboratories, 1973—. Mr. Keramidas has worked on III-V LEDs for displays and optoelectronics, and on ohmic contacts to III-V compound semiconductors. He is currently involved in the crystal growth and characterization of $Ga_{1-x}Al_xAs$ and $Ga_xIn_{1-x}As_yP_{1-y}$ LEDs for lightwave communications. Member, Electrochemical Society, American Association for the Advancement of Science, and American Association for Crystal Growth.

Kurt H. Mueller, Diploma (E.E.) 1961, and Ph.D., 1967, Swiss Federal Institute of Technology; Bell Laboratories, 1969-1978. Mr. Mueller was a supervisor in the Data Communications Laboratory. He has worked on a variety of problems in coding, modulation, signal processing and echo cancellation for high-speed data communication. During 1972-73 he was on leave at the Swiss Federal Institute of Technology where he lectured in information theory and was a member of the executive body of the European Informatics Networks. He is now with Gretag Ltd. in Switzerland. Senior Member IEEE.

Arun N. Netravali, B. Tech. (Honors), 1967, Indian Institute of Technology, Bombay, India; M.S., 1969, and Ph.D. (E.E), 1970, Rice University; Optimal Data Corporation, 1970-1972; Bell Laboratories, 1972-. Mr. Netravali has worked on problems related to filtering, guidance, and control of the space shuttle. At Bell Laboratories, he has worked on various aspects of signal processing, and is presently Head of the Visual Communication Research Department. He is also an adjunct professor of Electrical Engineering at Rutgers University. Member, Tau Beta Pi, Sigma Xi; Senior Member, IEEE.

Carolyn E. Schmidt, B.S. (Mathematics), 1974, Lafayette College; Bell Laboratories, 1974—. Miss Schmidt is a member of the Acoustic Research Department and is currently involved in work on speech communications. Member, Phi Beta Kappa.

Paul W. Shumate, Jr., B.S. (physics), 1963, College of William and Mary; Ph.D. (physics), 1968, University of Virginia; Bell Laboratories, 1969—. Mr. Shumate's first assignments at Bell Laboratories included research on the physical properties of magnetic bubble materials and magnetic bubble memory devices. He transferred to the Integrated Circuit Marketing and Applications Department in 1973, where he studied memory applications for integrated circuits. In 1975 he became Supervisor in the Lightwave Devices and Subsystems Department, where he directs the design and packaging of gallium-arsenide laser transmitters for use in future lightwave communications systems. Member, American Physical Society, Phi Beta Kappa, Sigma Xi, IEEE.

Raymond Steele, B.Sc. (E.E.), 1959, University of Durham, England; Ph.D., 1975, Loughborough University of Technology, England; Bell Laboratories, 1979—. Mr. Steele was a lecturer at the Royal Naval College, Greenwich, London from 1965 to 1968 when he became senior lecturer at Loughborough University of Technology. He has been engaged in source encoding of speech and picture signals and is the author of a book, *Delta Modulation Systems*. He was a consultant to the Acoustics Research Department at Bell Laboratories in the summers of 1975, 1977, and 1978, and is now a member of the Communications Methods Research Department.

John Stuller, B.S.E.E., 1963, Massachusetts Institute of Technology; M.S.E.E.; 1967, University of Southern California; Ph.D., 1971,

University of Connecticut; Northrop Nortronics Corp., 1963-1965; TRW Systems Group, 1965-1968; Perkin Elmer Corp. 1970-1971; Department of Electrical Engineering, University of New Brunswick, Fredericton, New Brunswick, Canada, 1971-1974; Bell Laboratories, 1977-1979. At Bell Laboratories, Mr. Stuller was engaged in analytical and simulation studies of both intraframe and interframe video coders. His primary area of interest is in the statistical theory of communications and its application.

Papers by Bell Laboratories Authors

BIOLOGY

Fluorescence X-Ray Absorption Studies of Rubredoxin, Its Model Compounds. R. G. Shulman, P. Eisenberger, B. K. Teo, B. M. Kincaid, and G. S. Brown, *Mol. Biol.*, 124 (1978), pp. 305-321.

High Resolution Carbon-13 Nuclear Magnetic Resonance Studies of Glucose Metabolism in E Coli. K. Ugurbil, T. R. Brown, J. A. Der Hollander, P. Glynn, and R. G. Shulman, *Proc. Nat. Acad. Sci.*, 75 (Aug. 1, 1978), pp. 3742-3746.

High Resolution Phosphorus-31 Nuclear Magnetic Resonance Study of Rat Liver Mitochondria. S. Ogawa, H. Rottenberg, T. R. Brown, R. G. Shulman, C. L. Castillo, and P. Glynn, *Proc. Nat. Acad. Sci.*, 75 (1978), pp. 1796-1800.

On the Measurement of PH in E Coli by Phosphorus-31 Nuclear Magnetic Resonance. S. Ogawa, R. G. Shulman, P. Glynn, T. Yamane, and G. Navone, *Biochem. Biophys. Acta*, 502 (1978), pp. 45-50.

Phosphate Metabolites in Lymphoid, Friend Erythroleukemic and Hela Cells Observed by High Resolution Phosphorus-31 Nuclear Magnetic Resonance. G. Navon, R. Navon, R. G. Shulman, and T. Yamane, *Proc. Nat. Acad. Sci.*, 75 (February 1978), pp. 891-895.

Response of the Bohr Group Salt Bridges to Ligation of the T-State of Haemoglobin Kansas. J. V. Kilmartin, N. L. Anderson, and S. Ogawa, *J. Mol. Biol.*, 123 (1978), pp. 71-87.

X-Ray Absorption Spectroscopy of Biological Molecules. R. G. Shulman, P. Eisenberger, and B. M. Kincaid, *Ann. Rev. Biophys. Bioeng.* 7 (1978), pp. 559-578.

CHEMISTRY

Analysis of the Electron Spin Echo Decay Envelope for Neodymium (3+): ATP Complexes. T. Shimizu, W. B. Mims, J. Peisach, and J. L. Davis, *J. Chem. Phys.*, March 1979.

Analysis of Nonlinear Stress Relaxation in Polymeric Glasses. S. Matsuoka, S. S. Bearder, and H. E. Earn, *Poly. Eng. Sci.*, 18, No. 4 (1978), pp. 1074-1080.

Analysis of Tin Nickel Electroplate by Secondary Ion Mass Spectrometry, Ion Scattering Spectrometry, and Rutherford Backscattering. R. Schubert, *J. Electrochem. Soc.*, 125, No. 8 (August 1978), pp. 1215-1218.

Azulene Revisited: Picosecond Decay of the Sulfur State in the Gas Phase and in Solution. D. Huppert, J. Jortner, and P. M. Rentzepis, *Israeli J. Chem.*, Volume on Excited States, 1978.

Consideration of Processing Conditions on the Properties of Solid Polymers. S. Matsuoka, *Poly. Eng. Sci.*, 18, No. 4 (1978), p. 264.

The Deuteron Quadrupole Coupling Constant in Carbon Deuterium (3) Fluoride. L. O. Snyder, *J. Chem. Phys.*, 68, No. 1 (1978), pp. 340-341.

Deuteron Quadrupole Coupling in Molecules. L. O. Snyder, *J. Chem. Phys.*, 68, No. 1 (1978), pp. 291-294.

The Linear Electric Field Effect in Stellacyanin, Azurin, and in Some Simple Copper Model Compounds. J. Peisach and W. B. Mims, *Eur. J. Biochem.*, 84 (1978), pp. 207-214.

Linear and Nonlinear Optical Properties of Sodium (2) Antimony Pentafluoride. J. G. Bergman, D. S. Chemla, R. Fourcose, and G. Mascherpa, *J. Solid State Chem.*, 23, No. 1-2 (1978), pp. 187-190.

Molecular Orbital Theory of the Compton Profile of Fluorine (2). L. O. Snyder and T. A. Weber, *J. Chem. Phys.*, **68**, No. 6 (March 15, 1978), pp. 2974-2979.

Molecular Recognition and Self-Organization in Fluorinated Hydrocarbons. F. J. Stillinger and Z. Wasserman, *J. Phys. Chem.*, **82**, No. 8 (April 20, 1978), pp. 929-940.

The Nuclear Modulation Effect in Electron Spin Echoes for Complexes of Copper (2+) and Imidazole with Nitrogen-14 and Nitrogen-15. W. B. Mims and J. Peisach, *J. Chem. Phys.*, **69** (1978), pp. 4921-4930.

Polarization Model Representation of Hydrogen Fluoride for Use in Gas and Condensed Phase Studies. F. H. Stillinger, *Int. J. Quantum Chem.*, **14** (1978), pp. 649-657.

Proton Transfer Reactions and Kinetics in Water. F. H. Stillinger in *Theoretical Chemistry Advanced Perspectives*, Vol. 3, H. Eyring and D. Henderson, eds., N.Y.: Academic Press, 1978, pp. 177-234.

Quantum Chemical Calculations to Model Borate Glass Electronic Structure and Properties. L. C. Snyder in *Borate Glasses: Structure, Properties, Applications*, L. D. Pye, V. D. Frechette, and N. J. Kreidl, eds., New York: Plenum, 1978, pp. 151-165.

The Reaction of Poly(vinyl Chloride) with Thiols. W. H. Starnes, I. M. Plitz, D. C. Hische, D. J. Freed, F. C. Schilling, and M. L. Schilling, *Amer. Chem. Soc. Polym. Preprints*, **19**, No. 1 (March 1978), pp. 623-628.

Recent Fundamental Developments in the Chemistry of Poly(vinyl Chloride) Degradation and Stabilization. W. H. Starnes, *ACS Advan. Chem. Ser.*, **169** (1978), pp. 309-323.

Reductive Dehalogenation with Tri(N-Butyl) Tin Hydride: A Powerful New Technique for Use in Poly(vinyl chloride) Microstructure Investigations. W. H. Starnes, R. L. Hartless, F. C. Schilling, and F. A. Bovey, *ACS Advan. Chem. Ser.*, **169** (1978), pp. 324-332.

Relaxation Processes in Matrix Isolated Carbon (3). V. E. Bondybey and J. H. English, *J. Chem. Phys.*, **68**, No. 10 (May 15, 1978), pp. 4841-4843.

Revised Central Force Potentials for Water. F. H. Stillinger and A. Rahman, *J. Chem. Phys.*, **68**, No. 2 (January 15, 1978), pp. 666-670.

Simulation of N-Butane Using a Skeletal Alkane Model. T. A. Weber, *J. Chem. Phys.*, **69** (September 15, 1978), pp. 2347-2354.

Stabilization of Poly(vinyl Chloride) by Thiols: A Mechanistic Study. W. H. Starnes, I. M. Plitz, D. D. Hische, D. J. Freed, F. C. Schilling, and M. L. Schilling, *Macromolecules*, **11** (Mar/Apr 1978), pp. 373-382.

Stochastic Classical Trajectory Approach to Relaxation Phenomena: I. Vibrational Relaxation of Impurity Molecules in Solid Matrices. M. Shugard, J. O. Tully, and A. Nitzan, *J. Chem. Phys.*, **69**, No. 1 (July 1, 1978), pp. 336-345.

Study of Melting and Freezing in the Gaussian Core Model by Molecular Dynamic Simulation. F. H. Stillinger and T. A. Weber, *J. Chem. Phys.*, **68**, No. 8 (April 15, 1978), pp. 3837-3844.

Subpicosecond Vibrational Relaxation in Calcium (2) in Rare Gas Solids. V. E. Bondybey and C. Albiston, *J. Chem. Phys.*, **63**, No. 7 (April 1, 1978), pp. 3172-3176.

Tantalum Ion Transport Number During the Anodic Oxidation of Beta-Tantalum Films. N. Schwartz and W. M. Augustyniak, *J. Electrochem. Soc.*, **125** (1978), p. 1812.

Tapered Windows in Phosphorus Doped Silicon Dioxide by Ion Implantation. J. C. North, T. G. McGahan, D. W. Rice, and A. C. Adams, *IEEE Trans. Electron. Dev.*, **25**, No 7 (July 1978), pp. 809-812.

Theory and Practice on Nonlinear Behaviour of Polymeric Solids, II. Programmed Calculation on Yield, Creep, and Notched Impact Strength. S. Matsumika and J. T. Ryan, *Soc. Plast. Eng., Tech. Conf. Preprint*, 1978, pp. 222-226.

COMPUTING

Analog Read Only Memory Using Gadolinium Molybdate. R. A. Lemons, M. M. Geary, L. A. Coloren, and H. G. Mattes, *Appl. Phys. Lett.*, **33**, No. 5 (September 1, 1978), p. 373-375.

An Assessment of the Technology for Data and Program Conversion. M. M.

Kaplan and N. H. Goguen, Proc. ACM/NBS Workshop Nat. Comput. Conf., 1978, pp. 887-907.

Coding Theorems for Individual Sequences. J. Ziv, IEEE Trans. Inform. Theory, 24, No. 4 (July 1978), pp. 405-412.

Computer Simulation of Charge Dynamics in Electron Irradiated Polymer Foils. D. A. Berkley, Proc. Elec. Insul. Dielec. Nat. Acad. Sci. Ann. Conf., Washington, D.C., 1978, pp. 26-32.

MAGIN—A Finite Element Computer Program for 2-D and Axisymmetric Magnetic Field Problems. E. Gauncial and S. Das Gupta, J. Inst. Eng. India, Electron. Telecommun. Eng. Div., 58, No. 3 (December 1977), pp. 166-170.

Mathematical Programming Approaches to System Partitioning. J. L. Uhrig, IEEE Trans. Syst. Man, Cybern., 8, No. 7 (July 1978), pp. 540-548.

A Microprocessor-Based Time Division Multiplexer. J. W. Smith, Proc. Electron. Nat. Conf., 32, October 1978.

Microprogrammed Spectrum Analysis. T. J. Mulrooney, Proc. 11th Microprogramming Annual Workshop, November 1978, pp. 90-99.

MVS/VM (1980-1985): A User Perspective. D. M. Becker, J. A. Jonekait, and J. Michelin, Proc. SHARE 5, August 1978, Vol. 1, pp. 124-142.

A Relative Entropy-Like Measure for System Organization. J. A. Chaikin and R. A. Orchard, Proc. 14th Comput. Perform. Eval. User Group Meeting, October 1978, p. 259-262.

Software Development for Microprocessors—A Case Study. F. A. Salomon, Proc. Comput. Software Appl. Conf., 1978, pp. 110-113.

A Workshop Report: The New and the Not-So-New. M. F. Siana and G. G. Dumas, Computer, 11, No. 3 (March 1978), pp. 47-51.

ELECTRICAL AND ELECTRONIC ENGINEERING

Adaptive Digital Radio Receivers for Combating Frequency Selective Fading. L. J. Greenstein and D. Vitello, IEEE Trans. Commun., April 1979.

Analog Viterbi Decoding for High Speed Digital Satellite Channels. A. S. Acampora and R. Gilmore, Proc. Nat. Telecommun. Conf. Rec., December 1977, pp. 34: 6/1-5.

Analysis of an Interference-Fit Pin Connection. R. P. Goel, Proc. 28th Electron. Compon. Conf. (April 1978), pp. 116-120.

A Comparative Study of Several General Convergence Conditions for Algorithms Modeled by Point-to-Set Maps. S. Tishyadhigama, E. Polak, and R. Klessig, Electron. Res. Lab. Memo, Univ. Calif., Berkeley, November 1977.

Control of Limit Cycles Oscillations in Second-Order Recursive Digital Filters Using Constrained Random Quantization. V. B. Lawrence and K. V. Mina, IEEE Trans. Acoust. Speech Signal Process., 26, No. 2 (April 1978), pp. 127-134.

A Digital Codec Simulation Facility. W. R. Daumer, IEEE Trans. Commun., 26, No. 5 (May 1978), pp. 665-669.

Electronic Micropositioning with Ferroelastic-Ferroelectrics. R. A. Lemons and L. A. Coloren, Rev. Sci. Instrum., 49, No. 12 (December 1978), pp. 1650-1652.

Envelope Fluctuation Statistics of Filtered PSK and Other Digital Modulations. L. J. Greenstein and P. J. Fitzgerald, IEEE Trans. Commun., April 1979.

From Analog to Digital—New Trends in Signaling and Switching. J. S. Ryan, Proc. ICC Conf. Rec. (1978), p.13.2.1.

High-Bit Rate Optical Fiber Data Links Using Indium Gallium Arsenide Phosphide LEDs Emitting at 1.3 Micron. A. G. Dentai, W. M. Muska, L. I. Tingye, and T. P. Lee, Proc. OSA/IEEE Laser Electroopt. Sys. Conf., February 1978, pp. 28, 30.

Light Emitting Diodes as Receivers with Avalanche Gain. W. M. Hubbard and H. E. Kehlenbeck, Electron. Lett., 14, No. 17 (August 17, 1978), pp. 553-554.

Maximum Likelihood Decoding of Binary Convolutional Codes on Band-Limited Satellite Channels. A. S. Acampora, IEEE Trans. Commun., 26, No. 6 (June 1978), pp. 766-776.

Measurement of Single-Mode Fiber Core-Cladding Concentricity. A. R. Tynes, R. M. Derosier, and W. G. French, Electron. Lett., 14 (February 16, 1978), pp. 113-115.

A Microprocessor Log PCM/ADPCM Code Converter. J. J. Dubnowski, *IEEE Trans. Commun.*, 26, No.5 (May 1978), pp. 660-664.

A Multipath Fading Channel Model for Terrestrial Digital Radio. L. J. Greenstein, *IEEE Trans. Commun.*, 26, No. 8 (August 1978), pp. 1247-1250.

An Optically Triggered Double Heterostructure Linear Bilateral Phototransistor. S. Knight, L. R. Dawson, V. G. Keramidis, and M. G. Spencer, *Proc. Int. Electron. Dev. Meeting* (December 1977), p. 472.

Optimal, Causal, Simultaneous Detection and Estimation of Random Signal Fields in Gaussian Noise Field. T. T. Kadota, *IEEE Trans. Inform. Theory*, 24, No. 3 (May 1978), pp. 297-308.

SEAT—Simulated Environment Accelerated Test. N. Schwartz and D. D. Bacon, *J. Electrochem. Soc.*, 125 (1978), p. 1495.

Simple Switching Control Method Changes Power Converter into a Current Source. C. W. Deusch, *Proc. IEEE PESC 1978 Rec.* (July 1978), p. 300-6.

A Stability RC Circuit Using High Nitrogen Doped Tantalum. O. J. Duff, G. J. Koerckel, R. A. DeLuca, E. H. Mayer, and W. Worobey, *Proc. 28th Electron. Compon. Conf.* (April 1978), pp. 229-233.

Variable Delay Using Acoustic Bulk Wave Reflection from a Ferroelastic-Ferroelectric Domain Wall. R. A. Lemons and L. A. Coloren, *Proc. IEEE Ultrasonics Symp.*, 1978, pp. 188-192.

MATERIALS SCIENCE

Applications of Ion Implantation to Magnetic Bubble Devices. J. C. North, R. Wolfe, and T. J. Nelson, *J. Vacuum Sci. Technol.*, 15 (September/October 1978), pp. 1675-1684.

Calcium Orthovanadate—A New High Temperature Ferroelectric. A. M. Glass, S. C. Abrahams, A. A. Ballman, and G. Loiacono, *Ferroelectrics*, 17 (April 1978), pp. 579-582.

Cleaning Photomasks with Scrubbers and High-Pressure Water. W. H. Kroeck, R. F. Doll, and E. L. Stokes, *Circuits Mfg.* (June 1978), pp. 16-21.

Copper Deactivators for Polyolefin Insulation. M. G. Chan, H. M. Gilroy, L. Johnson, and W. M. Martin, *Proc. 27th Wire Cable Int. Symp.*, November 1978, pp. 99-106.

Depth Profiles of Sodium and Calcium in Glasses—a Comparison of Secondary Ion Mass Analysis and Auger Spectrometry. D. L. Malm, M. J. Vasilie, F. J. Padden, D. B. Dove, and O. G. Pantano, *J. Vacuum Sci. Technol.*, 15, No. 1 (January/February 1978), pp. 35-38.

The Design and Optimization of Directly Heated Lanthanum Hexaboride Cathode Assemblies for Electron-Beam Instruments. P. H. Schmidt, L. D. Longinotti, D. C. Joy, S. D. Ferris, H. J. Leamy, and Z. Fisk, *J. Vacuum Sci. Technol.*, 15, No. 4 (July/August 1978), pp. 1554-1560.

Diffusion Effects of the Heat Treatment of Gold-Silver on Iron-Nickel Dry Sealed Reeds. C. A. Haque, *Proc. 9th Int. Elec. Contact Phenom. Conf.* (1978), pp. 605-609.

Diurnal Variation of Cosmic Ray Intensity, I: Two Approaches to the Study. J. W. Sari, D. Venkatesan, and L. J. Lanzerotti, *J. Geophys. Res.*, 83 (1978), pp. 5139-5150.

High Temperature Resistive Phase Transition in A15 High Temperature Superconductors. C. W. Chu, C. Y. Huang, P. H. Schmidt, and K. Sugawara in *Superconductivity in d- and f-Band Metals*, D. H. Douglass, ed., New York: Plenum, 1976, p. 453.

High Temperature Superconductivity in Copper Chlorine. E. I. Blount and J. C. Phillips, *J. Less Common Metals*, 62 (1978), pp. 457-461.

Impurity-Induced Exaggerated Grain Growth in Mn-Zn Ferrites. M. F. Yan and D. W. Johnson, *J. Amer. Ceram. Soc.*, 61, No. 7-8 (1978), pp. 342-349.

Interaction of Nitrogen Oxide and Copper at Various Relative Humidities. R. Schubert, *J. Electrochem. Soc.*, 125, No. 7 (July 1978), pp. 1114-1116.

Investigation of Titanium-Platinum Diffusion Barrier for Gold Beam Leads on Aluminum. S. P. Murarka, H. J. Leninstein, I. A. Blech, T. T. Sheng, and M. H. Read, *J. Electrochem. Soc.*, 125, No. 1 (1978), pp. 156-162.

Low Work Function Electron Emitter Hexaborides. P. H. Schmidt and D. C.

Joy, J. Vacuum Sci. Technol., 15, No. 6 (November/December 1978), p. 1309.

Measurement of MIS Capacitors with Oxygen-doped Aluminum (X) Gallium (1-X) Arsenic Insulating Layers. H. G. Casey, A. Y. Cho, D. V. Lang, and E. H. Nicollian, J. Vacuum Technol., 15 (1978), p. 1408.

Photograph of a Giant Structural Distortion in a Quasi-One-Dimensional Organic Crystal. J. B. Mock, G. A. Thomas, D. E. Moncton, and M. Kaplan, J. Phys. Chem. Solids, 39, No. 3 (1978), pp. 253-254.

Potential Problem Areas in Beta Backscatter Measurement Technology. R. R. Buckley, Proc. 1st Amer. Electroplating Soc. Test. Surf. Finish, New York City, March 1978, pp. 1-7; Proc. 65th Assoc. Tech. Ann. Conf., Washington, D. C., June 1978, pp. 3-10.

Predicting Stiffness, Creep, and Fracture Characteristics of Cable Jackets from Material Properties. G. M. Yanizeski, E. D. Nelson, and C. J. Alciscic, 25th Wire Cable Symp., Cherry Hill, N.J., November 1976.

Press Fit Pins in Printed Circuit Boards, Third, Fourth, Fifth, and Sixth Test Series. P. J. Tamburro, Proc. 10th Ann. Connect. Symp., October 1977, pp. 1-20.

Pulsed EPR Studies of Copper in Some Binary Oxide Glasses. W. B. Mims, G. E. Peterson, and C. R. Kurkjian, Phys. Chem. Glass, 19, No. 1 (February 1978), pp. 14-18.

Removal of Particulate Contamination from Hard Surface Photomasks by Various Cleaning Techniques. W. H. Kroeck, R. F. Doll, and E. L. Stokes, Proc. Kodak Microelectron. Sem., May 1978, pp. 26-32.

Silicon Monoxide Anti-Reflection Coatings for INP/CDS Solar Cells. D. H. Olson, J. Vac. Sci. Technol., 15, No. 1 (January/February 1978), pp. 130-132.

Sealed Extrusion—A Coating Method for Photoprinting UV Curable Liquid Resin Films. G. B. Fefferman and T. V. Lake, 4th Radiation Curing Conf. Tech. Paper, Soc. Manu. Eng., pp. 1-18.

Simple Pressurized Chamber for Liquid Encapsulated Czochralski Crystal Growth of Indium Phosphide and Several Chalcopyrites. E. Buehler, J. Cryst. Growth, 43, No. 5 (June 1978), pp. 584-588.

Sintering of High Density Ferrites. M. F. Yan and D. W. Johnson, in *Proceedings of Crystalline Ceramics*, H. Palmour, R. F. Davis, and T. M. Hare, eds., New York: Plenum, 1978, pp. 393-402.

A Stencil Technique for the Preparation of Thin Film Josephson Devices. L. N. Dunkleberger, J. Vacuum Sci. Tech., 15, No. 1 (January/February 1978), pp. 88-90.

MATHEMATICS AND STATISTICS

Discrete Balancing Games. J. C. Lagarias, Bull. Inst. Math. Acad., 5 (1977), pp. 363-373.

Signatures of Units and Congruences (Mod 4) in Certain Real Quadratic Fields. J. C. Lagarias, J. Reine Angew. Math., 301 (1978), pp. 142;146.

PHYSICS

Analysis and Operation of UHF Squid with Tapped Tank Circuit. L. D. Jackel, J. H. W. Stoeckenius, and R. W. Epworth, J. Appl. Phys., 49, No. 1 (January 1978), pp. 471-472.

Anisotropy of Thermionic Electron Emission Values for Lanthanum Hexaboride Single-Crystal Emitter Cathodes. P. H. Schmidt, D. C. Joy, L. D. Longinotti, H. J. Leamy, S. O. Ferris, and Z. Fisk, Appl. Phys. Lett., 28 (1976), pp. 400-401.

Anomalous Dielectric Absorption in Silicone Dioxide-Based Glasses. M. A. Bosch, in *The Physics of SiO₂ and Its Interfaces*, S. T. Pantelides, ed., New York: Pergamon, 1978, pp. 294-298.

Anomalous Transport Produced by Kinetic Alfvén Wave Turbulence. A. Hasegawa and K. Mima, J. Geophys. Res., 83, No. A3 (March 1, 1978), pp. 117-123.

Association of ULF Magnetic Variations and Changes in Ionospheric Conductivity During Substorms. L. J. Lanzerotti, C. G. MacLennan, and C. Evans, J. Geophys. Res., 83 (1978), pp. 2525-2532.

Carrier-Lattice Coupling in Titanium Sulfide and Titanium Selenide. J. A. Wilson, Proc. Latt. Dyn. Int. Conf., 1978.

- A Conductor Crossing Problem in Magnetic Bubble Memories.** W. Strauss, *J. Appl. Phys.*, 49, No. 3, Part 2 (March 1978), pp. 1897-1899.
- Crystallographic Factors Affecting the Structure of Polymeric Spherulites. Part I: Morphology of Directionally Solidified Polyamides.** A. J. Lovinger, *J. Appl. Phys.*, 49 (1978), pp. 5003-5013.
- Crystallographic Factors Affecting the Structure of Polymeric Spherulites. Part II: X-Ray Diffraction Analysis of Directionally Solidified Polyamides and General Conclusions.** A. J. Lovinger, *J. Appl. Phys.*, 49 (1978), pp. 5014-5028.
- Demountable Helium-3 Melting Curve Thermometer.** L. R. Corruccini, K. R. Mountfield, and W. C. Sprenger. *Rev. Sci. Instrum.*, 49, No. 3 (March 1978), pp. 314-318.
- Diagnostics with Series-Connected Josephson Tunnel Junctions.** A. F. Hebard and R. H. Eick, *J. Appl. Phys.*, 49, No. 1 (January 1978), pp. 338-343.
- Diffusion in a Palladium Copper-Silicon Metallic Glass.** H. S. Chen, L. C. Kimerling, J. M. Poate, and W. L. Brown. *Appl. Phys. Lett.*, 32, No. 8 (April 15, 1978), pp. 461-463.
- Direct Measurement of Quasiparticle Lifetime Broadening in a Strong Coupled Superconductor.** R. C. Dynes, V. Narayanamurti, and J. P. Garno, *Phys. Rev. Lett.*, 41 (1978), pp. 1509.
- Dynamics of the Charge Density Wave: I. Impurity Pinning in a Single Chain.** H. Fukuyama and P. A. Lee, *Phys. Rev. B*, 17, No. 2 (January 15, 1978), pp. 535-541.
- On the Effect of a Partial Sink in Binary Diffusion in Thin Films.** H. G. Tompkins, *J. Appl. Phys.*, 49, No. 1 (January 1978), pp. 223-228.
- The Electron-Hole Liquid in Semiconductors: Experimental Aspects.** J. C. Hensel, T. G. Phillips, and G. A. Thomas, *Solid State Phys.*, 32 (1978), pp. 87-314.
- Electrons in Surface States on Liquid Helium.** C. C. Grimes, *Surf. Sci.*, 73, No. 1 (September 19-22 1977), pp. 379-395.
- Electrons on the Vacuum Surface of Copper Oxide and the Screening of Patch Fields.** A. R. Hutson, *Phys. Rev. B*, 17, No. 4 (February 15, 1978), pp. 1934-1939.
- Energy Field Dependence of a Uniaxial Anisotropy Energy in Lithium Ferrite.** G. T. Rado, J. M. Ferrari, and J. P. Remeika, *J. Appl. Phys.*, 49, No. 3 (March 1978), p. 1953.
- Epitaxial Growth of Niobium (3) Germanium Niobium (3) Rhodium and Niobium (3) Iridium.** A. H. Dayen, T. H. Geballe, R. B. Zubeck, A. B. Hallak, and G. W. Hull, *J. Phys. Chem. Solids*, 39, No. 5 (1978), pp. 529-538.
- An Estimation of the Quantum Yield in the Cathodoluminescence Process by Low Energy Hydrogen (+) Ion Bombardment.** I. S. T. Tsong, A. Corredor, W. B. White, N. H. Tolk, and J. S. Kraus, *J. Electrochem. Soc.*, 125 (1978), p. 2015.
- Evidence for Chemical Annealing Effects in Indium Oxide Tunnel Junction Barriers.** A. F. Hebard, J. R. Arthur, and D. L. Allara, *J. Appl. Phys.*, 49, No. 12 (1978), pp. 3039-3044.
- Exciton Contribution to the Reflection Spectrum at High Excitation Intensities in Cadmium Sulfide.** R. F. Leheny, J. Shah, and G. O. Chiang, *Solid State Commun.*, 25, No. 9 (March 1978), pp. 621-624.
- Exciton Structure of Crystalline Potassium Cyanide.** M. A. Bosch and G. Zimofen, *Phys. Rev. Lett.*, 41, No. 8 (August 21, 1978), pp. 590-593.
- Experimental Ferroelectric Domain Wall Motion Devices.** J. M. Geary, *Appl. Phys. Lett.*, 32, No. 8 (April 15, 1978), pp. 455-457.
- Fast Time Heat Capacity in Amorphous Silicon Dioxide Using Heat Pulse Propagation.** R. B. Kummer, R. C. Dynes, and V. Narayanamurti, *Phys. Rev. Lett.*, 40, No. 18 (May 1, 1978), pp. 1187-1190.
- Fermi Surface Excitations in X-Ray Photoemission Lineshapes from Metals.** G. K. Wertheim and P. H. Citrin, Photoemissions in Solids, I, in *Topics in Applied Physics*, Vol. 26, M. Cardona and L. Ley, eds., New York: Springer-Verlag, 1978.
- Forced Magnetostriction of Dysprosium Aluminum Garnet (DAG).** J. F. Dillon, E. M. Gyorgy, E. I. Blount, E. Y. Chen, H. A. Brooks, and L. G. Van Uitert, *J. Appl. Phys.*, 49, No. 3, Part 2 (March 1978), pp. 1371-1373.
- Germanium-Tin Barrier Josephson Tunnel Junctions.** E. L. Hu, L. D. Jackel, A. R. Strand, R. W. Epworth, R. F. Lucey, C. A. Zogg, and E. Gornik, *Appl. Phys. Lett.*, 32, No. 9 (May 1, 1978), pp. 584-586.

- Hardness and Bonding in A-15 Superconducting Compounds.** G. Y. Chin, J. H. Wernick, T. H. Begalle, S. Mahajan, and S. Nakahara, *Appl. Phys. Lett.*, **33**, No. 1 (July 1, 1978), pp. 103-105.
- A High Resolution Electron Microscopy Study of the Silicon-Silicon Dioxide Interface.** O. L. Krivanek, T. T. Sheng, and D. C. Tsui, *Appl. Phys. Lett.*, **32**, No. 7 (April 1, 1978), pp. 437-439.
- High Temperature Resistive Phase Transition in Al5 High Temperature Superconductors.** C. W. Chu, C. Y. Huang, P. H. Schmidt, and K. Sugawara, in *Superconductivity in d- and f-Band Metals*, D. H. Douglass, ed., New York: Plenum, 1976, p. 453.
- Impact Strength of Polymers: 2. The Effect of Cold Working and Residual Stress.** B. S. Thakkar and L. J. Broutman, *Proc. 36th Ann. Tech. Conf., Soc. Plast. Engl.*, Vol. XXIV, April 1978, pp. 295-303. Inelastic Electron Scattering in Nickel. A. E. Meixner, M. Schluter, P. M. Platzman, and G. S. Brown, *Solid State Commun.*, **27** (1978), pp. 1255-1257.
- Loss in Cleaved Titanium-Diffused Lithium Niobate Waveguides.** I. P. Kaminow and L. W. Stulz, *Appl. Phys. Lett.*, **33**, No. 1 (July 1, 1978), pp. 62-64.
- The Magnetic Behavior of a Spin-1/2 Amorphous Antiferromagnet.** R. B. Kummer, R. E. Walstedt, S. Geschwind, V. Narayanamurti, and G. E. Devlin, *Phys. Rev. Lett.*, **40**, No. 16 (April 17, 1978), pp. 1098-1100.
- Magneto-optical Properties of Magnetic Garnets.** J. F. Dillon in *Physics of Magnetic Garnets*, A. Paoletti, ed., Amsterdam: North Holland, 1978, pp. 379-415.
- Measurements of the Absorption Edge in Fused Silica.** I. P. Kaminow, B. G. Bagley, and C. G. Olson, *Appl. Phys. Lett.*, **32**, No. 2 (January 15, 1978), pp. 98-99.
- A Memory Device Utilizing the Storage of Abrikosov Vortices at an Array of Pinning Sites in a Superconducting Film.** A. E. Hebard and A. T. Fiory, *Proc. 44th Amer. Inst. Phys. Conf.*, pp. 465-469.
- Modeling the Contrasting Semimetallic Characters of Titanium Disulfide and Titanium Diselenide.** J. A. Wilson, *Phys. Stat. Sol. B*, **86**, No. 1 (March 1, 1978), pp. 11-36.
- Molecular Dynamics Studies of a One-Dimensional Complex Order Parameter Model.** T. R. Koehler and P. A. Lee, *Phys. Rev.*, **B16** (1977), p. 5203.
- Molecular SCF Calculations to Model (III) Surface States and Relaxation of Diamond.** L. C. Snyder and Z. Wasserman, *Surf. Sci.*, **31**, No. 2 (February 1978), pp. 407-413.
- New Ductive Chromium-Cobalt Iron Permanent Magnet Alloys for Telephone Receiver Applications.** G. Y. Chin, J. T. Plewes, and B. C. Wonsiewicz, *J. Appl. Phys.*, **49**, No. 3, Part 2 (March 1978), pp. 2046-2048.
- Nonlinear Self-Modulation of Ion-Acoustic Waves.** H. Ikezi, K. Schwarzenegger, A. L. Simons, Y. Ohsawa, and T. Kaminura, *Phys. Fluids*, **21**, No. 2 (February 1978), pp. 239-245.
- Optical Birefringence of Ultrathin Aluminum (X) Gallium (1-X) Arsenide Gallium Arsenide Multilayer Heterostructures.** J. R. Van Der Ziel and A. C. Gossard, *J. Appl. Phys.*, **49**, No. 5 (May 1978), pp. 2919-2921.
- Optically Pumped Taper Coupled Gallium Arsenide-Aluminum (X) Gallium (1-X) Arsenide Laser with a Second Order Bragg Reflector.** R. C. Miller, W. A. Nordland, R. A. Logan, and L. F. Johnson, *J. Appl. Phys.*, **49**, No. 2 (February 1978), pp. 539-542.
- Optical Properties of Expanded Fluid Mercury.** H. Ikezi, A. L. Schwarzenegger, A. L. Simons, A. L. Passner, and S. L. McCall, *Phys. Rev. B*, **18**, No. 6 (September 1978), pp. 2494-2499.
- Optical Studies of the Magnetic Phase Diagram of Iron Chloride (2).** J. F. Dillon, Chen E. Yi, and H. J. Guggenheim, *Phys. Rev. B*, **18** (1978), pp. 377-387.
- Propagation of Nonlinear Waves on an Electron-Charged Surface of Liquid Helium.** K. Mina and H. Ikezi, *Phys. Rev. B*, **17**, No. 9 (May 1, 1978), pp. 3567-3575.
- Questions Concerning the Form Taken by the Cow and the Accompanying PSD in 2H-Tantalum Selenide and Closely Related Materials.** J. A. Wilson, *Phys. Rev. B*, **17**, No. 10 (May 15, 1978), pp. 3880-3898.
- Removal of Patriculate Contamination from Hard Surface Photomasks by Various Cleaning Techniques.** W. H. Kroeck, R. F. Doll, and E. L. Stokes, *Proc. Kodak Microelectron. Sem.*, May 1978, pp. 26-32.
- The Role of the Thermal Relaxation Time in Pulse Nonlinearity Measure-**

ments. A. F. Hebard and W. M. Steverson, *J. Appl. Phys.* 49, No. 10 (1978), pp. 5250-5255.

Screening of 3D Holes in the Rare Earths. G. K. Wertheim and M. Campagna, *Solid State Commun.*, 26, No. 8 (May 1978), pp. 553-556.

Search for Cluster Effects in X-Ray Production by Fast Hydrogen Molecules. A. Lurio, H. H. Andersen, and L. C. Feldman, *Phys. Rev. A*, 17, No. 1 (January 1978), pp. 90-92.

Sensitivity of the TC of Niobium (3) Germanium to Hydrogen Content. W. A. Lanford, P. H. Schmidt, J. M. Rowell, J. M. Poate, and R. C. Dynes, *Appl. Phys. Lett.*, 32, No. 5 (March 1978), pp. 339-341.

Shift in Cyclotron Resonance of Electrons on Liquid Helium Surfaces. A. F. Cheng and P. M. Platzman, *Solid State Commun.*, 25, No. 10 (March 1978), pp. 813-815.

Significance of the Channeling Surface Peak in Thin Film Analysis. R. L. Kauffman, L. C. Feldman, P. J. Silverman, and R. A. Zuhr, *Appl. Phys. Lett.*, 32, No. 2 (January 15, 1978), pp. 93-94.

Spin Polarization of Electrons Field Emitted from Single-Crystal Iron Surfaces. M. Landolf and Y. Yafet, *Phys. Rev. Lett.*, 40, No. 21 (May 22, 1978), pp. 1401-1403.

Stabilization of Drift Cyclotron Loss-Cone Mode by Low Frequency Density Fluctuations. A. Hasegawa, *Phys. Rev. Lett.*, 40, No. 14 (April 3, 1978), pp. 938-941.

Stochastic Controllability of Linear Discrete Systems with Multiplicative Noise. S. E. Harris, *Int. J. Contr.*, 27, No. 2 (February 1978), pp. 213-227.

Structure of the Electron Gas at the Surface of Liquid Helium. C. C. Grimes, *Journal de Physique*, 39, No. C6 (August 1978), pp. 1352-1357.

The Structure of Sputtered Superconducting Niobium (3) Germanium Films Revealed by X-Ray Diffraction, Transmission Electron Diffraction, and by Tunneling. P. H. Schmidt, E. G. Spencer, D. C. Joy, and J. M. Rowell, in *Superconductivity in d- and f-Band Metals*, D. H. Douglass, ed., New York: Plenum, p. 431.

Surface Effects on Valence in Rare Earth Intermetallic Compounds. G. K. Wertheim, J. H. Wernick, and G. Creelius, *Phys. Rev. B*, 18 (July 15, 1978), pp. 875-879.

Titanium Diselenide: Semiconductor, Semimetal or Excitonic Insulator. M. M. Traum, G. Margaritondo, N. V. Smith, J. E. Rowe, and F. J. DiSalvo, *Phys. Rev. B*, 17, No. 4 (February 15, 1978), pp. 1836-1838.

Two-Dimensional Electrical Conductivity in Quench-Condensed Metal Films. R. C. Dynes, J. P. Garno, and J. M. Rowell, *Phys. Rev. Lett.*, 40, No. 7 (February 12, 1978), pp. 479-482.

Two-Photon Absorption Spectrum of Aluminum Arsenide-Gallium Arsenide Monolayers. J. P. Van Der Ziel and A. C. Gossard, *Phys. Rev. B*, 17, No. 2 (January 15, 1978), pp. 765-768.

Verification and Design Aspects of True Concurrency. D. W. Mizell, 5th Ann. ACM Symp. Conf. Rec. Princ. Program Lang., January 1978, pp. 171-175.

Why a Session on LSI Testing? T. R. Finch, *Proc. Comcon*, February/March, 1978, p. 96.

PSYCHOLOGY

Perceptual Delay—A Consequence of Metacontrast and Apparent Motion. R. S. Didner, *Proc. 21st Hum. Factor Soc. Ann. Meet.*, October 1977, p. 33-37.

Personnel Subsystem Design Experience Under a Total System Development Philosophy. M. W. Soth and I. W. Grosso, *Proc. 21st Hum. Factor Soc. Ann. Meet.*, October 1977, pp. 20-23.

SYSTEMS ENGINEERING AND OPERATIONS RESEARCH

The Basic Principles of Teleconferencing. C. Stockbridge and R. J. Miller, in *Planning Exploratory Trials of New Interpersonal Telecommunications Evaluating New Telecommunications Services*, M. C. J. Elton, ed., New York: Plenum, 1978, pp. 721-746.

Bell System's Chicago Lightwave Communications Evaluation. D. D. Sell, *Proc. OSA/IEEE Laser Electroopt. Syst. Conf.* (February 1978), p. 26.

Centralized Maintenance of Telecommunication Equipment. A. Cuiwik and I.

R. Feuster, Proc. Nat. Telecommun. Conf. Rec., December 1977, p. 32:1/1-4.

CONECs: A Complete Splicing System for Aerial, Buried and Underground Telephone Plant. D. R. Frey and G. S. Cobb, Proc. 27th Wire Cable Int. Symp., November 1978, pp. 287-292.

The Dimension 2000 PBX: Network Design and Call Capacity Estimation. A. M. Gerrish and T. G. Lewis, Proc. Int. Priv. Electron. Switch. Syst. Conf., April 1978, pp. 44-47.

Horizon Communication System—An Advanced Technology System for Small Business Customers. T. M. Quinn, Proc. Int. Priv. Electron. Switch. Syst. Conf. (April) 1978, p. 70-75.

The New N4 Carrier Terminal. C. V. Fanuele, R. L. LeBlanc, and L. Y. Levesque, Telecommun. Conf., 1978, December 1978, Vol. 2, pp. 21.3.1-8.

Prospects for Lower Cost Telephone Exchange Service. D. Gillette, Public Util. Fortn. (May 11, 1978), pp. 39-40.

Recent Trends in Transmission Planning for Local Networks. R. W. Hatch and J. L. Sullivan, Proc. Subscriber Loop Serv. Int. Symp., March 1978, pp. 195-199.

The Role of the ITU and CCITT in Telecommunications. J. S. Ryan, Commun. High Speed Digital Trans. Tech. Switch. Syst. Seminar, Princeton U. (March 1978), pp. 2/1-2/6.

Telecommunications Is Going Digital. S. J. Buchsbaum, World Book Encyclopedia, Vol. T, February 1978, p. 73.

The Use of Thermal Analysis in the Selection of Stabilizers for Polyolefin Telecommunication Cable. M. G. Chan, H. M. Gilroy, I. P. Heyward, L. Johnson, and W. M. Martin, 36th Antec. Plast. Eng. Soc., 1978, pp. 381-383.



B.S.T.J. BRIEF

Optically Powered Speech Communication Over a Fiber Lightguide

By R. C. MILLER and R. B. LAWRY

(Manuscript received March 21, 1979)

I. INTRODUCTION

The photovoltaic conversion of optical power transmitted over a fiber lightguide can supply electrical power to low-drain semiconductor devices in remote locations. Acoustic powers comparable to those of conventional telephone ringers have been produced¹ in this way by using a fiber-coupled GaAlAs photovoltaic detector² to excite an electroacoustic tone generator. It was conjectured¹ that the electrical power for other telephone functions—transmit/receive, dialing, and hook-status recognition—could also be optically supplied, but the signaling techniques appropriate to a dielectric fiber were left unspecified. This note describes the implementation of two-way speech communication between an electrically powered local station and an optically powered station located at the remote end of a 1.1-km-long, single-strand, optical fiber. The remote-station sound alerter has also been operated over this link.

II. SPEECH SIGNALING METHOD

The method used for two-way optically powered speech signaling is illustrated in Figs. 1 and 2. Figure 1 depicts schematically the electronics in an optically powered remote station and the optics in an electrically powered local station. Speech-modulated optical power was launched into the local station end of the fiber from a GaAs injection laser emitting at wavelength λ_1 . The remote station contained a fiber-coupled GaAlAs double heterostructure transceiver,³ denoted PV/LED in Fig. 1, which functioned as a photovoltaic detector when

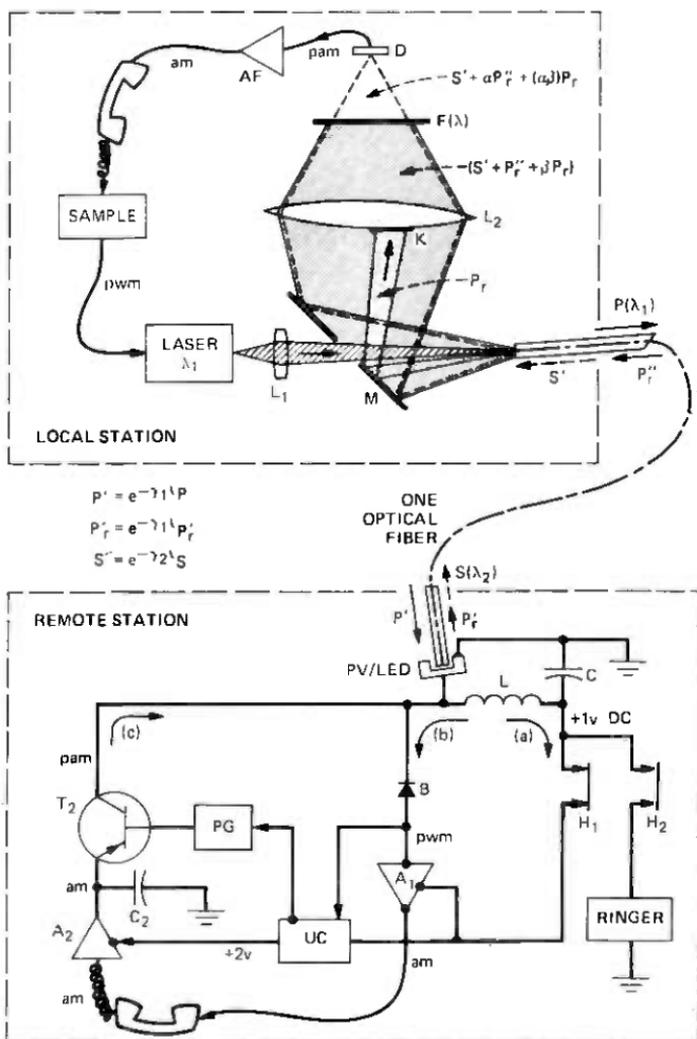


Fig. 1—Schematic of speech signaling method. The fiber length is l . Its transmittance is $\exp(-\gamma_1 l)$ for laser light and $\exp(-\gamma_2 l)$ for electroluminescence, in which γ_1, γ_2 are loss coefficients at the laser and electroluminescence wavelengths. Optical filter $F(\lambda)$ is transparent to electroluminescence but attenuates laser light by factor α . Specular component of reflection P_r is blocked by aperture stop K ; diffuse component is attenuated by $F(\lambda)$. Reflection P_r' is zero during electroluminescence pulse S' . Symbols am , pam , pwm refer, respectively, to amplitude modulated, pulse-amplitude modulated, or pulse-width modulated voltage or current. Switches H_1 and H_2 are shown in OFF-hook state; ON-hook, H_1 is open and H_2 closed. The 1 V and 2 V values indicated for the remote station dc voltages are nominal.

irradiated by light from the fiber and as an electroluminescent emitter when subjected to forward bias. One-volt dc electrical power was generated by photovoltaic conversion of the average power arriving over the fiber and was used to supply the remote station receive/

transmit circuits. The transceiver detected the speech modulation component of optical power, enabling its conversion by these circuits to acoustic power at the earphone, and it emitted electroluminescent pulses, amplitude-modulated by samples of the microphone output voltage. Other authors⁴⁻⁷ have discussed GaAlAs transceivers in which optical signals are generated and detected on a time-shared basis, with detection typically occurring under zero or reverse bias at very low optical powers. However, we are unaware of previous transceiver applications that combine injection electroluminescence with photovoltaic power conversion, even though these two phenomena have long been known⁸⁻¹⁰ as complementary effects in p-n junctions.

A form of pulse-width (or pulse-edge-position) modulation was used to transmit speech information from the local to the remote station. The local station laser was turned off periodically at a rate of $1/T$, equal at least to the Nyquist sampling rate for the speech bandwidth of interest, and was turned back on after a time T_M whose duration was modulated over an interval $\pm \Delta T_M$ in correspondence with samples of the analog output of the microphone. This type of modulation is illustrated in Fig. 2a by the time dependence of the laser power $P_r(\lambda_1)$ reflected from the local station end of the optical fiber. The off-time ($T_M + \Delta T_M$) was kept small compared to T to maximize the duty factor for transmitted power.

The optical pulses returned from the remote to the local station are illustrated in Figs. 2b and 2c. Laser power, reflected from the remote fiber-air interface and attenuated to level $P_r''(\lambda_1)$ by the fiber losses, arrived back at the local station after pulse round-trip time T_L . Electroluminescence pulse power, spectrally peaked at wavelength λ_2 , was generated by discharging capacitor C_2 through transistor switch T_2 with a slight delay T_S from the photovoltaic turn-off instant. A portion of this emission, proportional to the square of the fiber numerical aperture, entered the fiber-guided modes and reached the local station after attenuation to level $S'(\lambda_2)$. Thus, the time-varying optical power incident onto lens L_2 consisted of the laser reflections, $P_r(\lambda_1)$ and $P_r''(\lambda_1)$, and the very much smaller electroluminescence power, $S'(\lambda_2)$, modulated over a range $\Delta S'$ in correlation with the sampled audio-frequency voltage on capacitor C_2 . If this flux, depicted schematically in Fig. 2d, is allowed to impinge directly onto detector D , the output becomes noisy and is sensitive to microphonics. This impairment was greatly lessened by the dichroic filter $F(\lambda)$ which attenuated laser light by a large factor α relative to the longer wavelength electroluminescence. The specularly reflected power $P_r(\lambda_1)$ was attenuated an additional large factor β by the aperture stop K . The relative levels of these various powers at detector D , with filter $F(\lambda)$ and aperture stop K in place, are indicated in Fig. 2e.

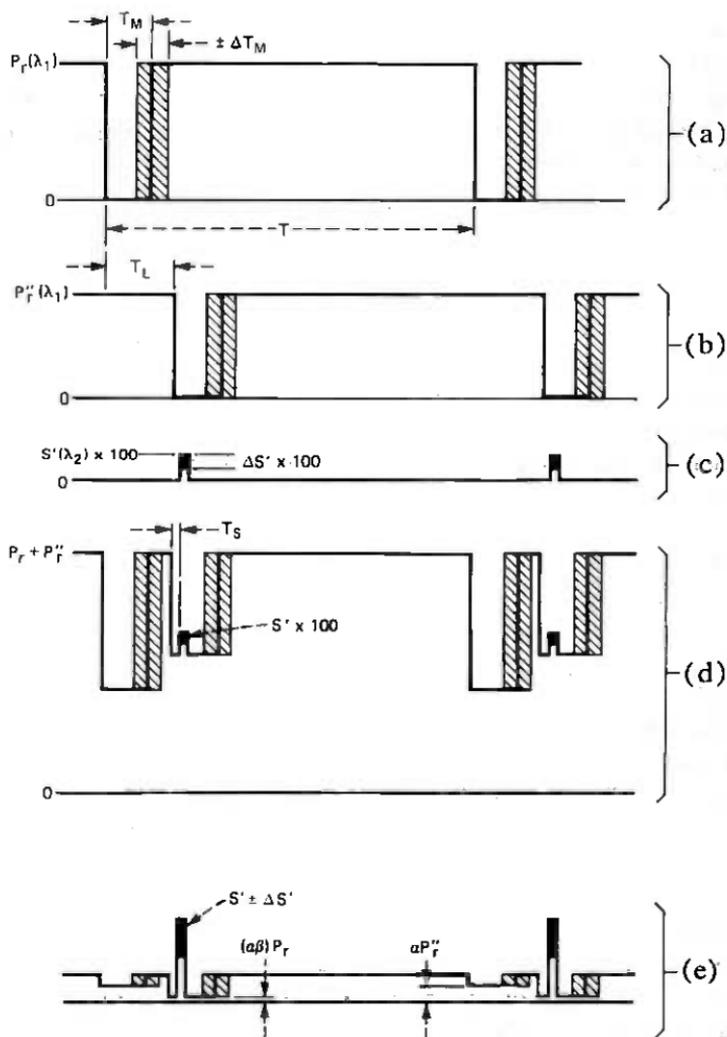


Fig. 2—Optical power levels in the local station. The laser is on most of the time. It is turned off at a rate $1/T$ equaling at least twice the highest speech frequency of interest, and is turned back on after a time T_M . The time T_M is variable over a range $\pm\Delta T_M$ (shaded area) in correspondence with the sampled, amplitude-modulated speech outgoing from the local station. (a) Laser power $P_r(\lambda_1)$ reflected from local station end of fiber. (b) Reflection $P_r''(\lambda_1)$ returned from remote end of fiber. The lightguide round-trip delay time is T_L . (c) Transmitted electroluminescence $S'(\lambda_2)$. The electroluminescence pulse is amplitude-modulated over a range $\Delta S'$. Generation of this pulse is slightly delayed (delay-time T_S) with respect to the remote transceiver turn-off time. (d) Superposition of P_r , P_r'' , and S' at lens L_2 . (e) Optical power incident onto photodetector D . Reflection P_r is reduced by a large factor β at aperture stop K . Filter $F(\lambda)$ reduces remaining light at wavelength λ_1 by a large factor α relative to wavelength λ_2 .

III. IMPLEMENTATION

The local and remote stations of Fig. 1 were linked by a 1.1-km-long, fused-silica fiber of 55- μm -diameter core and 0.22 numerical aperture,

whose attenuation was 6.3 dB at the 0.825- μm wavelength of the double-heterostructure GaAlAs laser, and 5.5 dB for the PV/LED electroluminescence spectrum peaked at 0.870- μm wavelength. The round-trip transmission time was $T_L = 11.5 \mu\text{s}$. Lens system L_1 converged the beam through a hole in mirror M to a focus at the fiber, with incidence angle such that specularly reflected light missed the mirror hole and did not refocus into the laser where it could cause cavity destabilization. Filter $F(\lambda)$ utilized the absorption band edge in p-type, $\text{Ga}_{0.97}\text{Al}_{0.03}\text{As}$ to obtain approximately 3-dB attenuation of electroluminescence and more than 20 dB attenuation of laser power.

The circuit branches labeled (a), (b), and (c) in the remote station section of Fig. 1 implement, respectively, the dc-powering, receive, and transmit functions defined earlier. The L-C filter of branch (a) provided ripple-free, 1.0-V power to the bipolar transistors in branches (b) and (c). Figure 3 depicts PV/LED voltage waveforms measured at an optical power sufficient to produce 1.61 mA of short-circuit current; the transceiver zero-volt level is indicated by horizontal arrows. The negative voltage excursion during the optical power OFF-interval is caused by the L-C filter. The clock period T was chosen equal to the 63.5- μs broadcast-television line period in anticipation of future experiments. Amplifier A_1 detected the modulation $T_M \pm \Delta T_M$ corresponding to the optical power OFF \rightarrow ON transition of Fig. 2a and used constant-current

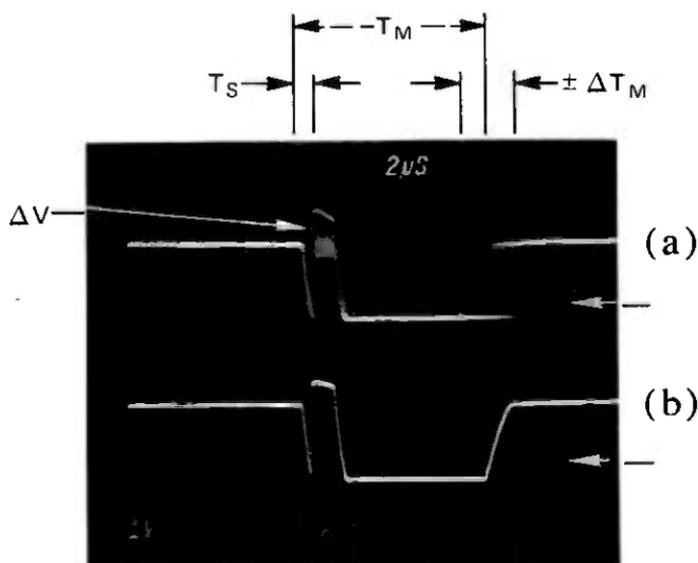


Fig. 3—Voltage waveforms at remote station transceiver. Audio-frequency modulation voltages were present at the local and remote station microphones for trace (a) and were absent for trace (b). Short horizontal arrows indicate the zero-voltage baselines. (Vertical scale 1 V/cm; horizontal scale, 2 μs /cm).

charging of a capacitor to convert the pulse-width variations to an analog audio-frequency voltage at the earphone. The forward voltage needed to produce electroluminescence was obtained by rectifying and filtering the output of a multivibrator-chopper voltage upconverter (UC). The *unrectified* chopper output was shaped by an R-L-C circuit (PG) into a 1.3- μ s wide pulse, delayed by $T_S = 0.2 \mu$ s from the optical turn-off instant, which discharged capacitor C_2 through transistor T_2 to produce the electroluminescence pulse. The forward-voltage pulse modulation, labeled ΔV in Fig. 3, corresponds to speech modulation of the C_2 voltage via the crystal microphone and audio amplifier A_2 . A compact tone generator (1.8-in. diameter \times 0.5-in. thickness) of approximately 35 percent electroacoustic efficiency at 2.0 kHz was incorporated into the remote station telephone base along with the sound-alert circuits.¹ These circuits were activated by depressing the telephone hook switch to close contacts H_2 and open contacts H_1 .

IV. RESULTS AND DISCUSSION

Satisfactory two-way speech transmission and vigorous sound alerting at the remote station have been obtained with 14 mW of dc-averaged laser power incident onto the fiber. This power is sufficient to produce short-circuit currents of 1.45 mA in the PV/LED transceiver at the remote end of the 1.1-km fiber. Quantitative studies of audio quality in this link have not been made; however, most observers describe the speech quality (noise, distortion, frequency response) at the remote and local earphones as excellent.

Noise in the remote earphone is inaudible, and excellent speech quality is obtained at the remote station with laser powers smaller than 10 mW incident onto the fiber. Speech volume at this earphone can easily be made uncomfortably loud, but distortion only appears if the local station speech-limiter circuits are maladjusted to permit ΔT_M to approach T_M . A low-level, white-noise-type hiss is present in the local station earphone under most operating conditions. This noise is associated with residual laser radiation present at detector D , and its volume can be varied from barely audible to intolerably loud by adjusting the position of optical filter $F(\lambda)$.

The use of a remote station detector with high photovoltaic efficiency² is essential to constructing a remotely located telephone all of whose power can be delivered from a central office. The optical complexity of the remote station is minimized by employing a unitary transceiver which takes advantage of the physical kinship between photovoltaic conversion and injection electroluminescence. Alternative arrangements in which the source and detector functions are performed in separately optimizable devices do, however, possess advantages, including the potential for greatly improving the source bright-

ness and for simplifying the time-sharing schemes. Several such arrangements are currently being investigated and will be reported later.

V. ACKNOWLEDGMENTS

The remote station A_1 amplifier was designed by H. R. Beurrier whose contributions to the later stages of this work have been important. The electroacoustic tone generator was provided by S. Kaufman. Use of GaAlAs "shelf" material for filter $F(\lambda)$ was suggested by R. W. Dixon; our filters were grown and etched by W. R. Wagner. We thank N. E. Schumaker, B. Schwartz, and L. A. Koszi for the PV/LED transceivers, R. L. Hartman for his selection of high performance lasers, and F. V. DiMarcello and J. C. Williams for the fiber.

REFERENCES

1. B. C. DeLoach, Jr., R. C. Miller, and S. Kaufman, "Sound Alerter Powered Over an Optical Fiber," *B.S.T.J.*, 57, No. 9 (November 1978), pp. 3309-3316.
2. R. C. Miller, B. Schwartz, L. A. Koszi, and W. R. Wagner, *Appl. Phys. Lett.*, 33 (1978), p. 721.
3. R. C. Miller, B. Schwartz, R. W. Dixon, and L. A. Koszi, unpublished work.
4. R. J. Lynch and D. C. T. Shang, *IBM Technical Disclosure Bulletin*, 12 (1969), p. 875.
5. J. J. Geddes, P. E. Petersen, D. Chen, and K. P. Koeneman, *Elec. Lett.*, 14 (1978), p. 214.
6. T. Ozeki, T. Uematsu, T. Ito, M. Yamamoto, and Y. Unno, *Opt. Lett.*, 2 (1978), p. 103.
7. W. M. Hubbard and H. E. Kehlenbeck, *Elec. Lett.*, 14 (1978), p. 553.
8. H. F. Ivey, *Electroluminescence and Related Effects*, New York: Academic Press, 1963, pp. 5-6.
9. E. E. Loebner in *Photoelectric Materials and Devices*, S. Larach, ed., New York: Van Nostrand, 1965, esp. p. 284.
10. W. van Roosbroeck and W. Shockley, *Phys. Rev.*, 94, 1954, p. 1558.

