RCA Review

September 1976 Volume 37 No. 3

RCARCI 37(3) 277–434 (1976)

RCA Review, published quarterly in March, June, September and December by RCA Research and Engineering, RCA Corporation, Princeton, New Jersey 08540. Entered as second class matter July 3, 1950 under the Act of March 3, 1879. Second-class postage paid at Princeton, New Jersey, and at additional mailing offices. Effective Jan. 1, 1971, subscription rates as follows: United States and Canada: one year $6.00, two years $10.50, three years $13.50; in other countries, one year $6.40, two years $11.30, three years $14.70. Single copies (except for special issues) up to five years old $3.00.

RCA Review

A technical journal published quarterly by RCA Research and Engineering in cooperation with the subsidiaries and divisions of RCA.

Contents

September 1976 Volume 37 Number 3

Performance Characteristics of Antennas for Direct Broadcasting Satellite Systems Including Effects of Rain Depolarization

I. P. Shkarofsky and H. J. Moody

RCA Ltd., Ste. Anne de Bellevue, Canada

Abstract—Phenomena affecting the design and operation of a broadcast satellite at 12 GHz are analyzed. We provide relations on attenuation and rain depolarization effects for incident linear and circular polarizations, allowing for imperfect antennas with finite isolation in clear weather and allowing for misalignment between transmitter and receiver polarizations. We then investigate interference due to an adjacent transmitter beam, including the relevant antenna patterns, and we deduce the net isolation. The relative advantages of linear and circular polarizations are outlined. The use of linear polarization at 12 GHz is recommended. We then comment on reference co-polarization and cross-polarization antenna patterns. Finally, systems aspects of a broadcast satellite system are discussed. Included in the discussion are limitations to antenna reflector size, the apparent change in shape of a ground area when the satellite is moved to different locations in the orbit, problems associated with channel assignment in a multibeam environment, and the geometrical aspects of using linear polarization in communication satellite systems.

1. Introduction

The European Broadcasting Union (EBU) is proposing to launch a broadcast satellite, having a high-power transmitter, beaming to individual home-type receivers at a frequency about 12 GHz. Various problem areas arise associated with interference problems caused by the satellite high power and by the relatively inexpensive receivers.

Neighboring satellite transmitters must be designed so that their transmitted beams provide minimum interference at a receiver on earth. Possibilities are spot beams directed to specific areas, frequency differences, and differing polarizations between beams. Since rain depolarization effects are minimized for incident horizontal polarization and especially for polarization in the vertical plane, it is desirable to construct

beams having one of these polarizations at the local spot on earth to which the beam is directed. The advantages of circular versus linear polarization must be considered before circular polarization is ruled out. To double the number of channels, each beam may be designed with two orthogonally polarized channels. Another problem is side lobes, which can provide interfering signals at large off-axis angles unless their power envelope is minimized. Finally, satellite antennas having very small beam widths require large apertures.

The earth receivers will probably be inexpensive, fixed, and small as compared to presently used earth stations. Their relative smallness means that the central beam width will be of the order of 1°–2°, which is sufficiently broad that adjacent satellite beams, even through received off-axis, may interfere with the signal from the desired satellite. The fact that the receivers are unadjustable means that it will be difficult to correct for misalignment between the polarization direction of the transmitted signal and that of the receiver. Misalignment can be caused by the wind, other weather conditions, mishandling, and by Faraday rotation in the ionosphere. (The last is negligible for frequencies above 8 GHz.) Misalignment introduces cross-polarization. Also the attempt to construct inexpensive receivers may run counter to means for bettering their inherent cross-polarization isolation in clear weather. One can expect the receiver antenna isolation to be worse than that of the satellite antenna. Another important factor is protection against rain dripping over the feed of the earth receiver. Rain dripping over the feed is known to greatly worsen the antenna isolation.

Section 2 discusses the theory of rain depolarization phenomena and gives the results of measurements that have been reported in the literature. The contribution of imperfections in the transmit and receive antennas to the coupling between the cross-polarized channels is included. The decrease in antenna response away from the antenna axis is also included, and standardized formulas for calculating this response are discussed.

Section 3 discusses system aspects of a broadcast satellite system. Included in the discussion are limitations to antenna reflector size, the apparent change in shape of a ground area when the satellite is moved to different locations in the orbit, problems associated with channel assignment in a multibeam environment, and the geometrical aspects of using linear polarization in communication satellite systems.

2. Cross-Polarization Effects

In this section we consider the effects of rain on attenuation, cross-polarization discrimination, and isolation. In Sec. 2.1, we provide relations

for calculating attenuation and rain depolarization effects for incident linear and circular polarizations, for the case of perfect antennas, and for exact alignment. In Sec. 2.2, we extend these results to allow for imperfect antennas with finite isolation in clear weather and for misalignment of the polarizations of the transmitter and receiver antennas. In Sec. 2.3, we investigate interference from neighboring satellites or between adjacent beams. The antenna patterns are included, and relations are given for the net isolation. Previous documents have either not analyzed the effect of rain properly or have omitted it altogether.

In Sec. 2.4, we consider the relative advantages and disadvantages of linear versus circular polarization and recommend the use of linear polarization at 12 GHz and higher frequencies. In Sec. 2.5, we comment on reference antenna patterns. Considerations related to antenna design problems are given in Sec. 3.

2.1 "Classical" Theory on Rain Depolarization

A crude estimate[1] has been proposed, for circular polarization, to specify rainfall depolarization values to be adopted for system planning. It does not however, allow for linear polarizations. Also it does not include the important effect of the elevation angle or the equivalent path length through the rainfall. Calculated results are available from Oguchi and Hosoya[2] and Chu[3], and these better tabulated results should be used.

The "classical" formulation, which we propose to follow, assumes that the following quantities are known.

(a) The slant path length distance through the rainfall, r' in km, between the satellite and earth receiver. This is usually deduced experimentally. Theoretical formulas[4] that give a good estimate for the vertical (V) and horizontal (H) extents of the rainfall are

$$H(\text{km}) = 5.34 - 1.67 \log_{10} R \text{ and } V(\text{km}) = 17.18 - 5.13 \log_{10} R, \quad [1]$$

where R is rainfall in mm/hr, assuming $R > 1$ and assuming that the values of R are available experimentally. Then the value of r' is given by whichever one of the following two expressions is less than $(H^2 + V^2)^{1/2}$.

$$r' = H/\sin\alpha \quad \text{or} \quad r' = V/\cos\alpha. \quad [2]$$

Here $90° - \alpha'$ is the elevation angle in degrees and $\alpha = \pi\alpha'/180$ in radians is the angle of incidence with respect to the vertical.

(b) The differential attenuation values per unit length $\Delta A' = A'_H - A'_V$ in dB/km.

Table 1—Calculated Values of Attenuation and Phase Shift Due to Rain (Angle of Incidence α' of 90°; Frequency 11 GHz) (From Chu[3])

Rain Rate mm/hr	Attenuation (dB/km)			Phase Shift (deg/km)		
	Vertical Polarization A_H'	Horizontal Polarization A_V'	Difference $\Delta A'$ (Horiz. — Vert.)	Vertical Polarization	Horizontal Polarization	Difference (Horiz. — Vert.)
0.25	0.002428	0.002669	0.000241	0.3985	0.4195	0.021
1.25	0.01592	0.01820	0.00228	1.579	1.697	0.118
2.5	0.03787	0.04399	0.00612	2.880	3.127	0.247
5.0	0.09144	0.1076	0.01616	5.266	5.783	0.517
12.5	0.2907	0.3470	0.0563	11.69	13.06	1.37
25.0	0.6898	0.8293	0.1395	21.32	24.18	2.86
50.0	1.605	1.945	0.340	38.94	44.93	5.99
100.0	3.586	4.392	0.806	70.25	82.58	12.33
150.0	5.605	6.919	1.314	99.26	118.3	19.04

(c) The differential phase shift values per unit length $\Delta\Phi'$ in degrees/km. Both $\Delta A'$ and $\Delta\Phi'$ are already averaged over the drop size distribution in space. Calculated values for them are available from Chu[3] for $\alpha' = 90°$ (see Table 1) and from Oguchi and Hosoya[2] for $\alpha' = 90°$, 70°, 50°, and 30° (see Table 2). Tabulated values are given for rain rates R ranging from 0.25 to 150 mm/hr.

(d) The canting angle τ' in degrees. Consider the plane perpendicular to the propagation direction and project onto this plane the average image ellipse of the raindrop spheroids. The angle τ' is defined to be the angle between the electric field and the nearby axis of this ellipse. For an incident wave with horizontal or vertical polarization, the effective τ' can be taken[5] to be between 2° to 4°. This approach is simpler than that of Chu,[3] who uses the averages $\langle|\tau'|\rangle \approx 25°$ and $\langle\sin^2 2\tau\rangle \approx (0.14)^2 \sin^2(2\langle|\tau|\rangle)$ in the equations. If the polarization angle (see Eq. [38]) is non-zero, then τ' has to be equated to the sum of the canting and polarization angles.

Before we use the suggested formulas, we require the following changes in units:

$$\tau = \frac{\pi}{180}\tau', \quad \alpha = \frac{\pi}{180}\alpha', \quad r = 10^3 r'$$

$$\Delta A = \Delta A' \times 10^{-3}/8.686 \text{ and } \Delta\Phi = \Delta\Phi' \times 10^{-3}\pi/180 \qquad [3]$$

Define the following quantities used by McCormick and Hendry[6]

$$p^2 = \frac{\cosh(r\Delta A) - \cos(r\Delta\Phi)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)},$$

$$p\cos\chi = \frac{\sinh(r\Delta A)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)}. \qquad [4]$$

Then the cross-polarization discrimination values ($XPD_{H,V}$) due to rain alone for linear polarizations in two orthogonal directions are given by

$$XPD_{H,V} = 10\log_{10}\left[\frac{1 \mp 2p\cos\chi\cos 2\tau + p^2\cos^2 2\tau}{p^2\sin^2 2\tau}\right]. \qquad [5]$$

The upper sign applies to H and the lower sign to V polarization. For circular polarization, either right-hand or left-hand, the result is the same as letting $\tau = \pi/4$ in the above relation, yielding

$$XPD_C = -10\log_{10} p^2. \qquad [6]$$

Table 2—Calculated Values of Attenuation and Phase Shift Due to Rain (Frequency 11 GHz) (From Oguchi and Hosoya[2])

Rain Rate (mm/hr)	Attenuation (dB/km)			Phase Shift (deg/km)		
	Vertical Polarization	Horizontal Polarization	Difference (Horizontal −Vertical)	Vertical Polarization	Horizontal Polarization	Difference (Horizontal −Vertical)
			Angle of incidence α': 90°			
0.25	0.002507	0.002731	0.000224	0.3962	0.4150	0.0188
1.25	0.01604	0.01809	0.00205	1.560	1.664	0.104
2.5	0.03781	0.04326	0.00545	2.844	3.064	0.220
12.5	0.2852	0.3349	0.0497	11.54	12.77	1.23
25.0	0.6745	0.7983	0.1238	21.09	23.66	2.57
50.0	1.554	1.855	0.301	38.47	43.85	5.38
100.0	3.504	4.230	0.726	69.95	81.13	11.18
150.0	5.467	6.637	1.17	98.99	116.1	17.11
			Angle of incidence α': 70°			
0.25	0.002524	0.002721	0.000197	0.3983	0.4150	0.0167
1.25	0.01614	0.01796	0.00182	1.572	1.664	0.092
2.5	0.03805	0.04287	0.00482	2.869	3.063	0.194
12.5	0.2873	0.3312	0.0439	11.69	12.78	1.09
25.0	0.6801	0.7895	0.1094	21.42	23.69	2.27
50.0	1.569	1.835	0.266	39.18	43.93	4.75
100.0	3.547	4.189	0.642	71.52	81.40	9.88
150.0	5.543	6.580	1.037	101.4	116.6	15.2
			Angle of incidence α': 50°			
0.25	0.002567	0.002698	0.000131	0.4038	0.4148	0.011
1.25	0.01641	0.01762	0.00121	1.602	1.663	0.061
2.5	0.03868	0.04188	0.0032	2.933	3.062	0.129
12.5	0.2925	0.3217	0.0292	12.07	12.79	0.72
25.0	0.6944	0.7673	0.0729	22.24	23.75	1.51
50.0	1.609	1.786	0.177	40.99	44.15	3.16
100.0	3.658	4.087	0.429	75.50	82.08	6.58
150.0	5.740	6.432	0.692	107.7	117.8	10.1
			Angle of incidence α': 30°			
0.25	0.002616	0.002672	0.000056	0.4100	0.4147	0.0047
1.25	0.01672	0.01762	0.00051	1.636	1.662	0.026
2.5	0.03939	0.04076	0.00137	3.005	3.060	0.055
12.5	0.2985	0.3110	0.0125	12.49	12.80	0.31
25.0	0.7109	0.7420	0.0311	23.17	23.81	0.64
50.0	1.654	1.730	0.076	43.04	44.39	1.35
100.0	3.787	3.971	0.184	80.05	82.85	2.80
150.0	5.968	6.265	0.297	114.9	119.2	4.3

The attenuation in dB of the signals with respect to their clear weather values are given by

$$\mathrm{ATT}_{H,V} = (A'_H + A'_V)\frac{r'}{2} + 5\log_{10}[(1+p^2)^2 - 4p^2\cos^2\chi]$$

$$- 10\log_{10}[1 \mp 2p\cos\chi\cos 2\tau + p^2\cos^2 2\tau] \qquad [7]$$

for linear polarization where $A'_{H,V}$ are the attenuation values in dB/km given in Tables 1 and 2 for the respective two linear polarizations. For circular polarization

$$\mathrm{ATT}_C = (A'_H + A'_V)\frac{r'}{2} + 5\log_{10}[(1+p^2)^2 - 4p^2\cos^2\chi]. \qquad [8]$$

Obviously, even XPD_C is not simply related to ATT_C, so that the equation suggested[1] at 12 GHz, namely $\mathrm{XPD}_C = 31 - 2(\mathrm{ATT}_C)$ can at best be only approximate. The suggestion is that it be used for $\mathrm{ATT}_C \leq 6.1$ dB, valid for 99% of the worst month. However, attenuations greater than 6 dB and in fact up to 20 dB do occur[5] at 11 GHz for the remaining 1% of the worst month. Consequently, for a more reliable system, better theoretical predictions should be applied. The relations given here can be applied to satisfy this requirement.

2.2 Combined Effect of Clear-Weather Antenna Isolation and Rain Depolarization (Worst-Case Limit)

Antennas in clear weather have finite isolation determined by the difference in dB between the co-polarization and cross-polarization patterns. On-axis and off-axis cross-polarization discrimination values have been plotted both for linear and circular polarizations.[7] The patterns also show that in many cases there is no minimum on-axis for the cross-polarization pattern.

Polarizers in the feed, which are required to generate circular polarization, are available with an ellipticity ratio of 0.2 dB, giving 39 dB isolation at 4–6 GHz. Low ellipticity ratios (<0.5 dB) can probably be obtained at 12 GHz.

Of the various antennas for satellite use, the offset-fed paraboloid is advantageous, especially if multiple beams are to be used. An offset feed provides (1) less return into the horn, (2) less blockage loss associated with the feed and support structures, (3) generally lighter weight, and (4) higher cross-polarization discrimination, since a major degrader is the supporting structure of the feed which causes polarization changeover.

The state of the art on cross-polarization discrimination is about 33

dB for linear and 30 dB for circular polarization up to the 3-dB contour coverage area. Two possible methods for improvement are (1) to use a circular waveguide feeding the horn with the TE_{11} mode, which after reflection from the paraboloid gives a purer linear wave or (2) use a longer focal length to diameter ratio, of the order of 0.7 or larger. Extensions of method (1) are discussed by Tseng.[8]

The clear-weather isolation of a link is determined by the effects of misalignment of polarization, finite antenna ellipticity ratios for circular polarization, finite antenna isolations for linear polarization, and the effect of rain.

In the following, we generalize previous work[7] in order to include the effect of rain along the propagation path. The works of Oguchi, Chu, McCormick, and others are used. We consider first linear polarization and then circular polarization.

Linear Polarization

We develop here theoretical equations to calculate the cross-polarization discrimination ($XPD_{H,V}$) between two receiver channels for a single transmitted linearly polarized wave, either horizontal or vertical, along a one-way path. These equations provide results for the *worst case*, where the clear weather phases of ellipticity of the transmitter and receiver antennas are of opposite sign. The following effects are included:

(a) Off beam-axis transmission and reception, where ϕ_T and ϕ_R are the angles that the line-of-sight propagation path makes with the respective normals to the antenna apertures.

(b) The clear-weather finite polarization discrimination, $D'_T(\phi_T)$ in dB, of the satellite transmitter antenna between the co-polarization antenna radiation pattern, $F'_T(\phi_T)$ in dB (negative quantity), and the cross-polarization radiation pattern, $f''_T(\phi_T)$ in dB (negative quantity), with $D'_T(\phi_T) = F'_T(\phi_T) - f''_T(\phi_T)$.

(c) The clear-weather finite polarization discriminations, $D'_R(\phi_R)$ and $D''_R(\phi_R)$ in dB, of the ground antenna. D'_R refers to discrimination in the co-polarization receiver channel between reception of a co-polarization radiation pattern, $F'_R(\phi_R)$ in dB, and a cross-polarization pattern, $f'_R(\phi_R)$ in dB, with $D'_R(\phi_R) = F'_R(\phi_R) - f'_R(\phi_R)$. D''_R refers to discrimination in the cross-polarization receiver channel between reception of the direct radiation pattern, $F''_R(\phi_R)$ in dB in the cross-polarization direction, and the pattern $f''_R(\phi_R)$ in dB in the co-polarization orthogonal direction, with $D''_R(\phi_R) = F''_R(\phi_R) - f''_R(\phi_R)$. Also let the respective on-axis antenna gains of the two receiver channels be G'_R in dB and G''_R in dB.

(d) Angular difference of the polarization direction between the satellite

and the earth antenna, θ' in degrees, and/or the Faraday rotation angle.

(e) Depolarization due to rain, assuming a constant or effective rainfall rate along the path. We use
   (1) the slant path length distance through the rainfall, r' in km, from the satellite to earth receiver,
   (2) the differential attenuation values per unit length, $\Delta A'$ in dB/km,
   (3) the differential phase shift values per unit length, $\Delta\Phi'$ in deg/km,
   (4) the effective canting angle of the raindrops, τ' in degrees.

   Note that (1), (2) and (3) depend on elevation angle ($90° - \alpha'$) in degrees, where α' is the angle of incidence with respect to the vertical, and on the rain rate R in mm/hr. Also (2) and (3) depend on frequency f in GHz. For a transmitter using horizontal or vertical polarization, τ can be taken as a few degrees, say 4° as an example of a bad case.[5] Values of (2) and (3) versus R, α', and f are obtainable from Chu[3] and from Oguchi and Hosoya.[2] At 11 GHz, close to the frequencies of interest here, Oguchi and Hosoya give $\Delta A'$ and $\Delta\Phi'$ values for $\alpha' = 90°, 70°, 50°, 30°$, (elevation angles = 0°, 20°, 40°, and 60°, respectively) and for eight R values from 0.25 to 150mm/hr.

Before we use the formulas, we require the following changes in units:

$$D_T = 10^{D_T'/20}, \quad D_{R1} = 10^{D_R'/20}, \quad D_{R2} = 10^{D_R''/20}$$

$$\theta = \frac{\pi}{180}\theta', \quad \tau = \frac{\pi}{180}\tau', \quad \alpha = \frac{\pi}{180}\alpha', \quad r = 10^3 r'$$

$$\Delta A = \Delta A' \times 10^{-3}/8.686, \quad \Delta\Phi = \Delta\Phi' \times 10^{-3}\pi/180 \qquad [9]$$

We also define the following quantities used by McCormick and Hendry.[6]

$$p^2 = \frac{\cosh(r\Delta A) - \cos(r\Delta\Phi)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)}, \quad \tan\chi = \frac{\sin(r\Delta\Phi)}{\sinh(r\Delta A)}$$

$$p\cos\chi = \frac{\sinh(r\Delta A)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)},$$

$$p\sin\chi = \frac{\sin(r\Delta\Phi)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)} \qquad [10]$$

Then the values of XPD for a one-way path are given by

$$\mathrm{XPD}_{H,V} = 10\log_{10}\frac{N_{\mp}}{D_{\mp}} + [f_R'(\phi_R) - f_R''(\phi_R)]_{H,V} + (G_R' - G_R'')_{H,V} \qquad [11]$$

where

$$N_{\mp} = [(D_{R1} - D_T^{-1})\cos\theta \mp (D_{R1} + D_T^{-1})p\cos\chi\cos(2\tau + \theta) - (1 + D_{R1}D_T^{-1})p\sin\chi\sin(2\tau + \theta)]^2 + [(1 - D_{R1}D_T^{-1})\sin\theta + (D_{R1} + D_T^{-1})p\sin\chi\cos(2\tau + \theta) \mp (1 + D_{R1}D_T^{-1})p\cos\chi\sin(2\tau + \theta)]^2 \quad [12]$$

and

$$D_{\mp} = [(1 + D_{R2}D_T^{-1})\cos\theta \mp (1 - D_{R2}D_T^{-1})p\cos\chi\cos(2\tau + \theta) + (D_{R2} - D_T^{-1})p\sin\chi\sin(2\tau + \theta)]^2 + [(D_{R2} + D_T^{-1})\sin\theta - (1 - D_{R2}D_T^{-1})p\sin\chi\cos(2\tau + \theta) \mp (D_{R2} - D_T^{-1})p\cos\chi\sin(2\tau + \theta)]^2 \quad [13]$$

The upper signs refer to XPD_H, where the horizontal is the co-polarization channel, and the lower signs refer to XPD_V, where the vertical is the co-polarization channel.

Examples are now given. In the absence of rain, $p = \chi = 0$, and for $F_R' = F_R''$, $f_R' = f_R''$, and $G_R' = G_R''$, Eq. [11] reduces to[7]

$$XPD = 10\log_{10}\left[\frac{(D_R - D_T^{-1})^2 - (D_R^2 - 1)(1 - D_T^{-2})\sin^2\theta}{(1 + D_RD_T^{-1})^2 + (D_R^2 - 1)(1 - D_T^{-2})\sin^2\theta}\right]. \quad [14]$$

Eq. [14] gives the reduction in XPD due to causes (a) to (d) above. Obviously, as D_R and D_T go to infinity, we obtain[7]

$$XPD = 10\log_{10}\cot^2\theta \quad [15]$$

For the situation of perfect antennas, D_R and D_T infinite, and for exact alignment $\theta = 0$, we obtain Eq. [5]:

$$XPD_{H,V} = 10\log_{10}\left[\frac{1 \mp 2p\cos\chi\cos2\tau + p^2\cos^2 2\tau}{p^2\sin^2 2\tau}\right], \quad [16]$$

which is the relation given by McCormick and Hendry[6] for linear polarization. This gives the reduction in XPD due to rain alone (cause (e) above). If we include effects (d) and (e), again for D_R and D_T infinite but θ finite, we obtain

$$XPD_{H,V} = 10\log_{10} \times \left[\frac{\cos^2\theta \mp 2p\cos\chi\cos\theta\cos(2\tau + \theta) + p^2\cos^2(2\tau + \theta)}{\sin^2\theta \mp 2p\cos\chi\sin\theta\sin(2\tau + \theta) + p^2\sin^2(2\tau + \theta)}\right]. \quad [17]$$

The more general Eq. [11] includes all causes (a) to (e).

Circular Polarization

We propose the following equations to calculate the cross-polarization discrimination ($XPD_{R,L}$) for circular polarized waves, either right or left hand, along a one-way path. We consider the worst case, where the clear-weather ellipticity axes of the transmitter and receiver antennas are at right angles to each other. The following effects are included:

(a) Off beam-axis transmission and reception, where ϕ_T and ϕ_R are the angles the line-of-sight propagation path makes with the respective normals to the antenna apertures.

(b) The clear-weather finite ellipticity, $E'_T(\phi_T)$ in dB, of the satellite transmitter antenna. This is related to the difference between the co-polarization antenna radiation pattern, $F'_T(\phi_T)$ in dB, and the cross-polarization radiation pattern, $f''_T(\phi_T)$ in dB, by

$$F'_T(\phi_T) - f''_T(\phi_T) = 20 \log_{10} [(10^{E'_T/20} + 1)/(10^{E'_T/20} - 1)]$$
$$\approx 24.797 - 20 \log_{10} (E'_T) + 0.009594 (E'_T)^2, \qquad [18]$$

where the approximation holds for $E'_T/20 \ll 1$.

(c) The clear-weather finite ellipticities, $E'_R(\phi_R)$ and $E''_R(\phi_R)$ in dB, of the ground antenna. $E'_R(\phi_R)$ refers to the co-polarization receiver channel and is related (similar to the above formula) to the difference between the co-polarization radiation pattern, $F'_R(\phi_R)$ in dB, and the cross-polarization pattern, $f'_R(\phi_R)$ in dB. Similarly $E''_R(\phi_R)$ refers to the cross-polarization receiver channel and is related to the difference between the direct radiation pattern, $F''_R(\phi_R)$ in dB in the cross-polarization direction, and the pattern $f''_R(\phi_R)$ in dB in the co-polarization orthogonal direction. Because the axes of the receiver ellipticity are orthogonal to E'_T, the E_R's are negative if E'_T is taken as positive. Let the respective on-axis antenna gains of the two receiver channels be G'_R in dB and G''_R in dB.

(d) Depolarization due to rain, similar to (e) above for linear polarization. Before we use the equations, we require the following changes in units:

$$E_T = 10^{E'_T/20}, \quad E_{R1} = 10^{|E'_R|/20}, \quad E_{R2} = 10^{|E''_R|/20}$$

$$\tau = \frac{\pi}{180}\tau', \quad \alpha = \frac{\pi}{180}\alpha', \quad r = 10^3 r'$$

$$\Delta A = \Delta A' \times 10^{-3}/8.686, \quad \Delta\Phi = \Delta\Phi' \times 10^{-3}\pi/180 \qquad [19]$$

$$p^2 = \frac{\cosh(r\Delta A) - \cos(r\Delta\Phi)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)},$$

$$p\cos\chi = \frac{\sinh(r\Delta A)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)}$$

$$p\sin\chi = \frac{\sin(r\Delta\Phi)}{\cosh(r\Delta A) + \cos(r\Delta\Phi)} \qquad [20]$$

Then the values of $XPD_{R,L}$ for a one-way path are given by

$$XPD_{R,L} = 10\log_{10}\frac{N_{\mp}^{c}}{D_{\mp}^{c}} + (G_R' - G_R'')_{R,L} + 10\log_{10}\left[\frac{10^{F'R/10} + 10^{f'R/10}}{10^{F''R/10} + 10^{f''R/10}}\right] \qquad [21]$$

where

$$N_{\mp}^{c} = [(E_T + E_{R1})^2 \mp 2p\cos\chi\cos 2\tau(E_T^2 - E_{R1}^2) - 2p\sin\chi\sin 2\tau(E_T + E_{R1})(1 - E_T E_{R1}) + p^2\cos^2 2\tau(E_T - E_{R1})^2 + p^2\sin^2 2\tau(1 - E_T E_{R1})^2]/[(1 + E_T^2)(1 + E_{R1}^2)] \qquad [22]$$

and

$$D_{\mp}^{c} = [(1 - E_T E_{R2})^2 \pm 2p\cos\chi\cos 2\tau(1 - E_T^2 E_{R2}^2) + 2p\sin\chi\sin 2\tau(E_T + E_{R2})(1 - E_T E_{R2}) + p^2\cos^2 2\tau(1 + E_T E_{R2})^2 + p^2\sin^2 2\tau(E_T + E_{R2})^2]/[(1 + E_T^2)(1 + E_{R2}^2)] \qquad [23]$$

The upper signs refers to XPD_R, where right-hand circular is the co-polarization channel, and the lower signs refer to XPD_L, where left-hand circular is the co-polarization channel.

Examples are now given. In the absence of rain, $p = \chi = 0$, and for $E_{R1} = E_{R2}$ and $G_R' = G_R''$, Eq. [21] reduces to[7]

$$XPD = 10\log_{10}\left[\frac{E_T + E_R}{1 - E_T E_R}\right] \approx 24.797 - 20\log_{10}(E_T' + |E_R'|) + 0.009594\,[(E_T')^2 + (E_R')^2 - 4E_T'|E_R'|] \qquad [24]$$

where the approximation holds for $E_T'/20 \ll 1$ and $|E_R'|/20 \ll 1$. For the situation of perfect antennas, E_T', E_R', and E_R'' are zero, E_T, E_{R1}, and E_{R2} are one, so that $XPD = 10\log_{10} p^{-2} = -20\log_{10} p$, as in Eq. [6], which is the relation given by McCormick and Hendry[6] for circular polarization. This relation is independent of τ and is the same for right- and left-hand

circular polarizations. The more general Eq. [21] includes all causes (a) to (d) and, as can be seen, depends on τ.

2.3 Combined Effect Based on Random Phase Limit with Application to Interference

Interference arises from signals from an unwanted satellite arriving at an earth station, ordinarily receiving signals from a wanted closer satellite. The effect of depolarization induced by rain is to change the polarization of the interfering radiation to the polarization of the receiver. The basic assumption, which is reasonable, is that the powers of the contributions to the interfering signal can be added, since the relative phases are random with respect to each other. Simple formulas that have been suggested[9] omit (a) the dependence of the cross-polarization on the type of incident polarization (linear horizontal, linear vertical or circular), and (b) attenuation due to rain, which as we see below alters the relations appreciably.

The notation below follows that used in the previous two sections. The off-axis angle ϕ_T allows the radiation from the interfering satellite to reach the receiver, which detects it at its off-axis angle ϕ_R. We again include all the effects given in Sec. 2.2 for generality. In contrast to Sec. 2.2, where we aligned the phases of transmitter and receiver for the worst result, we now consider the phases to be random. We consider several situations below.

Case (1a)—Receiving Antenna with Nearly the Same Linear Polarization as that of the Interfering Satellite

Define

$$P_{c\pm} = \cos^2\theta + p^2\cos^2(2\tau + \theta) \pm 2p\cos\theta\cos\chi\cos(2\tau + \theta),$$

$$P_{s\pm} = \sin^2\theta + p^2\sin^2(2\tau + \theta) \pm 2p\sin\theta\cos\chi\sin(2\tau + \theta). \quad [25]$$

Here θ is the misalignment angle between the transmitter and receiver polarization directions, τ is the canting angle, and the rain parameters p and χ are defined in Sec. 2.1 and 2.2. These P_c and P_s factors allow for attenuation and cross-polarization due to rain.

The transmitter co-polarization radiation pattern is denoted by $F_T'(\phi_T)$ in dB and its cross-polarization radiation pattern is denoted by $f_T''(\phi_T)$ in dB. Later we also consider the transmitter radiating with orthogonal polarizations and then let $F_T''(\phi_T)$ in dB be the direct radiation pattern in the cross-polarization direction and let $f_T'(\phi_T)$ in dB be the cross-polarization pattern in the co-polarization orthogonal direction. For the receiver, we only use the co-polarization radiation pattern $F_R'(\phi_R)$

in dB and its cross-polarization pattern $f'_R(\phi_R)$ in dB, both for reception in the co-polarization receiver channel. We also define

$$F_{T1} = 10^{F_T'/20}, \quad F_{T2} = 10^{F_T''/20}, \quad f_{T1} = 10^{f_T''/20},$$
$$f_{T2} = 10^{f_T'/20}$$
$$F_R = 10^{F_R'/20} \text{ and } f_R = 10^{f_R'/20}. \qquad [26]$$

Here F'_T and f''_T are normalized to $F'_T(0) = 0$. Similarly, F''_T and f'_T are normalized to $F''_T(0) = 0$ and F'_R and f'_R are normalized to $F'_R(0) = 0$. For other off-axis angles, these quantities in dB are negative. The on-axis gains in dB are denoted respectively by G'_T, G''_T, and G'_R.

Let P_R be the received power, P_T the transmitted power, λ the wavelength, and d the satellite to earth distance. As before, r' in km is the propagation path through the rain. The other symbols used below are defined in Sec. 2.1 and 2.2.

For the case considered, the corrected result that we propose is given by

$$\begin{aligned}
&(P_R - P_T - G'_T - G'_R)_{H,V} + 20 \log_{10}(4\pi d/\lambda) \\
&+ (A'_H + A'_V)r'/2 + 5 \log_{10}[(1 + p^2)^2 - 4p^2 \cos^2\chi] \\
&= 10 \log_{10}(F^2_{T1}F^2_R P_{c\mp} + f^2_{T1}f^2_R P_{c\pm} + F^2_{T1}f^2_R P_{s\mp} + f^2_{T1}F^2_R P_{s\pm}) \\
&= 10 \log_{10}[10^{-(F'_T+F'_R)/10}P_{c\mp} + 10^{-(f''_T+f'_R)/10}P_{c\pm} \\
&+ 10^{-(F'_T+f'_R)/10}P_{s\mp} + 10^{-(f''_T+F'_R)/10}P_{s\pm}]
\end{aligned} \qquad [27]$$

The upper sign refers to H and the lower sign to V polarization. We now let $\theta = 0$ so that $P_s = P_{s\pm} = p^2 \sin^2 2\tau$ and we define the cross-polarization discrimination ratio as in Eq. [5]:

$$\begin{aligned}
\text{XPD}_{H,V} &= 10 \log_{10} \frac{P_{c\mp}}{P_s} \\
&= 10 \log_{10}\left[\frac{1 + p^2 \cos^2 2\tau \mp 2p \cos\chi \cos 2\tau}{p^2 \sin^2 2\tau}\right].
\end{aligned}$$

Then the right hand side of Eq. [27] for the received power becomes

$$\begin{aligned}
(\text{RHS})_{H,V} &= 10 \log_{10} P_{c\mp} \\
&+ 10 \log_{10}[10^{-(F_T'+F_R')/10} + 10^{-(f_T''+f_R')/10}P_{c\pm}/P_{c\mp} \\
&+ 10^{-(F_T'+f_R'+\text{XPD}_{H,V})/10} + 10^{-(F_R'+f_T''+\text{XPD}_{H,V})/10}].
\end{aligned} \qquad [28]$$

A simplified relation that has been suggested[9] omits attenuation (so that $A'_H = A'_V = 0$, $P_{c\pm} = 1$, and $p = 0$ except in P_s) and does not distinguish between incident polarizations. The results obtained with these simplifications are obviously incorrect. The P_c factors introduce important corrections. Also $\text{XPD}_{H,V}$ and the double signs distinguish between linear

H and linear V incident polarization.

Case (1b)—Same as Case (1a) but with Circular Polarization

The result follows directly from Eqs. [25] and [27] upon setting $\theta = 0$ and $\tau = \pi/4$. We obtain the same expression for right- and left-hand circular polarization upon using the assumption of random phases. The formula is

$$\begin{aligned}&(P_R - P_T - G'_T - G'_R)_C + 20 \log_{10} (4\pi d/\lambda)\\&+ (A'_H + A'_V)r'/2 + 5 \log_{10} [(1 + p^2)^2 - 4p^2 \cos^2\chi]\\&= 10 \log_{10} [10^{-(F_T' + F_R')/10} + 10^{-(f_T'' + f_R')/10}\\&+ 10^{-(F_T' + f_R' + \mathrm{XPD}_c)/10} + 10^{-(F_R' + f_T'' + \mathrm{XPD}_c)/10}] \qquad [29]\end{aligned}$$

where $\mathrm{XPD}_c = -10 \log_{10} p^2$. The right-hand side of this relation agrees with the form suggested in Ref. [9]. The left-hand side of Eq. [29], however, retains the attenuation factors, A'_H, A'_V, and the log argument involving p, which are omitted incorrectly in Ref. [9].

Case (2a)—Interfering Satellite with Polarization Nearly Orthogonal to that of the Receiving Earth Antenna

The result for this case is as follows:

$$\begin{aligned}&(P_R - G'_R)_{H,V} - (P_T + G''_T)_{V,H} + 20 \log_{10} (4\pi d/\lambda)\\&+ (A'_H + A'_V)r'/2 + 5 \log_{10} [(1 + p^2)^2 - 4p^2 \cos^2\chi]\\&= 10 \log_{10} (f_{T2}^2 F_R^2 P_{c\mp} + F_{T2}^2 f_R^2 P_{c\pm} + f_{T2}^2 f_R^2 P_{s\mp} + F_{T2}^2 F_R^2 P_{s\pm})\\&= 10 \log_{10} [10^{-(f'_T + F'_R)/10} P_{c\mp} + 10^{-(F''_T + f'_R)/10} P_{c\pm}\\&+ 10^{-(f'_T + f'_R)/10} P_{s\mp} + 10^{-(F''_T + F'_R)/10} P_{s\pm}]. \qquad [30]\end{aligned}$$

When $\theta = 0$, we obtain for the right hand side

$$\begin{aligned}&(\mathrm{RHS})_{H,V} = 10 \log_{10} P_{c\mp}\\&+ 10 \log_{10} [10^{-(f_T' + F_R')/10} + 10^{-(F_T'' + f_R')/10} P_{c\pm}/P_{c\mp}\\&+ 10^{-(f_T' + f_R' + \mathrm{XPD}_{H,V})/10} + 10^{-(F_T'' + F_R' + \mathrm{XPD}_{H,V})/10}] \qquad [31]\end{aligned}$$

In the relation suggested in Ref. [9], the authors set $A'_H = A'_V = 0$, $P_{c\pm} = 1$, and $p = 0$ except in P_s. They also do not distinguish between F''_T and F'_T and between f''_T and f'_T. Their results obtained with these simplifications are incorrect.

Case (2b)—Same as (2a) but with Circular Polarization

Again we obtain this result by letting $\theta = 0$ and $\tau = \pi/4$, yielding

$$
\begin{aligned}
&(P_R - G'_R)_{R,L} - (P_T + G''_T)_{L,R} + 20\log_{10}(4\pi d/\lambda) \\
&+ (A'_H + A'_V)r'/2 + 5\log_{10}[(1+p^2)^2 - 4p^2\cos^2\chi] \\
&= 10\log_{10}[10^{-(f_T' + F_R')/10} + 10^{-(F_T'' + f_R')/10} \\
&+ 10^{-(f_T' + f_R' + \mathrm{XPD}_c)/10} + 10^{-(F_T'' + F_R' + \mathrm{XPD}_c)/10}]
\end{aligned}
\qquad [32]
$$

The right-hand side of Eq. [32] agrees with the form suggested in Ref. [9], except that Eq. [32] also includes the attenuation factors on the left-hand side.

Cross-Polarization Isolation (XPI)

The cross-polarization isolation of an antenna (XPI) against interfering signals can be derived by making use of the results for cases (1) and (2). XPI is simply the difference in dB between these two sets of results. It refers to the cross talk in a single receiver channel when two cross-polarized signals are transmitted.

For linear polarization, we find from cases (1a) and (2a) that

$$
\begin{aligned}
(\mathrm{XPI})_{H,V} &= G'_T - G''_T \\
&+ 10\log_{10}\left[\frac{F_{T1}^2F_R^2P_{c\mp} + f_{T1}^2f_R^2P_{c\pm} + F_{T1}^2f_R^2P_{s\mp} + f_{T1}^2F_R^2P_{s\pm}}{f_{T2}^2F_R^2P_{c\mp} + F_{T2}^2f_R^2P_{c\pm} + f_{T2}^2f_R^2P_{s\mp} + F_{T2}^2F_R^2P_{s\pm}}\right] \\
&= G'_T - G''_T + F'_T - F''_T \\
&+ 10\log_{10}\left[\frac{D_R^2P_{c\mp} + D_{T1}^{-2}P_{c\pm} + P_{s\mp} + D_R^2D_{T1}^{-2}P_{s\pm}}{D_R^2D_{T2}^{-2}P_{c\mp} + P_{c\pm} + D_{T2}^{-2}P_{s\mp} + D_R^2P_{s\pm}}\right]
\end{aligned}
\qquad [33]
$$

where $D_{T1} = F_{T1}/f_{T1}$, $D_{T2} = F_{T2}/f_{T2}$, and $D_R = F_R/f_R$ are the discrimination ratios of the antennas in clear weather.

For circular polarization, we find from cases (1b) and (2b) that

$$
\begin{aligned}
(\mathrm{XPI})_C &= G'_T - g''_T \\
&+ 10\log\left[\frac{F_{T1}^2F_R^2 + f_{T1}^2f_R^2 + p^2(F_{T1}^2f_R^2 + f_{T1}^2F_R^2)}{f_{T2}^2F_R^2 + F_{T2}^2f_R^2 + p^2(f_{T2}^2f_R^2 + F_{T2}^2F_R^2)}\right] \\
&= G'_T - G''_T + 10\log_{10}\left[\frac{F_{T1}^2 + f_{T1}^2}{F_{T2}^2 + f_{T2}^2}\right] \\
&+ 10\log_{10}\left[\frac{1 + \left(\frac{1-p^2}{1+p^2}\right)\frac{4E_{T1}E_R}{(1+E_{T1}^2)(1+E_R^2)}}{1 - \left(\frac{1-p^2}{1+p^2}\right)\frac{4E_{T2}E_R}{(1+E_{T2}^2)(1+E_R^2)}}\right].
\end{aligned}
\qquad [34]
$$

Similar to Eq. [9], we introduce here E'_T, E''_T and E'_R, the clear-weather ellipticity values in dB of the circularly polarized antennas. We then define

$$E_{T1} = 10^{E'_T/20}, \quad E_{T2} = 10^{E''_T/20}, \quad E_R = 10^{|E'_R|/20},$$
$$\frac{F_{T1}}{f_{T1}} = \frac{E_{T1}+1}{E_{T1}-1}, \quad \frac{F_{T2}}{f_{T2}} = \frac{E_{T2}+1}{E_{T2}-1} \text{ and } \frac{F_R}{f_R} = \frac{1+E_R^{-1}}{1-E_R^{-1}} \qquad [35]$$

These parameters are used in Eq. [34].

Cross Polarization Discrimination (XPD)

XPD refers to the discrimination of received signals between the two cross-polarized receiver channels when a single dominantly polarized signal is transmitted. We now provide results similar to those in Sec. 2.2 (using the same notation as there) but based on the random phase limit.

For linear polarization, we have

$$(\mathrm{XPD})_{\mathrm{H,V}} = G'_R - G''_R + f'_R - f''_R + 10\log_{10}\left[\frac{D_{R1}^2 P_{c\mp} + D_T^{-2} P_{c\pm} + P_{s\mp} + D_{R1}^2 D_T^{-2} P_{s\pm}}{D_{R2}^2 D_T^{-2} P_{c\pm} + P_{c\mp} + D_T^{-2} P_{s\pm} + D_{R2}^2 P_{s\mp}}\right]. \qquad [36]$$

For circular polarization, the result is obtained by letting $\theta = 0$ and $\tau = 45°$ in the above, giving

$$(\mathrm{XPD})_{\mathrm{C}} = G'_T - G''_T + 10\log_{10}\left[\frac{F_T^2 F_{R1}^2 + f_T^2 f_{R1}^2 + p^2(F_T^2 f_{R1}^2 + f_T^2 F_{R1}^2)}{f_T^2 F_{R2}^2 + F_T^2 f_{R2}^2 + p^2(f_T^2 f_{R2}^2 + F_T^2 F_{R2}^2)}\right], \qquad [37]$$

where $f_{R1} = 10^{f_R'/20}$ and $f_{R2} = 10^{f_R''/20}$.

Numerical Example and Comparison with an Experiment

Fig. 1, from Oguchi,[10] shows measured values of $\mathrm{XPD_{H,V}}$ and $\mathrm{XPI_{H,V}}$ versus attenuation (relative to clear-weather conditions), obtained by Shimba et al[11] at 20 GHz over a terrestrial path. At sufficiently large attenuations, the order of the curves is $\mathrm{XPD_V} > \mathrm{XPI_V} > \mathrm{XPI_H} > \mathrm{XPD_H}$. We have performed computer calculations based on the random phase limit (with $F_{R1} = F_{R2}$, $f_{R1} = f_{R2}$, $F_{T1} = F_{T2}$, $f_{T1} = f_{T2}$, $G_{T1} = G_{T2}$, and $G_{R1} = G_{R2}$) using the equations in this section. We adopt the differential

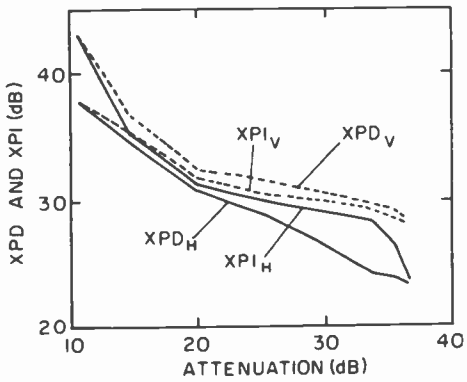


Fig. 1—Variation of $XPD_{H,V}$ and $XPI_{H,V}$ values with signal attenuation.[10,11]

phase and attenuation results of Oguchi and Hosoya[2] at 19.3 GHz. We find that the above order of the four curves occurs for exact alignment when $D_R' > D_T'$, i.e., when the clear weather isolation of the receiver is better than that of the transmitter. Otherwise, when $D_T' > D_R'$, the order is $XPI_V > XPD_V > XPD_H > XPI_H$. The best match to the experimental

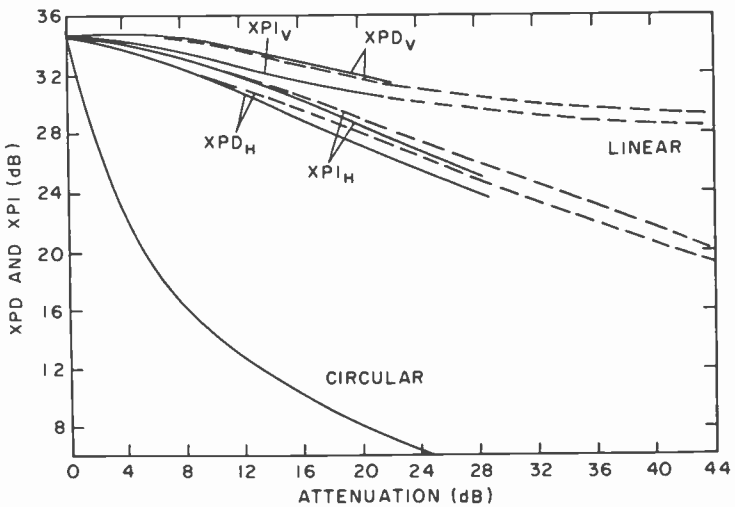


Fig. 2—Theoretical results to compare with Fig. 1. Here, $\alpha' = 90°$, $\tau' = 2°$, $\theta' = 0°$, frequency = 19.3 GHz, $D_R' = 45$ dB, and $D_T' = 35$ dB. Solid curves are for path length through rain given by H in Eq. [1]; dashed curves are for H double that in Eq. [1].

results is obtained with a canting angle of 2°, and by letting D'_R = 45 dB and D'_T = 35 dB. If the canting angle is larger, the ordinate values are too low, and if the difference between D'_T and D'_R is less, the four curves are too close together. The numerical results are plotted in Fig. 2. The solid curves are based on the H value given in Eq. [1], and the dashed curves are calculated with twice this path length. A plot of XPD and XPI versus attenuation is not very sensitive to the value of H. Good agreement is obtained for the magnitude, shape, and separation of the experimental curves for the higher attenuation range above 15 dB. We thus see that theory can act as a diagnostic on the experimental curves, providing an indication of whether $D'_R > D'_T$ or vice versa, of the difference between D'_R and D'_T, and of the magnitude of the effective canting angle.

The theoretical curves for circular polarization are also plotted in Fig. 2. Circular polarization is much worse than linear polarization at 20 GHz.

2.4 Considerations of Linear Versus Circular Polarization

We first discuss considerations in the use of circular polarization and then those of linear polarization. Circular polarization provides the following advantages:

(a) There is no need for polarization tracking. By contrast, linear polarization requires that adequate alignment be maintained between the polarization directions of the satellite and earth antennas.
(b) Circular polarization is insensitive to Faraday rotation. For linear polarization, rotation of the polarization angle introduces cross polarization. This is an important consideration at frequencies below 6 GHz. At frequencies above 10 GHz, however, Faraday rotation is negligible and this argument is no longer pertinent.
(c) Two adjacent areas serviced by separate satellites can maintain orthogonality using orthogonal right- and left-hand circular polarizations, respectively. This is not so for linear polarization. This problem is discussed in Sec. 3.
(d) The angle of polarization of the antenna may vary over its design bandwidth.[7] This presents a possible disadvantage in a frequency reuse system employing linear polarization where the orthogonal polarization is also used. Then, the overall isolation may not be maintained over the frequency bandwidth. This is not of concern if the isolation is sufficient throughout. Due to the variation in the angle of polarization, however, methods proposed to cancel out

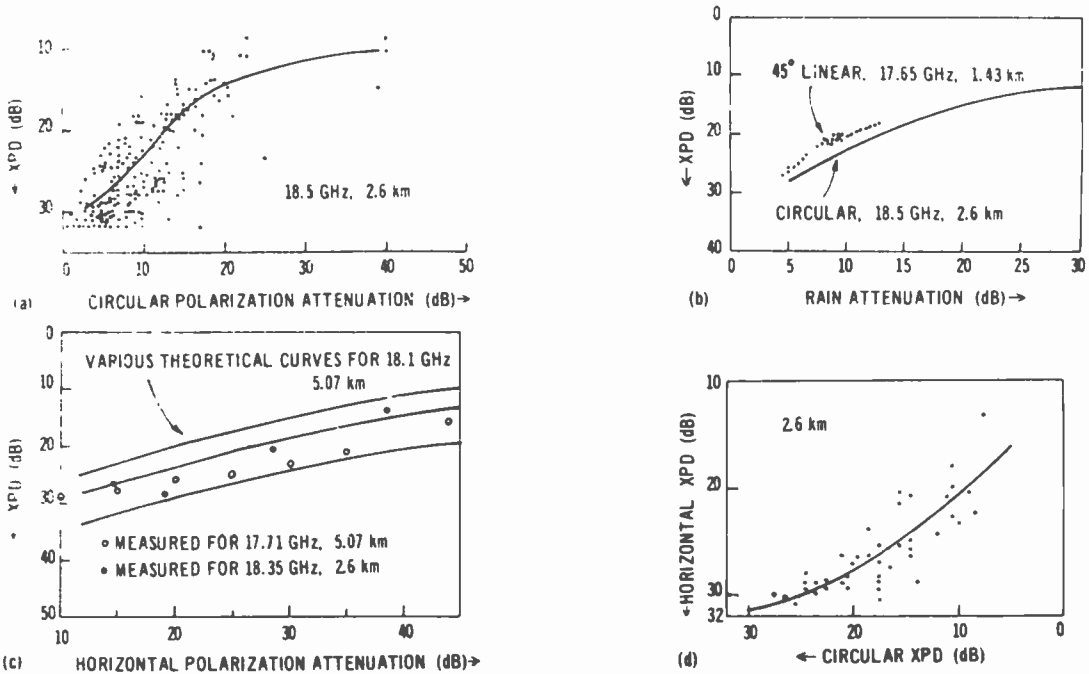


Fig. 3—Comparison of linear and circular polarizations: (a) theory versus experiment for circular polarization;[13] (b) experiments with 45° linear and circular polarizations;[14,15] (c) theory versus experiment for horizontal polarization;[3] and (d) experiments with horizontal and circular polarizations[16] with respective frequencies 18.35 and 18.65 GHz.

cross-polarization components induced by rain may not be possible in the r-f stage over the entire bandwidth, but will be limited to the i-f stage, separately for each transponder.

In the 4–6 GHz band, (a) and (b) are the main advantages of circular polarization. In the 12 GHz band, (a) and (c) are the main advantages.

The main advantage of linear polarization arises from considerations on propagation effects through rain. We denote by XPD the cross-polarization discrimination or the ratio of direct-to cross-polarized received power in dB. Depolarization due to rain is less and XPD is greater for both vertical and horizontal linear polarizations than it is for circular polarization, except for a canting angle of 45°, where circular and linear polarizations give about the same XPD. The theory, given in Sec. 2.1 above, indicates this. Since a distribution of canting angles exists, averaging over the canting angles favors linear polarization for most of the time.[12] A vertically polarized electric field experiences better XPD and less attenuation than a horizontally polarized field and both provide better XPD than circular polarization.

Experimental evidence substantiates the theory. Fig. 3 illustrates experimental results over terrestrial paths in the 17–19 GHz band. The top left is a plot of theory versus experiment using circular polarization.[13] The XPD value is plotted versus attenuation of the co-polarized signal. The attenuation is the excess over the clear weather value. The top right[14] compares the same experimental results with another experiment[15] using 45° linear polarization. We see that a 45° polarized wave is more or less equivalent to circular polarization, as theory (see Sec. 2.1) predicts for antennas with good clear-weather isolation. The plot on the bottom left compares theory and experiment for an incident horizontally polarized wave for various path length.[3] Finally, the plot in the bottom right shows the experimental XPD's for circular polarization replotted against those for horizontal polarization and for identical path lengths.[16] We clearly see that depolarization is worse for an incident circular polarization.

Fig. 4 illustrates experimental results obtained at 20 GHz with the ATS-6 satellite[14] and comparison with theory.[3] The polarization angle for these results is between 20° and 22° with respect to the vertical. For these polarization angles, the results lie as expected between theoretical predictions of XPD for vertical polarization (corresponding to 0°) and circular polarization (equivalent to 45°). This again indicates distinct advantages in the use of linear over circular polarization, even when the polarization angle is as large as 20°.

We have mentioned above that if the polarization angle at the beam

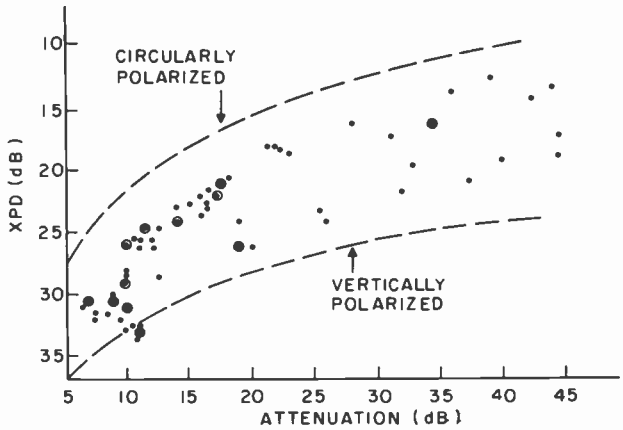


Fig. 4—Bell Laboratories measurements[14] on cross-polarization versus attenuation of a 20-GHz signal from ATS-6. The incident linear polarization is oriented 20° from the plane containing the local vertical. Curves are calculated values. Circled dots represent multiple points.

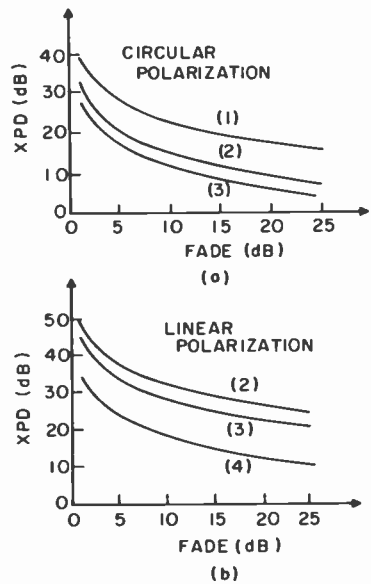


Fig. 5—Calculated[17] XPD versus attenuation for a satellite-to-earth link at 11 GHz. Satellite elevation angle: (1) 60°; (2) 40° (beam center); (3) 20° (beam edge, north); (4) 40° (beam edge, east-west, and polarization angle = 17°).

center is vertical or horizontal, it is different at the beam edges, and consequently XPD at the beam edges is degraded from that at beam center. In contrast, for circular polarization, XPD doesn't vary much in the east-west direction at a more or less constant elevation angle. One should not infer from this that circular polarization is better. As seen from Fig. 4, XPD for all finite polarization angles other than 45° is better for linear polarization than for circular polarization.

Fig. 5 gives theoretical predictions in relation to the planned European Orbital Test Satellite (OTS) operating in the 11 to 12 GHz band,[17] based on the theory in Sec. 2.1. This graph plots XPD versus attenuation using the Eurobeam A antenna on this particular satellite. The left-hand calculations are for circular polarization and the right-hand ones are for linear polarization, presumably horizontal. The canting angle on the beam axis is assumed to be 4°. Different curves result depending on the elevation angle and on the depolarization angle determined by the offset of the earth station from beam center. The best XPD values occur for the largest elevation angle, with the receiver on beam axis and with the beam using linear polarization. Curve (1) shows results for an elevation angle of 60° with circular polarization. Curves (2) for a 40° elevation angle compare linear to circular and show that circular is worse than linear polarization. The same would be true for 60° or any other angle. Curves (3) are for a 20° elevation angle, which occurs towards the beam edge in the north, and the XPD values are now lower for both polarizations. At 40° elevation, the polarization angle at the receiver changes from linear horizontal at beam center to 17° with respect to horizontal as one moves in the east-west direction to the 3-dB beam edge. This further decreases XPD for linear polarization as indicated by curve (4). Curve (2) still applies for circular polarization, but this curve, as implied above, is still worse than curve (4) for linear polarization.

Consider the following example for Canada. Let a satellite be situated at about 100°W longitude, radiating vertically polarized waves in the meridian plane. The polarization angle covering all of Canada with a single beam would vary by about ±30° with respect to the plane on earth containing the local vertical. With several spot beams, this variation in polarization angle can be greatly reduced and in all cases, it is below the 45° equivalent of circular polarization.

For reference, the following formula can be used with good accuracy to calculate the polarization angle ξ. Let the satellite be situated at longitude L_{os}. Let the on-axis beam from the satellite antenna intersect the earth at latitude L_{oa} and longitude L_{oa}. Define $L_{da} = L_{oa} - L_{os}$. Also let the off-axis ray intersect the earth at latitude L_a and longitude L_o and define $L_d = L_o - L_{os}$. The polarization angle ξ, as measured from that at beam center, is then given by

$$\tan\xi = \frac{\sin L_{aa} \cos L_a \sin L_d - \cos L_{aa} \sin L_{da} \sin L_a}{\sin L_{aa} \sin L_a - \cos L_{aa} \sin L_{da} \cos L_a \sin L_d} \qquad [38]$$

In particular when $L_{da} = 0$, we obtain $\tan\xi = \cot L_a \sin L_d$.

We can summarize the advantages of linear polarization as follows:

(a) Both from terrestrial links and satellite-to-earth links and from theoretical predictions and experimental data, linearly polarized systems perform better than circular ones in the presence of rain and provide better XPD.

(b) A polarizer is required to generate circular polarization and this device introduces ellipticity. The isolation for the antenna cannot be made better than say 33 dB if the polarizer ellipticity is 0.4 dB. This advantage is not appreciable, since 33 dB isolation is sufficient in many cases. (The design of high performance polarizers is given by Soma et al.[18]) Furthermore, linearly polarized ground antennas may also employ an adjustable polarizer (instead of using mechanical means) in order to align the receiver electric field direction to that of the incoming wave.

(c) In the off-axis directions and in the side lobes, the antenna clear-weather isolation is better for linear than for circular polarization, provided alignment can be maintained.[7] Also, cross-polarization for linear polarization is usually concentrated in four lobes in the 45° off-axis planes, whereas for circular polarization, circular symmetry is maintained and the depolarization covers a larger area.[17,19]

Let us reconsider the previously mentioned disadvantages of linear polarization. Misalignment remains a problem as far as the earth station is concerned, where it can be caused by winds or mishandling. A misalignment of 2° decreases XPD to 29 dB and 8° decreases it to 17 dB. Occasional checking of the earth antenna alignment may be necessary. As far as the satellite is concerned, it is generally maintained that its attitude control can be kept to within 0.25°, so that a variable polarization should not occur. If problems arise due to improper station keeping, the plane of polarization will change as well, presenting problems both to linear and circular polarization systems.

To overcome the cross-talk problem mentioned under advantages of circular polarization, one of the proposed solutions (see Sec. 3) is to use alternate frequency channels as well as orthogonal polarizations in adjacent beams. By the use of different frequencies, discrimination can be maintained even at the beam edges.

The OTS is being designed with several antenna systems[20]. Two are the Eurobeam A beam and the steerable spot beam, both of which use two orthogonally polarized linear waves, and another is the Eurobeam

B beam, which uses a single circularly polarized wave. The reason for including the latter is the uncertainty presented by polarization misalignment with linear polarization. One of the purposes of CTS is to enable a comparison between the two types of polarization.

As discussed above, at frequencies above 10 GHz, propagation effects favor linear polarization. Misalignment problems favor circular polarization. We feel that the advantages of linear polarization outweigh the disadvantages if the number of earth stations that have to be serviced occasionally for alignment are not too many. Watson and Soutter[17] also show that linearly polarized systems perform significantly better at 12 GHz in the presence of rain and offer greater capacity with multi-level PSK modulation. This conclusion, however, has to be qualified. In the future, the situation may arise where ground stations may proliferate greatly with a multitude of individual home receivers. Proper alignment cannot be expected to be performed or maintained by all the operators. Then the economics of servicing all these stations will be a significant consideration. Circular polarization may be the easy way out of this problem, even at the expense of fading and depolarization occurring during precipitation. However, a better solution is to use linear polarization and insert easily adjustable polarizers to align the polarizations.

2.5 Reference Antenna Patterns

Various reference pattern envelopes have been proposed[21] for each of the co-polarization and cross-polarization components. In addition, different patterns apply to satellite transmitting antennas and to individual receiver antennas.

It is difficult to suggest reference envelopes that apply to all different situations and types of antennas, since each type of antenna has a somewhat different envelope decay. The co-polarization pattern envelope also depends on whether methods are used for side-lobe control, such as special designs of the aperture edges. The cross-polarization pattern envelope also differs on the means taken to suppress or filter the cross-polarization component, such as grid reflectors, long-focal-length paraboloids, or optimum modes for feeding the dish which cancel out the cross-polarization pattern. Blockage obstacles and antenna supports alter the envelope shapes and cause spurious side lobes. Furthermore, the envelopes in the E- and H-planes can differ, so that any reference has to include both. Also the pattern envelopes vary depending on the frequency within the bandwidth.

In view of all these variables, at best one can only propose reference

envelopes for the average type of antenna and require actual performances to surpass these requirements by being below the envelopes. Even so, it is conceivable for the actual cross-polarization pattern in the side lobes to be above the co-polarization pattern, for example at null points in the co-polarization pattern.

Satellite Co-Polarization Pattern Envelope

Suggestions for the co-polarization pattern envelope of satellite antennas have evolved through various stages. One report[22] suggests the following for fixed service satellite antennas beaming power towards large earth terminals.

$$\begin{aligned} F'_T(\phi) &= -3(2\phi/\phi_0)^2 \quad \text{for} \quad 0 \leqslant \phi/\phi_0 \leqslant 1.29 \\ &= -20 \quad \text{for} \quad 1.29 \leqslant \phi/\phi_0 \leqslant 3.15 \\ &= -25 \log_{10}(2\phi/\phi_0) \quad \text{for} \quad 3.15 \leqslant \phi/\phi_0 \end{aligned} \qquad [39]$$

with a limit at the isotropic gain. In the above, $F'_T(\phi)$ is the gain in dB relative to its on-axis value, namely $F'_T(0) = 0$. Also, ϕ is the angle away from the beam axis and ϕ_0 is the total beamwidth, so that $F'_T(\pm\phi_0/2) = -3$ dB.

In another report[23], referring to broadcast service satellites beaming power towards individual earth receivers, a different reference envelope is proposed, namely

$$\begin{aligned} F'_T(\phi) &= -12(\phi/\phi_0)^2 \quad \text{for} \quad 0 \leqslant \phi/\phi_0 \leqslant 0.50 \\ &= -10.5 - 25 \log_{10}(\phi/\phi_0) \quad \text{for} \quad 0.50 \leqslant \phi/\phi_0 \leqslant 0.82 \\ &= -20 - 135 \log_{10}(\phi/\phi_0) \quad \text{for} \quad 0.82 \leqslant \phi/\phi_0 \leqslant 1.09 \\ &= -25 \quad \text{for} \quad 1.09 \leqslant \phi/\phi_0 \leqslant 3.80 \\ &= -10.5 - 25 \log_{10}(\phi/\phi_0) \quad \text{for} \quad 3.80 \leqslant \phi/\phi_0 \end{aligned} \qquad [40]$$

with a limit at the isotropic gain. This envelope, representative of normal designs with lobe control, includes a Gaussian pattern for near zero angles.

A recent report[21] suggests having a single reference envelope that includes antenna patterns for both fixed and broadcast service satellites. Towards this end, the following simple envelope is proposed for the co-polarization pattern.

$$\begin{aligned} F'_T(\phi) &= -12(\phi/\phi_0)^2 \quad \text{for} \quad 0 \leqslant \phi/\phi_0 \leqslant 1.44 \\ &= -25 \quad \text{for} \quad 1.44 \leqslant \phi/\phi_0 \leqslant 3.80 \\ &= -10.5 - 25 \log_{10}(\phi/\phi_0) \quad \text{for} \quad 3.80 \leqslant \phi/\phi_0 \end{aligned} \qquad [41]$$

with a limit at the isotropic gain. It is seen that the envelope pattern for the fixed service satellite is adopted up to ϕ/ϕ_0 greater than unity, and the previous broadcast satellite envelope is adopted for ϕ/ϕ_0 greater than 1.44. In many cases, one can do better than required by this envelope, especially if precautions are taken to reduce side lobes. This final suggestion nonetheless looks reasonable for most types of antennas.

Satellite Cross-Polarization Pattern Envelope

Different reference cross-polarization envelopes apply depending on whether special care is taken to increase the antenna isolation. Various methods for increasing the isolation are discussed by Raab[24] and one method using linearly polarized filtering screens has been applied to the RCA-SATCOM satellite. With care, the reference[21] envelope given below should be attainable. The satellite cross-polarization envelope in dB, denoted by $f_T'(\phi)$, is given relative to $F_T'(0) = 0$, by

$$\begin{aligned} f_T'(\phi) &= -36 - 25 \log_{10} \left| \frac{\phi}{\phi_0} - 1 \right| \quad \text{for} \quad 0 \leq \phi/\phi_0 \leq 0.42 \\ &= -30 \quad \text{for} \quad 0.42 \leq \phi/\phi_0 \leq 1.58 \\ &= -36 - 25 \log_{10} \left| \frac{\phi}{\phi_0} - 1 \right| \quad \text{for} \quad 1.58 \leq \phi/\phi_0 \end{aligned} \qquad [42]$$

with a limit at the isotropic gain. If no special effort is taken to minimize the cross-polarization component, Reference [21] suggests that the factor 25 be increased to 40; the corresponding ϕ/ϕ_0 limits would then be 0.29 and 1.71.

Our opinion is that there is a greater difference in the envelopes between the two cases whether the cross-polarization is suppressed or not by additional means. With a solid dish, it is difficult to obtain 30-dB isolation up to the 3-dB points ($\phi/\phi_0 = 0.5$) over the whole frequency bandwidth, although it may occur at specific frequencies. A gridded reflector can make a big difference, even though the number of reflectors has to be doubled in a dual-polarized system. Only one reflector can be used per polarization. With grids, one can filter the cross-polarization level, theoretically to 50-dB isolation, but practically to about 36 dB up to the 3-dB points. We thus feel that with gridded systems one can do better than the reference envelope shown. Without special care, the cross-polarization envelope is usually worse than the modified expressions suggested above.

A possible modification that we suggest is to alter the middle part of Eq. [42] from −30 to −33 dB and change the corresponding ϕ/ϕ_0 limits

to 0.24 and 1.76. This applies to a proper design minimizing cross-polarization.

A good antenna should display a deep minimum in cross-polarization on axis. This is not always observed.[7] Sometimes, this failure is not inherent in the antenna but is due to an imperfect measuring antenna used to derive the cross-polarization pattern of the test antenna. Also, scatter from supporting structures may actually fill in the minimum. Both the magnitude in dB of the minimum and its position in space may also vary as the frequency is changed. Nonetheless, we agree with Eq. [42]. Although it shows only a relatively small increase in cross-polarization from on-axis to $\phi/\phi_0 = 0.5$, it still suggests to the designer the desire to obtain even a lower on-axis minimum.

For large off-axis angles beyond the second side lobe, it is questionable whether the envelopes for direct- and cross-polarization have any meaning. First the symmetry about the axis is not maintained. Secondly, in relation to isolation considerations, the power in cross-polarization may surpass the co-polarization power at minimum points in the co-polarization pattern. Nonetheless, the reference pattern envelopes serve to indicate the need to decrease the envelopes of the side lobes for both direct- and cross-polarization to be sufficiently below their values at $\phi/\phi_0 = 0.5$. These references should be kept for this reason but, perhaps, made constant at the values attained at about the position of the third side lobe.

Individual Receiver Co-Polarization Pattern Envelope

For individual reception, the following co-polarization pattern envelope has been suggested[23] for $F'_R(\phi)$, the receiver gain in dB, relative to its on-axis value of $F'_R(\phi) = 0$:

$$F'_R(\phi) = -9 - 20 \log_{10} (\phi/\phi_0) \quad \text{for} \quad 0.5 \leqslant \phi/\phi_0 \leqslant 11.22$$
$$= -30 \quad \text{for} \quad 11.22 \leqslant \phi/\phi_0 \qquad [43]$$

with a limit at the isotropic gain. For community reception where the antenna is designed with better side-lobe suppression, the above is replaced by

$$F'_R(\phi) = -10.5 - 25 \log_{10} (\phi/\phi_0) \qquad \text{for} \qquad 0.5 \leqslant \phi/\phi_0. \qquad [44]$$

In a later proposal,[21] it has been suggested that the above relations for individual receivers be retained only for $\phi/\phi_0 \geqslant 0.707$. For small off-axis angles, a flat-top is allowed up to $\phi/\phi_0 = 0.25$, in order to account for possible pointing errors of $\pm 0.5°$, and a Gaussian is adopted for ϕ/ϕ_0 between 0.25 and 0.707. This gives the following for the reference envelope:

$$F'_R(\phi) = 0 \quad \text{for} \quad 0 \leq \phi/\phi_0 < 0.25$$
$$= -12(\phi/\phi_0)^2 \quad \text{for} \quad 0.25 < \phi/\phi_0 \leq 0.707$$
$$= -9 - 20 \log_{10}(\phi/\phi_0) \quad \text{for} \quad 0.707 \leq \phi/\phi_0 \leq 11.22$$
$$= -30 \quad \text{for} \quad 11.22 \leq \phi/\phi_0 \qquad [45]$$

with a limit at the isotropic gain.

It seems to us that the proposed envelope, which decreases as $-9 - 20 \log_{10} \phi/\phi_0$ is more representative of an antenna having a specially designed uniform illumination over the aperture. Usual antennas have side lobes somewhat lower than the envelope and then the envelope decreases more rapidly. We would suggest consideration of the following modification replacing the last two dependences.

$$F'_R(\phi) = -9.52 - 23.4 \log_{10}(\phi/\phi_0) \quad \text{for} \quad 0.707 \leq \phi/\phi_0 \leq 7.5$$
$$= -30 \quad \text{for} \quad 7.5 \leq \phi/\phi_0 \qquad [46]$$

Individual Receiver Cross-Polarization Pattern Envelope

A form for $f'_R(\phi)$, the cross-polarization pattern envelope in dB relative to $F'_R(0) = 0$, has recently been suggested. It allows for the fact that the minimum on-axis value of -30 dB may be difficult to obtain in view of possible pointing errors, and that it may be partly filled in if no special care is taken. The suggested envelope is[21]

$$f'_R(\phi) = -25 \quad \text{for} \quad 0 \leq \phi/\phi_0 \leq 0.25$$
$$= -30 - 40 \log_{10} |\phi/\phi_0 - 1| \quad \text{for} \quad 0.25 \leq \phi/\phi_0 \leq 0.44$$
$$= -20 \quad \text{for} \quad 0.44 \leq \phi/\phi_0 \leq 1.40$$
$$= -30 - 25 \log_{10} |\phi/\phi_0 - 1| \quad \text{for} \quad 1.40 \leq \phi/\phi_0 \leq 2.0$$
$$= -30 \quad \text{for} \quad 2.0 \leq \phi/\phi_0 \qquad [47]$$

with a limit at the isotropic gain. We generally agree with this form of the pattern.

Our previous remarks on the far off-axis patterns for satellites also apply here. To improve on the cross-polarization pattern for linear polarization up to $\phi/\phi_0 = 1.4$, one should use a long-focal-length paraboloid. With good design, both the co-polarization and cross-polarization envelopes can be improved.

A final suggestion is to require that the reference envelopes be met throughout the frequency bandwidth and for both E- and H-planes.

3. Systems and Implementation Problems

The "broadcast", or broadcast service, satellite is intended to provide radio or television service to many small inexpensive earth terminals. This requires a very high EIRP on the spacecraft provided by high-power transponders and high-gain antennas. By way of contrast, fixed service communications satellites provide a low EIRP to strategically located, large, high-gain, expensive earth terminals. Some of the problems associated with providing the high EIRP required of broadcast service satellites are discussed in the literature.[8,22–34] A few additional problems are discussed here.

3.1 Limitation to Antenna Beam Sizes

In designing a spacecraft antenna, allowance must be made for antenna beam-pointing errors, that is, the antenna beam must be made larger, by an amount $\pm\Delta\theta$, than the minimum required to illuminate the desired ground area. This allows the antenna to be misdirected by the amount $\Delta\theta$ and still keep the ground coverage area within the edge contour of the antenna beam. However, because the beam dimensions are larger than the minimum, the gain at the edge contour will be reduced.

If it is desired to cover a large ground area with a series of adjacent spot beams, then the parameter that must be maximized is the product of coverage area and antenna gain. For a circular beam, it can be shown that the gain is inversely proportional to the square of the 3-dB beam width (θ_3). The coverage area, reduced by the pointing errors, is proportional to $(\theta_3 - 2\Delta\theta)^2$ as shown in the insert in Fig. 6. Thus

$$\text{product} \propto \frac{1}{{\theta_3}^2}(\theta_3 - 2\Delta\theta)^2$$

$$= 1 - 4\frac{\Delta\theta}{\theta_3} + 4\left(\frac{\Delta\theta}{\theta_3}\right)^2; 0 < \Delta\theta < \frac{\theta_3}{2}. \qquad [48]$$

In the limits, $\Delta\theta$ cannot be less than zero and must not be greater than $\theta_3/2$.

A plot of Eq. [48] is shown in Fig. 6. It shows that in the case of zero pointing errors, the limit is unity. If the pointing errors increase to one tenth the beam width, there is already a loss in coverage area (or gain) of 36%, or about 2 dB. This is probably the largest ratio of $\Delta\theta/\theta_3$ that can be tolerated. That is, for a pointing error of $\pm 0.2°$, the minimum beam width should not be smaller than 2°. Since this is approximately the state of the art in spacecraft attitude control, any smaller beam size must be accompanied by antenna fine-steering capabilities to remove the attitude

errors of the spacecraft. The actual trade-off point between fixed antennas and steerable antennas must be determined by detailed trade-off studies in each case. Since some of the antenna pointing errors are a result of thermal distortion in the antenna itself, an appropriate means of antenna pointing is r-f tracking of a ground beacon (located on the axis of the spacecraft antenna beam) using a monopulse feed incorporated into the actual antenna. The dominant source of pointing error should then be the signal-to-noise ratio of the beacon, and a considerable reduction in pointing error should result. There should be no intractable technical problems in implementing a tracking antenna, with the possible

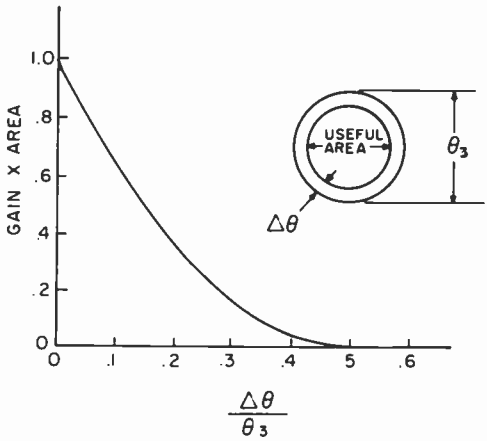


Fig. 6—Gain coverage product versus antenna pointing errors normalized to the 3-dB beam width.

exception of the reliability of the steering mechanism for a seven-year mission. There is however a logistics problem associated with placing a ground beacon in an appropriate location.

Another limitation on the antenna beam size is the shroud dimensions. The 8-foot Thor-Delta shroud has an inside clear diameter of approximately 7 feet. An antenna aperture limited to this diameter will have a beam width of just under 1° at 12 GHz. The 10-foot Atlas Centaur shroud would contain an antenna with a beam width of about 0.75° at 12 GHz. To obtain narrower beam widths, it would be necessary to build the antenna with fold out sections. The usual type of deployable antenna used in space is a fold-out mesh reflector suitable for low frequencies. The mesh stretches straight between the ribs introducing a surface error. To make the errors tolerable at 12 GHz, the number of ribs must be in-

creased, and it is anticipated that the surface errors introduced by the mesh would limit the beam size to approximately 0.5°. Only one such reflector could be carried, but it could be used for a number of spot beams by means of more than one feed horn.

3.2 System Implications

Before considering the different antenna types and their suitability, it is necessary first to consider the overall system and to establish the antenna requirements. The discussion that follows is not intended to be definitive but to indicate the direction a broadcast satellite system might take. In addition, it outlines the context in which the antenna discussion must be viewed.

We will assume that the broadcast satellite system requires an eirp at 12 GHz of 58 dBW, equivalent to that of CTS with the 200-watt TWTA. CTS uses a steerable 2.5°-beam-width antenna. The equivalent EIRP can be obtained with higher gain antennas and lower power TWTAs as indicated in Table 3.

Table 3—Gain/Power Requirements to Obtain EIRP of 58 dBW at 12 GHz

Beam Width (deg)	Gain (dB)	Power (watts)
2.5	36	200
2.1	37.5	142
1.75	39	100
1.5	40.4	72
1.0	44	32
0.75	46.5	18
0.5	50	8

To avoid the complication of antenna steering, a beam width of about 2° is chosen with a corresponding TWTA power of 120 watts. This matches the power of the Hughes TWT being built for the Japanese Broadcast satellite. Using a 3914 Thor-Delta with 2000 lbs launch capability, the transponder would consist of two active TWTAs, each with a redundant unit.

Two TWTAs would operate in daylight and one during eclipse. Four 2° antenna beams are needed to cover Canada, and the problem is to cover the country with a spacecraft with only 2 active TWTA's, taking into consideration the necessity for spares.

We assume that it is necessary to provide service uniformly to the whole country rather than only to the populous areas. It is necessary then

to launch two active spacecraft plus a spare, all of which must be identical. The two active satellites can be placed in the same slot (so that a single uplink antenna may be used) with a negligible probability of collision. Each spacecraft must carry antenna feeds for all four beams with facilities for switching the TWTA to the desired beam. There is space within the Thor-Delta shroud to provide two reflectors of the required size so that beams 1 and 3 could be provided by one reflector and beams 2 and 4 by the other. The minimum-weight solution is to use one reflector for all four beams and four adjacent feed horns. The single reflector

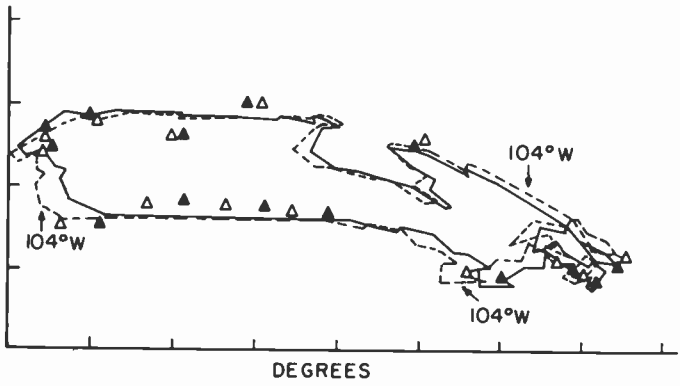


Fig. 7—The apparent size and shape of Canada as viewed from 104° and 114° west including an attitude bias to optimize antenna gain. Antenna contours are not shown; △ indicate locations of cities for 104° W and ▲ locations for 114° W.

approach would probably be followed in spite of a problem in overlapping the beams sufficiently to provide continuous coverage.

One of the problems associated with fixed-beam antennas is that of operating from different orbital locations. When the satellite is moved from one orbital location to another, the shape and size of the ground coverage area changes as viewed from the satellite. In addition, the ground area moves from west to east as the satellite is moved from east to west and vice-versa. This latter effect is accommodated by biasing the attitude of the satellite to point the antenna axis at the apparent center of the coverage area. To accommodate the change in size and shape, it is customary to oversize the antenna beam to include the whole coverage area for all expected orbit locations. This would also be done in the case of a multibeam antenna system, but an additional problem appears at the boundary between adjacent beams. Fig. 7 shows the outline of Canada as it appears when viewed from two different slot locations. The

spacecraft attitude has been biased in both cases to optimize the coverage at the east and west extremities of the country as specified by the city locations. For this reason the biasing is sensitive to the antenna beam shape and would change slightly for a different antenna beam configuration. The trend is very obvious however. The southern part of the country has a larger apparent shift than the northern part. The change in attitude bias approximately compensates for the shift at the northern continental coast and more than compensates for the shift in the arctic islands. Individual cities in the south move by nearly 0.5° in spacecraft coordinates. This is sufficient for a large fraction of the country to move from one spot beam to an adjacent spot beam when switching to an alternate satellite located 10° away in the orbital arc. This would mean that the ground station would have to switch channel numbers if the alternate spacecraft is to be used, implying additional cost in the ground terminal. An obvious solution is to keep the operational spacecraft close together in the orbital arc thus keeping this effect to a minimum.

3.3 Polarization Alignment of Adjacent Satellites

Assume that it is required by mutual agreement that all the spacecraft at 12 GHz have linear polarization aligned with the north-south spin axis of the earth. In the case of pencil beams, such as we are considering, the beam axis can be tilted towards the north or the south so as to optimally cover the desired ground area. Then, the polarization vector is not precisely aligned with the earth's axis but is tilted by the same amount as the antenna. Different spacecraft around the orbital arc all pointing at the same latitude have polarization vectors that appear as on the surface of a cone with half angle equal to the tilt angle. Thus two satellites at different locations in the orbital arc but illuminating the same ground area will have a polarization misalignment due to the tilt angle. This misalignment is calculated and presented in Fig. 8 as a function of angular separation between the cooperating satellite and the interfering satellite. The lower curve is for zero attitude error and the upper curve for an attitude error (for the interfering satellite) of 0.25°.

Also shown in Fig. 8 is the level of interference (in dB) resulting from the polarization error. This isolation is in addition to that offered by the off-axis response of the ground antenna. Since all the satellites illuminating the same ground area would normally be located within 10° of arc, the polarization isolation is 32 dB plus the off-axis response of the ground antenna to cross-polarized radiation. This is considered to be a negligible source of interference provided the polarization vector on the spacecraft is nominally aligned with the earth's spin axis.

3.4 Spacecraft Antennas[22]

Cassegrain Reflector

The standard center-fed Cassegrain geometry is very good for large antennas. For smaller antennas with beam widths of 1° or more, the subreflector is either too small to form a good reflecting surface or so large that the blockage effects are excessive. This geometry has never been used on a spacecraft.

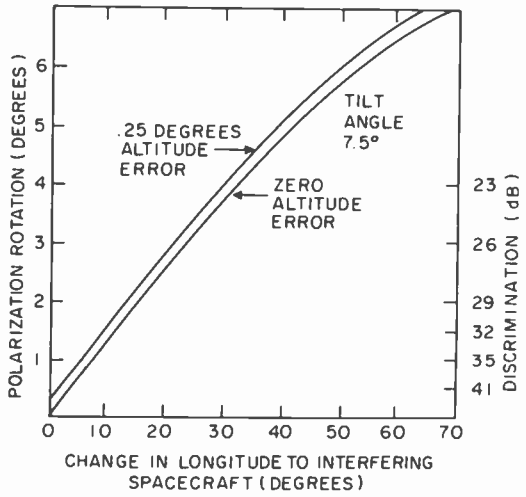


Fig. 8—Discrimination level due to antenna tilt angle (not including earth antenna discrimination).

Cassegrain Horn Feed

This geometry eliminates blockage effects by offsetting the reflector; subreflector can then be large enough to have satisfactory beam-forming properties. However, it has two drawbacks. First, the increased area of precision reflecting surface results in a corresponding increase in weight. Second, the center of gravity of the antenna is pushed farther away from the spacecraft upper deck, thus increasing the mechanical design and fabrication problems of the antenna and magnifying the dynamic stability problem of the spacecraft.

Conical Horn Reflector

Because of the large size of the conical horn, this type of antenna is considered to be too large and heavy for application on a satellite. For earth-coverage beams at microwave frequencies, a simple horn without

the reflector is frequently used, but for beam sizes of the order of 1°, the horn would become excessive in size and weight.

Prime Focus Reflector

This simple geometry generally has minimum weight and size and is simple from both an electrical and mechanical design standpoint. It suffers from a loss of gain due to blockage by the horn and an increase in cross polarization due to scattering from the horn support structure.

Off-Set Fed Reflector

This configuration is generally used for spacecraft antennas. It removes all, or nearly all, blockage from the antenna aperture, thus giving maximum antenna gain. In addition it reduces the scattering from the horn and support structure into the cross-polarized component.

Modifications to the basic configuration include circular aperture feed horns, multi-horn feeds, gridded reflectors, and gridded screens.

Circular feed horns, if fed with the appropriate mode, are used to reduce the cross-polarization of the antenna.[8] At the appropriate F/d ratio, the cross-polarized component introduced by the horn can be used to cancel the component introduced by the reflector thus effecting an improvement in cross-polarization isolation. Multiple feed horns are used to provide a carefully contoured far-field antenna beam to match the desired ground coverage area.[25,26,28,31–34] This is accomplished by using a large aperture giving a pencil beam for each of the feed horns. By driving an appropriately selected subset of the feed horns with equal amplitude and phase, the desired antenna beam shape can be built up. Other shapes of antenna beams can be obtained by selecting another subset of the feed horns. This antenna design optimizes the antenna gain over the coverage area for a number of reasons. First, the gain of the antenna is fairly uniform across the beam. The energy maximum normally appearing at the beam center has been spread uniformly over the beam, increasing the gain above that which would be obtained from a diffraction limited beam. Second, the antenna beam has a sharp fall off outside the coverage area. Third, by contouring the antenna beam to match the desired coverage area, a large amount of energy that would normally be directed toward unpopulated areas can be directed into the main beam to increase the gain in the desired direction. The magnitude of the gain increase will depend upon the complexity of the desired coverage area and the number of spot beam used to make up the antenna beam.

In another application of multiple feed horns, the horns are not paralleled but are used individually to carry different signals to adjacent

areas on the ground by connecting the signals to adjacent feed horns on the satellite. Because adjacent spot beams overlap on the ground, it is necessary to provide isolation by using different frequencies or crossed-polarization or both. As an example, consider an antenna with side lobes such that a given frequency can be used on beams 1 and 4 with an acceptably low level of interference. If the band is divided into eight 60 MHz channels, then the channels can be assigned to the spot beams in the grid as indicated in Fig. 9. Two 60-MHz channels can be assigned to each beam and each beam can carry different information. In this way the full antenna gain of the spot beam is maintained for each channel, but many power amplifiers are required to feed all the spot beams. To cover a country such as Canada from the geostationary orbit, only four beams 2° in diameter are required.

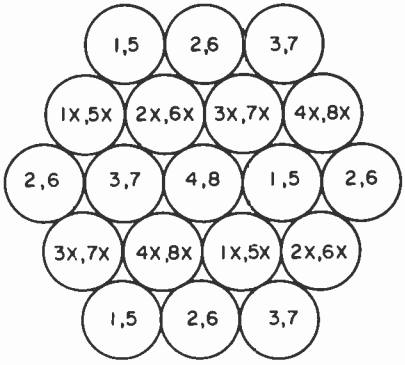


Fig. 9—A possible channel assignment scheme for a multiple-beam antenna system.

3.5 Apparent Polarization Rotation of Satellite Signals

A problem associated with rain depolarization occurs when a single satellite is used to illuminate large areas of the ground. If a single large antenna beam is used to illuminate the whole area, the field vectors are everywhere essentially parallel in the beam. However, when viewed from the ground the apparent polarization orientation is not the same everywhere. If the satellite polarization is such that it appears vertical at one location, then it will not appear vertical at other locations. This is because of a change in the local vertical rather than a rotation of the polarization vector. This effect has been calculated using the formula appearing after Eq. [38], namely

$$\tan\xi = \sin(\text{Long})/\tan(\text{Lat}) \qquad [49]$$

where Lat is the latitude of the earth station, Long is the longitude of the earth station relative to the subsatellite longitude (earth position directly below satellite), and ξ is the apparent rotation of the polarization vector with reference to the local vertical compared to the orientation on the subsatellite longitude. This formula does not include the contribution due to the field vectors not being precisely parallel everywhere.

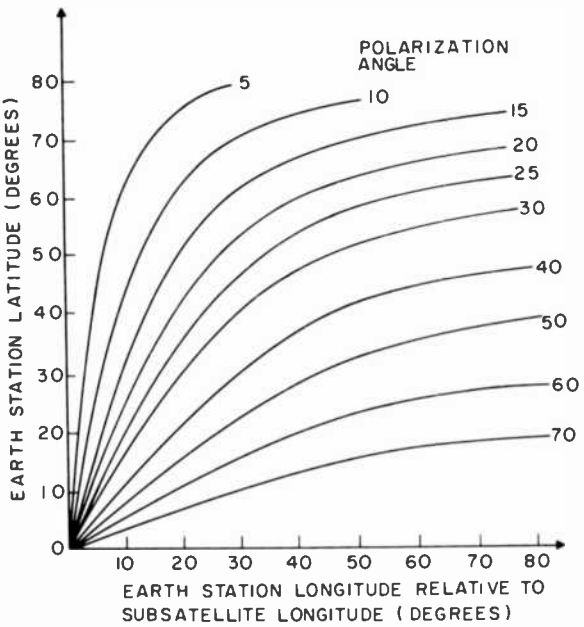


Fig. 10—Change in apparent polarization angle relative to an earth station located on the subsatellite longitude.

This contribution is very small and can be neglected for beam widths of a few degrees, as considered here.

Results of the calculation are shown in Fig. 10. It is seen that, for a single beam covering Canada and polarized vertically on the subsatellite longitude, the polarization vector near the edges of the country are rotated from the local vertical by as much as 25°–30°. Thus, rain depolarization, while it is a minimum for vertical polarization, is considerable worse at the edges of the country where the polarization is no longer vertical.

In the multibeam situation there are two alternatives:

(1) Each individual beam can have its polarization vector rotated so that it appears parallel to the local vertical at the beam center.

(2) The polarization vectors of every beam can be made parallel (or perpendicular) at the satellite, irrespective of the apparent orientation at the earth stations.

In the first case, the rain depolarization is minimized everywhere and the resulting cross-coupling between polarizations is accepted. In the second case, the cross-coupling betweem beams is minimized and then

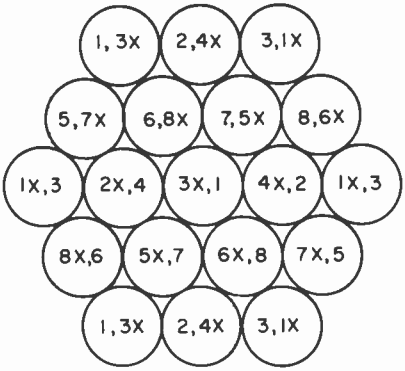


Fig. 11—A channel assignment scheme suitable for a system maintaining the polarization at each beam center to be locally vertical and horizontal.

the rain depolarization effect is accepted. The channel assignments shown in Fig. 9 have been made under the assumption of case 2. A suitable channel assignment scheme for case 1 is shown in Fig. 11. A detailed analysis would be required in any particular case to minimize cross-coupling between beams including all factors, such as that deliberate polarization rotation (as in case 1) gives a continuous coupling factor, while rain depolarization only occurs for a fraction of the total time.

Acknowledgment

This work was sponsored by the Department of Communications, Ottawa, Canada, under DSS Contract No. 12ST.36100-5-2101, DSS Serial No. OST5-0105. The authors wish to acknowledge helpful discussions and consultations with F. J. F. Osborne, P. Foldes, A. R. Raab, J. R. G. Cox, C. M. Kudsia, and H. Staras.

References:

[1] "Values of Rainfall Depolarization to be Adopted for System Planning," EBU Proposal Draft Report (K3) 162-E, Dec. 1975.

[2] T. Oguchi and Y. Hosoya, "Scattering Properties of Oblate Raindrops and Cross Polarization of Radio Waves Due to Rain (Part II): Calculations at Microwave and Millimeter Wave Regions," *J. Radio Res. Labs.* (Japan), **21**, No. 105, p. 191 (1974).

[3] T. S. Chu, "Rain-Induced Cross-Polarization at Centimeter and Millimeter Wavelengths," *Bell System Tech. J.*, **53**, No. 8, p. 1557, Oct. (1974).

[4] M. K. Lee and H. Staras, Private Communication.

[5] P. A. Watson and M. Arbabi, "Semiempirical Law Relating Cross-Polarization Discrimination to Fade Depth for Rainfall," *Electronics Lett.*, **11**, No. 2, p. 42, Jan. 23 (1975).

[6] G. C. McCormick and A. Hendry, "Method for Measuring the Anisotropy of Precipitation Media," *Electronics lett.*, **9**, No. 10, p. 216, May 17 (1973).

[7] "Polarization Discrimination by Means of Orthogonal Circular and Linear Polarization," CCIR Report 555 (1974).

[8] F. T. Tseng, "Effective Spectrum Utilization in Satellite Communications by Improving Antenna Radiation Performance," IEEE Publ. 75-CH-0971-2CSCB, Int. Conf. on Communications, p. 10–16 (1975).

[9] "Use of Antenna Reference Patterns in Deducing Levels of Interference," EBU Proposal Report (K3) 135-rev. E, Dec. (1975).

[10] T. Oguchi, "Rain Depolarization Studies at Centimeter and Millimeter Wavelengths: Theory and Measurement," presented at XVIIIth Gen. Assembly of URSI, Com. II, Lima, Peru (1975).

[11] M. Shimba, K. Morita, and Y. Hosoya, "Degradation of Cross Polarization Discrimination due to Rain," *Joint Conv. Record—Four Institutes of Electrical Engineers*, Japan, (in Japanese) p. 5.61 (1974).

[12] P. A. Watson and M. Arbabi, "Rainfall Cross Polarization of Linearly and Circularly Polarized Waves at Microwave Frequencies," *Electronics Lett.*, **8**, No. 11, p. 283, June 1 (1972).

[13] R. A. Semplak, "The Effect of Rain on Circular Polarization at 18 GHz," *Bell System Tech. J.*, **52**, No. 6, p. 1029, July–Aug. (1973).

[14] D. C. Hogg and T. S. Chu, "The Role of Rain in Satellite Communications," *Proc. IEEE*, **63**, No. 9, p. 1308, Sept. (1975).

[15] C. W. Bostian, W. L. Stutzman, P. H. Wiley, and R. E. Marshall, "Millimeter Wave Rain Depolarization: Some Recent 17.65 GHz Measurements," *IEEE/AP Int. Symp. Digest*, p. 289 (1973).

[16] R. A. Semplak, "Simultaneous Measurements of Depolarization by Rain Using Linear and Circular Polarizations at 18 GHz," *Bell System Tech. J.*, **53**, No. 2, p. 400, Feb. (1974).

[17] P. A. Watson and C. J. Soutter, "Dual Polarization Frequency Re-Use in Satellite Communications Systems at 11 GHz," IEE Publ. No. 126 on Satellite Communication Systems Technology, p. 294 (1975).

[18] S. Soma, I. Sato, S. Tamagawa, and M. Jida, "A High Performance Mode Transducer/Polarizer Assembly in the 4 and 6 GHz Bands," *IEEE/AP Int. Symp. Digest*, p. 387 (1974).

[19] S. I. Ghobrial, "Cross-Polarization Effects of Paraboloidal Reflector Antennas Surface Errors," IEE Publ. No. 126 on Satellite Communication Systems Technology, p. 246 (1975).

[20] I. D. Clough and J. Susplugas, "A Review of Some Problems of In-Orbit Testing of a Communication Satellite," IEE Pub. No. 126 on Satellite Communication Systems Technology, p. 63 (1975).

[21] "Reference Diagrams for Antennas to be Used in Systems for Broadcasting from Satellites," EBU Proposal Report (K3) 139 rev. E, Jan. (1976).

[22] "Satellite Antenna Patterns in the Fixed Satellite Service," CCIR Report 558 (1974).

[23] "Broadcasting-Satellite Service—Sound and Television," CCIR Report 215-3 (1974).

[24] A. R. Raab, "Cross-Polarization Performance of the RCA Satcom Frequency Re-Use Antenna," To be presented at IEEE/AP Symp., Oct. (1976).

[25] L. J. Ricardi et al, "Some Characteristics of a Communication Satellite Multiple Beam Antenna," Lincoln Lab. Technical Note 1975-3, Jan. 28 (1975).

[26] A. R. Dion, "Optimization of a Communication Satellite Multiple-Beam Antenna," Lincoln Lab. Technical Note 1975-39, May 27 (1975).

[27] L. K. DeSize, "Measured Characteristics of Variable Power Dividers," Lincoln Lab. Technical Note 1975-38, May 27 (1975).

[28] B. M. Elson, "Multi-Beam 'Lens' Studied for Satcoms," *Aviation Week and Space Technology*, **103**, p. 75, Oct. 6 (1975).

[29] P. A. Watson and S. I. Ghobrial, "Cross-Polarization in Cassegrain and Front-Fed Antennas," *Electronics Lett.*, **9**, p. 297, July 12 (1973).

[30] P. J. Wood, "Depolarization with Cassegrain and Front-Fed Reflectors," *Electronics Lett.*, **9**, p. 181, May 3 (1973).

[31] W. G. Scott, H. S. Luh and E. W. Matthews, "Design Tradeoffs for Multiple Beam Antennas in Communications Satellites," International Conf. on Communications, Philadelphia, Penn., June (1976).

[32] J. W. Duncan, S. J. Hamada and P. G. Ingerson, "Dual Polarization Multiple Beam Antenna for Frequency ReUse Satellites," AIAA Satellite Communication Systems Conf. Paper 76-248, Montreal, April (1976).

[33] E. A. Ohm and M. J. Gans, "Numerical Analysis of Multiple-Beam Offset Cassegrainian Antennas," AIAA Satellite Communication Systems Conf. Paper 76-301, Montreal, April (1976).

[34] R. Kreutel, "A Multiple-Beam Torus Reflector Antenna for 20/30 GHz Satellite Communications Systems," AIAA Satellite Communication Systems Conf. Paper 76-302, Montreal, April (1976).

Subjective Effects of Bit Errors in a PCM Color Television System

Richard A. Ulene

RCA Laboratories, Princeton, N.J. 08540

Abstract—Subjective testing was made of human observers' response to PCM television pictures degraded with varying amounts of binary symmetric channel noise. Several approaches are discussed for evaluating picture quality measurements, the final choice being the Relative Preference method based on Thurstone's Law of Comparative Judgment. A suitable model for human response to visual stimuli is developed, as well as the analytical methods for analyzing the data from the subjective experiments. The experiments employ the method of paired comparisons to generate the appropriate data for the model. A description of the experiments is given and of the equipment used to generate the impaired television images. Of particular importance is the generation device used to inject known amounts of errors onto the PCM communication lines. This device generates eight nearly uncorrelated error streams by sampling the outputs of three random noise generators. A computer program was used to estimate the parameters of interest from the observed preferences, calculate confidence intervals on the estimates, and also to perform a goodness-of-fit test on the estimated parameters using the hypothesized model. The parameter estimation method is the maximum-likelihood procedure. The results indicate that PCM television picture quality is more dependent on the brightness level and the amount of detail in the picture than conventional analog television picture quality. The results of the TASO tests of 1958 are cited for purposes of comparison. Limitations of the procedures are discussed and suggestions for future work are made.

1. Introduction

Transmission of television pictures by pulse-code modulation (PCM) offers several advantages[1-3] over transmission by analog methods, including higher signal-to-noise ratios and time-division multiplexing capabilities. Because of its digital nature, a high-quality signal can be maintained even though the signal may pass through a long chain of

repeater amplifiers. Due to its well-defined timing characteristics, PCM is also ideally suited for multiplexing signals when high degrees of quality and reliability are required, such as in network broadcast links.

Briefly, PCM is a technique for encoding an analog signal in the form of eight-bit "words" where each word corresponds, within some small increment, to the value of the original analog signal voltage at that particular sampling time. The size of the increment is determined both by the range of voltages being encoded and the number of discrete levels that can be coded for, the latter being given by 2^n where n is the number of bits in each word. As long as the sampling interval is short enough* the received signal can be decoded to yield an almost exact replica of the original signal.

The most important artifact of the process described above is an error signal called "quantization noise." Quantization noise results because the process of representing a signal composed of an infinite number of levels (that is, the original continuous signal voltage) by a finite number of discrete levels introduces an error in the reproduced message. Quantization noise, while very important in actual system design, will not be considered extensively in this paper, since the system under study is relatively free of this type of noise (as shown in a later section).

In the environments that the system under study might be employed, there is yet another type of interference that is of major importance, since it bears directly on system complexity and cost. This interference is produced by external noise in the communications channel and is referred to as binary symmetric channel noise or "bit error" noise. The noise takes the form of occasional errors in the code words of the digitized message. The errors arise when the received bits are misread—a "0" is interpreted as a "1," and vice versa. The presence of high levels of noise in the channel can confuse the decoding device by changing the well-defined digital signals from the transmitter into ill-defined signals. There are several ways to quantify a certain level of bit error noise. Among the most useful parameters is the average error rate, defined as the average number of errors per bit, or the probability that a particular bit will be in error when received. This number is usually most easily determined by finding the average over time of the time intervals between successive errors and dividing its reciprocal by the sampling frequency.

---

* In any sampling process where one wishes to recover a signal that is band-limited to a frequency f_0, it is necessary only to retain the instantaneous amplitude of the signal at intervals that are spaced no farther than $1/(2f_0)$ apart in time.[2]

† If the errors are generated locally, and they are recoverable as a random pulse train, then there is a rather convenient method for monitoring the error rate. A frequency counter may be used where the input signal is the error pulses and the counter time base is the sampling pulse train. This effectively causes the counter to display the ratio between the time-averaged bit error frequency and the sampling frequency, giving the time-averaged bit error rate. This was the method used in the equipment employed in this work.

The remainder of this paper is concerned with how a typical television viewer perceives, subjectively, different levels of bit error noise and how the subject matter affects these perceptions. Part 2 of the paper is concerned with the development of a satisfactory model that adequately characterizes responses of human observers in a quantitative form. Part 3 describes an experiment based on the model proposed in Part 2, and Part 4 summarizes the results of the experiment and suggests possible interpretations of the outcome. Part 5 presents conclusions and discusses the limitations imposed by time and equipment as well as those due to the experimental method chosen.

2. Available Models

The problems associated with subjective evaluation of picture quality are in many ways identical to those encountered when speech quality is subjectively evaluated. For this reason, it would appear that many of the methods used for speech quality evaluation might also be applied to television picture quality.

Methods for measuring speech quality can be divided into several distinguishable groups,[5] the two major classifications being "analytic" and "utilitarian." Analytic methods attempt to analyze speech quality with the intention of determining its psychological components; they commonly use factor analyses of similarity and preference data, as well as other types of multivariate techniques. Utilitarian methods attempt to determine a measure of speech quality that is suitable for the evaluation of speech transmission systems; these methods normally lead to unidimensional solutions and include pair-comparison procedures.

An IEEE report[4] on recommended practice has found that the utilitarian methods appear to be best suited for engineering studies. It recommends three approaches for speech-quality measurements: (a) the isopreference method; (b) the relative preference method; and (c) the category-judgment method. In the isopreference method, a test signal is compared directly with a reference signal for different levels of degradation, and an isopreference level of the test signal is determined to be the impairment level (usually the signal-to-noise ratio) of that reference signal at which the preference votes of an observer group are equally divided. From many comparisons, psychometric curves that give the percentage of judgments preferring the reference signal over the test signal for different levels of impairment can be constructed, and these may be used to obtain sets of isopreference curves. References [6] and [7] make use of this method. A modified version of the isopreference method has been used in a few studies.[8–10] In this procedure the observers are asked to specifically adjust the impairment level of the test

signal (which in these studies is a television picture) until it cannot be distinguished from the reference picture in terms of its interfering effect. Isopreference curves may be obtained directly from these experiments.

In the relative preference method the quality of the test signal is placed (using preference testing) on a unidimensional continuum that has the properties of an equal-interval scale. This is sometimes called a standard unit variance scale of esthetic acceptability. The scale is defined by several reference signals, which are generally chosen to spread approximately evenly across the scale. The reference and test signals are scaled by considering how often each is preferred to the others. This method is employed in Reference [11].

The category-judgment method requires observers to relate the test signal to some subjective standard of quality. A set of categories bearing descriptions such as "good" or "objectionable" are presented to the observer, and he is asked to place the test signal in one of these categories. The result of this technique is that the test signals are defined on this standard scale. This method is currently the most widely used,[1,12–17] and it can be used in conjunction with other methods to help anchor test results to more familiar measures of quality.

It seems reasonable to extend these methods to television, as many of these studies have done. Each of the methods has advantages and disadvantages. Looking first at the isopreference method, it would appear to be the simplest in terms of experimental method and data reduction. One need simply make a sufficient number of comparisons to allow the construction of psychometric curves. The modified version of this method only requires that the observer try to match two signals. Unfortunately, both versions have drawbacks in that they *must* be anchored to reference signals to be very meaningful, and they give results in terms of isopreference curves from which it is not always easy to obtain useful conclusions. In addition, the modified version also suffers because each observer must be tested separately, and this is time consuming; furthermore, this version requires that the test signals and reference signals be very similar so that the observer will have little trouble in deciding when two signals are equally interfering.

The category-judgment method gives the most aesthetically pleasing output. The major problems with this method are that it requires the largest number of observers (particularly in regard to memory) and it also may be influenced by the set of signals that is judged.[4]

The relative preference method is experimentally similar to the unmodified version of the isopreference method, i.e., it does not demand too much of the observers. While the output is more meaningful in its own right, the data reduction is somewhat more complicated, and there

is some sensitivity to the amounts of degradation in the reference signals (if the results are to be anchored with these reference signals).[4]

In light of the above considerations the author has chosen the relative preference method whose results, if the reader wishes, may be anchored by a category-judgment experiment, which was performed by Panel 6 of the Television Allocations Study Organization (TASO) in 1958.[12-14]

2.1 The Model

The relative preference method is derived from Louis L. Thurstone's Law of Comparative Judgment.[18] The law assumes that inter-individual differences in preference response distribute essentially normally on a unidimensional equal-interval scale.[11] Briefly, the law specifies the following model.* For any stimulus X_j in a judgmental experiment, there is an associated "discriminal process" that can be represented as a random variable v_j, composed of a fixed component μ_j common to all observers and a random component e_j, i.e.,

$$v_j = \mu_j + e_j. \qquad [1]$$

The process above is hypothetical—we cannot measure it directly. We can, however, measure the difference between discriminal processes by comparison of stimuli. For any two stimuli X_a and X_b, the associated discriminal processes v_a and v_b, whose probability densities are portrayed in Fig. 1a, are

$$v_a = \mu_a + e_a, \qquad [2a]$$

and

$$v_b = \mu_b + e_b. \qquad [2b]$$

These distributions have means μ_a and μ_b, variances $\sigma_a{}^2$ and $\sigma_b{}^2$, and correlation ρ_{ab}. For the moment the distributions are assumed to be normal. This is not a necessary restriction on the procedure, and another possible distribution will be discussed shortly. The response of observer i is assumed to be the following:

$$(A)\ X_a > X_b \text{ when } (v_{ai} - v_{bi}) > 0,$$

* Thurstone actually described several cases of his model. The one developed here is basically Case II, which characterizes discriminal process variation over a population rather than over repeated trials with one individual, which is Case I. Cases III–V make assumptions about the similarity in type of stimuli which need to be made later in this work.[28]

or

$$(B)\ X_b > X_a \text{ when } (v_{ai} - v_{bi}) < 0,$$

$$(X_a = X_b \text{ is not permitted}).$$

The model can therefore be expressed by defining a "discriminal difference" of the form

$$v_{ab} = v_a - v_b = (\mu_a - \mu_b) + (e_a - e_b)$$
$$= \mu_{ab} + e_{ab}. \qquad [3]$$

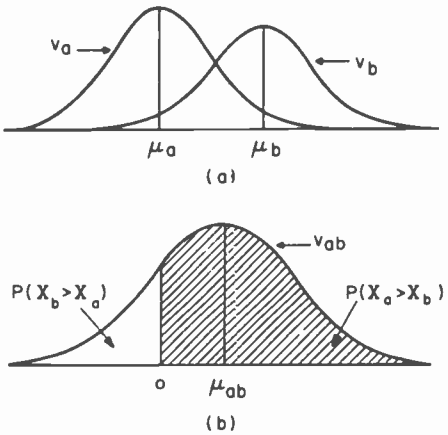


Fig. 1—(a) Marginal probability densities of v_a and v_b and (b) probability density of v_{ab}.

The probability density of v_{ab} is shown in Fig. 1b. It has a mean $\mu_{ab} = \mu_a - \mu_b$ and a variance $\sigma_{ab}{}^2 = \sigma_a{}^2 + \sigma_b{}^2 - 2\rho_{ab}\sigma_a\sigma_b$.* When v_a and v_b are assumed to be normally distributed, then v_{ab} is also, and the probability of X_a being judged greater than X_b is

$$P_{ab} = \frac{1}{\sqrt{2\pi}\,\sigma_{ab}} \int_0^\infty \exp\left[-\frac{1}{2}\left(\frac{y - \mu_{ab}}{\sigma_{ab}}\right)^2\right] dy. \qquad [4]$$

For notational convenience, it helps to introduce the change of variable

$$z = \frac{y - \mu_{ab}}{\sigma_{ab}};$$

* The standard deviation σ_{ab} corresponds to the error component of random variable v_{ab}. In contrast to μ_{ab}, which tells how similar two stimuli are in terms of quality, σ_{ab} gives an indication of how similar two stimuli are in terms of the *type* of impairment; i.e., it implies the level of difficulty in choosing which stimulus is of higher quality due to the dissimilar characters of the stimuli.

so now we have

$$P_{ab} = \frac{1}{\sqrt{2\pi}} \int_{-\mu_{ab}/\sigma_{ab}}^{\infty} \exp\left(-\frac{1}{2} z^2\right) dz. \qquad [5]$$

Eq. [5] is the integral of the unit normal density function with the argument μ_{ab}/σ_{ab}. For the sake of generality, we can express the above equation as

$$P_{ab} = \Phi(\mu_{ab}/\sigma_{ab}). \qquad [6]$$

Eq. (6) is referred to as the "response function" and P_{ab} is frequently called the "response probability." The response function transforms the parameters of the model (μ and σ) to the expected proportion of times that stimulus X_a is chosen over stimulus X_b.

Up to this point, we have concentrated exclusively on the judgmental model without making any attempt to relate the psychological variate of the model μ_{ab} to the characteristic(s) of the stimuli that interest us (such as quality). We must now hypothesize a "psychophysical model" that relates the two. In Thurstone's original paper,[18] he introduces parameters α_a and α_b, called "affective values," which represent the characteristic of the objects participating in a single pairing. He then proposes the relationship

$$\mu_{ab} = f(\alpha_a) - f(\alpha_b), \qquad [7]$$

where the terms on the right are single-valued functions of the affective values. This relation is used to find the affective values given the means. Eqs. [6] and [7] together constitute a statement of the law of comparative judgment.

Once armed with the above relationships, we are finally prepared to tackle the fundamental problem of estimating the affective values and the discriminal dispersions, given the set of observed proportions (preferences) $\hat{P}_{ab}$, which are obtained through the experimental method of paired comparisons. Before proceeding, however, it is useful to reflect on several points. First, since the ratio μ_{ab}/σ_{ab} is invariant with changes in the unit of measurement for the discriminal processes, the model gives us the freedom of choosing that unit arbitrarily. One of the most convenient ways of doing this is to assign $\sigma_{ab} = 1$ for some particular value of a and of b. In addition to this degree of freedom, the model allows us to choose the origin for the affective values; this is an even simpler matter, and it is handled by arbitrarily assigning some value to one of the affective values.

Another point to remember is that although the integral of the unit normal density function is usually used as the response function,[18-21] many other functions may be implemented, provided the one chosen is

differentiable as many times as is necessary in a particular application.[21,22] The choice of response function depends on several factors. Among the most important are the estimation criterion used (i.e., the function that is being minimized or maximized), the goodness-of-fit to the data (i.e., the closeness with which the estimated parameters can predict the originally observed proportions), and physical interpretability (i.e., the likelihood of the response function corresponding to some actual physical process). In the present case, two response functions were investigated—the standard unit normal density function and the logistic function.

The logistic response function takes the form

$$P_{ab} = \frac{1}{4} \int_{-\mu_{ab}/\sigma_{ab}}^{\infty} \operatorname{sech}^2 \frac{z}{2}\, dz.$$

It is similar to the standard normal response function, but it is derived from the following model (Bradley–Terry–Luce Model).[21] Assume that the discriminal processes have an actual positive physical intensity, and that the observer responds as follows:

$$\text{(A) } X_a > X_b \text{ when } \frac{v_a}{v_b} > 1,$$

$$\text{(B) } X_b > X_a \text{ when } \frac{v_a}{v_b} < 1.$$

If v_a and v_b are random variables with medians m_a and m_b (which correspond to affective values), then the probability that X_a will be judged greater than X_b approaches zero as m_b increases without limit for fixed m_a. If $m_a = m_b$, the probability is one-half. When m_a increases without limit for fixed m_b, the probability approaches one. A simple function which has the above properties is

$$P(X_a > X_b) = \frac{m_a/m_b}{1 + (m_a/m_b)}.$$

Because it is not convenient to express the ratio m_a/m_b by the psychophysical model of Eq. [7] directly (the ratio may not be assigned a positive value by this equation), an exponential reparameterization of the simple function above is performed, yielding

$$P(X_a > X_b) = \frac{\exp(\mu_{ab})}{1 + \exp(\mu_{ab})} = P_{ab}$$

When σ_{ab} is assigned the value "one," then the logistic response function reduces to the simple ratio above. Logistic deviates can be obtained by an inverse transformation, and these may be used in much the same

way as normal deviates. As it happens, the logistic and normal response functions are very similar and yield similar results in parameter estimation schemes.

As a final point, the reader may have noticed a possible flaw in the preceding development. Isn't the above model vulnerable to effects caused by the observers' criteria for choosing one stimulus over another? As it turns out (see Appendix 1), the method of paired comparisons (in which all reasonable comparisons between stimuli are investigated) is a forced-choice procedure, which means that the results obtained from it will be independent of the observers' criteria.[25]

2.2 The Estimation Procedure

The problem of parameter estimation can be attacked with any one of a variety of methods; among the most popular are the least-squares solutions and the maximum-likelihood solution.

The classic principle of least squares[21] is that estimates of the parameters of a model should be chosen so as to minimize an appropriate linear function of squared residual elements. Due to statistical considerations, it is usually necessary that this function be a formal goodness-of-fit criterion whose sampling distribution is known; the most popular least-squares solution is called the "minimum normit χ^2" (Urban's solution). It is not necessary to give a formal derivation here (for the interested reader, a detailed formulation may be found in Bock and Jones[21]). In brief, since we assume the inverse transform of the observed proportions

$$\hat{y}_{ab} = \Phi^{-1}(\hat{P}_{ab})$$

to be a random variable of the form

$$\hat{y}_{ab} = f(\alpha_a) - f(\alpha_b) + \epsilon_{ab}, \qquad [8]$$

where ϵ_{ab} is the random sampling error, then

$$Q = \sum_{a<b} N_{ab} W_{ab} [\hat{y}_{ab} - f(\alpha_a) + f(\alpha_b)]^2 \qquad [9]$$

is distributed as χ^2 in the limit. Here, N_{ab} is the number of observers who make the comparison, and W_{ab} is the weighting value for that comparison. The fundamental aim of the minimum normit χ^2 solution is to estimate the affective values of the stimuli by minimizing the right-hand expression in Eq. [9] with respect to the variation of the parameter values, i.e., by adjusting the affective values until the minimum squared error is obtained. This least-squares estimation criterion was not used in this work primarily because it does not easily lend itself to

the determination of confidence intervals on the affective value estimates.

The maximum-likelihood solution, unlike the least-squares solution, allows one to obtain confidence intervals with little extra computation; for this reason, it was chosen over the minimum normit χ^2 as the estimation criterion for the data analysis in this study. The principle of maximum likelihood can be explained as follows.[21,26] The likelihood function is defined as the joint frequency distribution function of the set of randomly sampled observations $\hat{P}_{ab}$ obtained from the experiment; i.e., it gives the probability of obtaining any particular matrix of observed proportions. The data from a paired-comparison experiment may be summarized as the number of times that X_a is judged greater than X_b for the ordered pair (X_a,X_b). If N_{ab} observers make this comparison, then this number is

$$r_{ab} = \sum_{i=1}^{N_{ab}} s_{abi}, \qquad [10]$$

where the response score s_{abi} is the value "one" if observer i judged stimulus X_a greater than stimulus X_b and "zero" if he did not. Note that $\hat{P}_{ab} = r_{ab}/N_{ab}$. If we now assume* that the judgments made by the observers are independent (within, and between, individuals) and that the spatial and temporal order of the stimuli have no effect on the judgments, then we may determine from the binomial probability mass function the a priori probability that the sum of the response scores takes the value r_{ab} for any one pair of stimuli:

$$\text{Prob}(r_{ab}) = \frac{N_{ab}!}{r_{ab}!(N_{ab} - r_{ab})!} P_{ab}{}^{r_{ab}}(1 - P_{ab})^{(N_{ab}-r_{ab})}. \qquad [11]$$

Remember that P_{ab} is the actual, not observed, response probability. Still assuming independence, we can now determine the joint frequency distribution function of the set (matrix) of experimental observations (the likelihood function) to be simply the product of the a priori probabilities of the ordered pairs;

$$\text{Prob(all } r\text{'s)} = L$$

$$= \prod_{a<b} \frac{N_{ab}!}{r_{ab}!(N_{ab} - r_{ab})!} P_{ab}{}^{r_{ab}}(1 - P_{ab})^{(N_{ab}-r_{ab})}. \qquad [12]$$

The likelihood function L is regarded as a function of the P_{ab}'s, with the r's given by the experiment. The maximum-likelihood solution is

* The assumptions in this procedure are reasonable so long as precautions are taken to ensure that the "identity" of each stimulus is withheld from the observers and that the order of presentation within, and between, pairs of stimuli is counterbalanced. Also implicitly assumed in the maximum likelihood procedure is that the psychophysical model chosen is correct; this is a standard assumption in most methods of parameter estimation.

obtained by maximizing Eq. [12] with respect to P_{ab}; but since the parameters of the model (see Eqs. [6] and [7]) determine the response probabilities, this is actually equivalent to finding the maximum-likelihood estimators for the affective values and the discriminal dispersions.

As an example of the way a solution is obtained according to the above procedure, we shall look at a hypothetical set of data given in the matrix $\tilde{P}$ of Fig. 2(a). For simplicity we shall assume Thurstone's Case V assumptions hold (all discriminal dispersions are equal). Given each $\hat{P}_{ab}$, we can write the matrix of response scores r_{ab}, which is shown in Fig. 2(b).

$\tilde{P}$

$\hat{P}_{ab}$		Stimulus b		
		1	2	3
Stimulus a	1	0.5	0.7	0.8
	2	0.3	0.5	0.6
	3	0.2	0.4	0.5

(a)

$\{N_{ab} = 10\}$

r_{ab}		Stimulus b		
		1	2	3
Stimulus a	1	5	7	8
	2	3	5	6
	3	2	4	5

(b)

Fig. 2—Hypothetical data: (a) matrix of observed proportions and (b) matrix of response scores.

Although we will not perform a maximization of the likelihood function here, we can choose some trial response probabilities and observe how the likelihood function behaves. If, for example, we let $P_{12} = 0.6$ and $P_{23} = 0.55$, then, if N_{12} and N_{23} both equal ten, the likelihood function has the value 0.016. If, on the other hand, we let $P_{12} = 0.7$ and $P_{23} = 0.55$, then L has the value 0.019. Now, for ideal data under Case V assumptions (into which category this falls), the best estimate of P_{ab} is clearly $\hat{P}_{ab}$, so if we let $P_{12} = 0.7$ and $P_{23} = 0.8$, then L has its maximum value of 0.080, and we have found the maximum-likelihood estimates of P_{12} and P_{23}. From Eq. [6] we can determine μ_{12} and μ_{23}; if we assume a normal response function, then $\mu_{12} = 0.52$ and $\mu_{23} = 0.84$ (the unit normal deviates of P_{12} and P_{23}), also assuming that $\sigma_{12} = \sigma_{23} = 1$. To get the affective values, we assume an origin such as $\alpha_1 = 10$ and we also assume a unit by, say, letting $\alpha_1 - \alpha_2 = 2$; by scaling everything accordingly, we find that $\hat{\alpha}_1 = 10$, $\hat{\alpha}_2 = 8$ and $\hat{\alpha}_3 = 4.8$.

Arbuckle and Nugent[22] describe a method for maximizing the function L in the presence of nonideal data and "missing data" (where $N_{ab} = 0$).

Basically, it depends on numerical descent procedures, such as the method of steepest descent, the method of conjugate gradients, the Fletcher and Powell adaptation of the variable metric method, and the Newton–Raphson method. For these methods it is sufficient to find the function value and its first- and second-order derivatives.

In addition to the parameter estimates, the maximum-likelihood procedure also provides limiting variances of these estimates (from which confidence intervals may be obtained). The limiting variances are given by the diagonal elements of the expected value of the negative inverse of the matrix of second derivatives, whose ab entry is given by

$$c_{ab} = E\left[-\left(\frac{\partial^2 \log L}{\partial \alpha_a \partial \alpha_b}\right)^{-1} \right]. \qquad [13]$$

Because maximum-likelihood estimators are asymptotically normally distributed, a large-sample 95% confidence region, say, could be determined by

$$\hat{\alpha}_i - 1.96\sqrt{c_{aa}} < \alpha_i < \hat{\alpha}_i + 1.96\sqrt{c_{aa}}, \qquad [14]$$

where c_{aa} is the ath diagonal element of the above matrix and $\hat{\alpha}_i$ is the maximum-likelihood estimate of the single parameter α_i.

The entire maximum-likelihood procedure described above has been implemented in a FORTRAN IV computer program developed by Arbuckle and Nugent[27] and revised by the author to run on a VM/370 (IBM) computer facility under the control of a time-sharing system. This program provides a measure of goodness-of-fit in the form of a likelihood ratio chi-square, which gives an idea of how closely the estimated parameters can predict the observed data when fed back into the model. This was employed to check the validity of the underlying model.*

Having developed an adequate model and the analytical procedures for using it, we are now prepared to turn to the experimental aspects of the study, where analog television pictures impaired with Gaussian white noise were compared with PCM television pictures impaired with known levels of bit error noise. The experiments, employing the method of paired comparisons, yield the response probabilities $\hat{P}_{ab}$, which may be used as input for the model described in this section. The results of the analysis will enable us to relate bit error noise levels to levels of band-limited Gaussian white noise, whose approximate subjective effects have been revealed by studies such as that of TASO.

* Although the maximum-likelihood method is capable of handling response probabilities of zero and one, these values were not retained in this analysis since it then becomes impossible to obtain a likelihood ratio chi-square to check for goodness-of-fit. In the present instance, these response probabilities were simply not included, i.e., they were considered to be missing data.

3. The Subjective Tests

The subjective testing was performed over a period of three days in December 1974. The observers were volunteers from RCA Laboratories, Princeton, N.J., and included two major groups: (1) clerks, secretaries, patent attorneys, maintenance personnel, and technical people from areas other than television and communication, who were classed as "non-experts," and (2) people active in image processing and the technical evaluation of television images, who were classed as "experts." No screening of the non-experts was carried out, since the aim was to make this group a somewhat representative random sample of the television-viewing public. Each group was tested and analyzed separately.

3.1 Experimental Procedure

Five people participated at one time for a twenty-minute test period, including a one-minute introduction. Each group of observers was briefly told that they would be shown pairs of television pictures (on two side-by-side receivers) and that they should indicate on the forms provided which of the two pictures they thought was "better," i.e., which one they would prefer to watch. The choice of "equally preferred" was not allowed. All observers were allowed to view each pair for ten seconds, and then they were provided with ten seconds to write down their choice of the most preferred picture.

Two different slides were used. One, referred to as "couch," was a picture of a woman standing behind a sofa in a darkened room having little contrast or detail. The other, referred to as "straw," was a much more detailed picture of a women's face in a background of straw. Each picture was shown twenty-seven times to every group of observers, each time with a different combination of analog and digital noise imposed on the images. Thus, each observer made fifty-four decisions in about twenty minutes. Counterbalancing was effected by randomly changing the temporal order in which the pairs of stimuli were shown and also on which receiver they appeared.

The room that was selected had been modified to serve as a general purpose subjective testing area. Fluorescent lighting was replaced with incandescent lamps which could be selected in various combinations to provide several different lighting conditions. Drapes were used generously to hide instrumentation and other visual distractions. The room was linked by video cable and telephone with the Laboratory's television studio to facilitate smooth operation of the experiment. The room illumination was measured as three foot-lamberts at the front-center observer's position (see the layout in Fig. 3), while the screen highlight

brightness was maintained at about fifty foot-lamberts.

The experiment was controlled from the far left side of the room by means of three control boxes (described in more detail in the next section). One of these was a switching box that allowed the instantaneous selection of three out of four possible combinations of stimulus types:

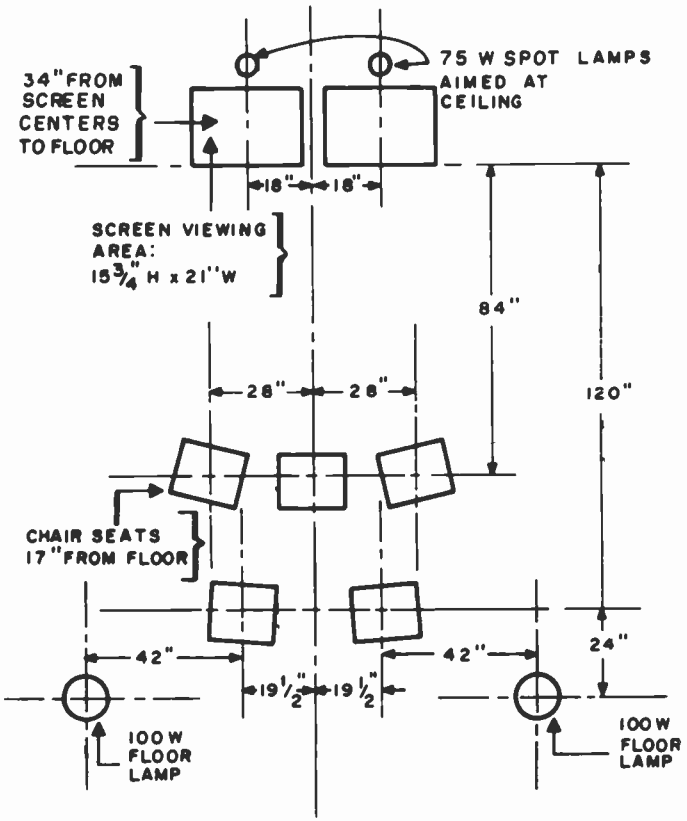


Fig. 3—Room layout for experiments.

(1) analog/analog, (2) digital/analog, (3) analog/digital, and (4) a black-out position. Digital/digital combinations were not shown since the method of paired comparisons allows for incomplete designs (within bounds), and the extra data is not needed in this instance.

The other controls were a threshold selector, which enables the instantaneous selection of all combinations of nine preselected bit error

rates on the eight parallel channels of the PCM system and the selection of a "no errors" position, and a two-channel five-position attenuator used for the selection of five preselected signal-to-noise ratios in the analog signal circuits.

During the experiments the comparisons were identified by number when first displayed on the screens. This number was also located on the questionnaire that each observer had been given. Records were maintained on the order of presentation of the pairs for purposes of tallying the results. At the conclusion of the subjective testing experiments, sixty-eight non-experts and twenty experts had participated. The results are described in Part 4; the equipment used in the tests is described below.

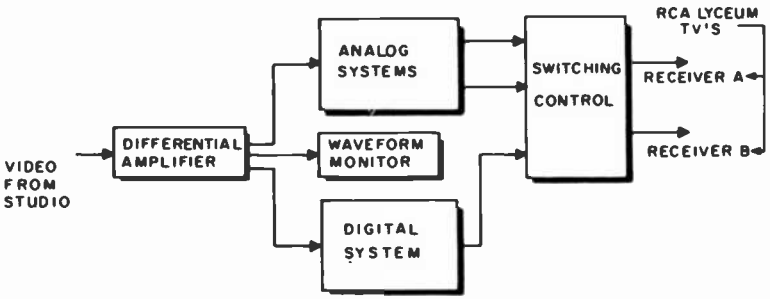


Fig. 4—Block diagram of equipment layout.

3.2 The Two Television Systems

The analog and digital television systems are similar in several respects. Both employ processing equipment that operates on a video composite signal (includes sync as well as video) containing frequencies in the range of zero to about 4.5 MHz. The composite signal was generated by an RCA slide projector/television camera system; this can produce motionless television images with a nominal picture-to-noise ratio (PNR)* of 45 dB.

Preceding the equipment is a specially designed differential amplifier (see Fig. 4) that effectively removes residual hum and noise that might

* The picture-to-noise ratio is equal to peak video divided by rms noise, where the peak video is defined as the voltage difference between the highest white level and the video blanking level (nominally 0.714 peak volts), and the rms noise voltage is the unweighted rms value of the Gaussian white noise signal. Although this is a popular way of measuring the signal-to-noise ratio of a composite video signal, it is not the only measure. To convert from the PNR to a TASO signal-to-noise ratio, for example, it has been shown by Gibson[29] and Beakley[35] that for the system used in this experimental setup, approximately 4.6 dB is added to the PNR. Conversion from a PNR to other measures of signal-to-noise ratio, such as CCIR weighted SNR, require other adjustment factors which have been determined.

be picked up in the video signal distribution system. The output of this amplifier is then driven to the two separate television systems.

The conventional analog television system (see Fig. 5) consists of a distribution amplifier (DA) with two isolated low-impedance outputs, each of which is connected to a 4.2 MHz active low-pass filter followed by another distribution amplifier. The noise contribution of this basic system is negligible. Two individually controllable noise signals are added by feeding the output of a 20 Hz – 20 MHz random noise generator through a dual-output DA, with each output connected through a specially built step-attenuator to an isolating DA. These DA outputs are summed, through a suitable resistor, with the corresponding video signal. All video signal levels are maintained at 1.0 volt peak-to-peak and are measured with a waveform monitor and a true-reading rms voltmeter.

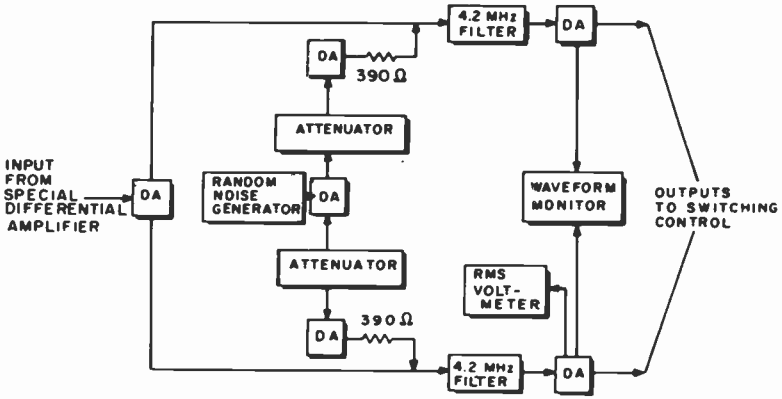


Fig. 5—Dual analog television system block diagram.

The basic PCM television system (shown in Fig. 6) is an eight-bit parallel channel pulse-code modulation communications link running at a sampling rate of 10.738635 MHz (three times the color subcarrier frequency). It consists of a 4.5-MHz active low-pass filter, followed by a DA that drives an eight-bit parallel analog-to-digital (A/D) converter to produce the (TTL compatible) PCM communication lines. The clock signal for the A/D converter is generated by a 3.579545-MHz crystal oscillator that can be optionally synchronized with the color burst in the video composite signal by a phase-locked loop circuit (the clock was kept synchronized with the color-burst signal in the subjective tests). This timing signal is finally tripled using a frequency-multiplier circuit, yielding the clock signal.

At the receiving end of the eight-channel PCM link is an eight-bit parallel digital-to-analog (D/A) converter whose analog output is fed to another 4.5-MHz active low-pass filter to eliminate spurious components from the recovered signal. The filtered signal is then passed through a multiple-output DA, which serves as the output driver for the digital system, both to a television monitor and to a waveform monitor.

Several methods of bit error generation were explored. Of particular interest were those that make use of pseudorandom sequences, as produced by a shift register with modulo two feedback.[30–32] The main advantages of this type of system are its ability to repeat exactly, its sta-

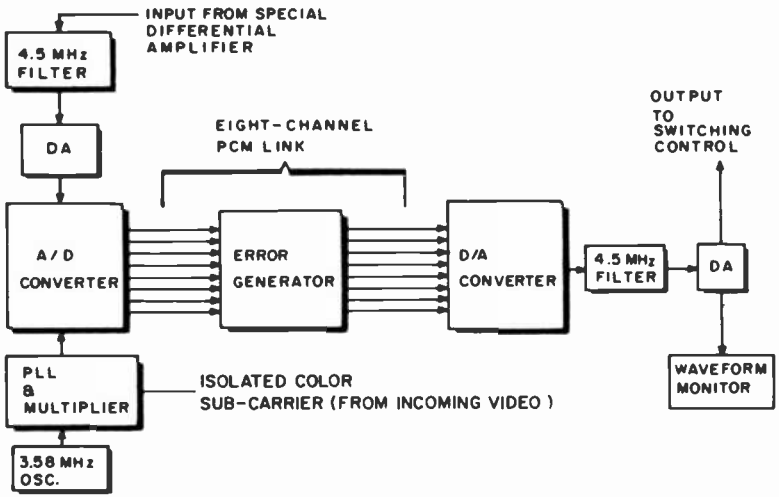


Fig. 6—PCM digital television system block diagram.

tistical distribution, and its spectral characteristics. Unfortunately, this approach is not very practical in the present instance because it requires that the shift register be completely cleared (i.e., shifted over as many times as the number of stages in the register) between each data-sampling interval to ensure that the resulting error streams are uncorrelated. At sampling rates of around 10 MHz, this implies shifting speeds of over 200 MHz if one register is to be used, or a prohibitively large number of shift registers if parallel generation is attempted.

A cheaper, yet quite satisfactory approach, implemented in the final system, employs emitter-coupled logic (ECL) gates and three random-noise generators of the type used in the analog television system. A block

diagram of this error-generation system is shown in Fig. 7.

First, all of the TTL signals from the A/D converter are translated to ECL levels; these include the eight output data lines and a delayed version of the clock signal. The error stream for each of the eight channels is derived by feeding the output of a random-noise generator with a fixed

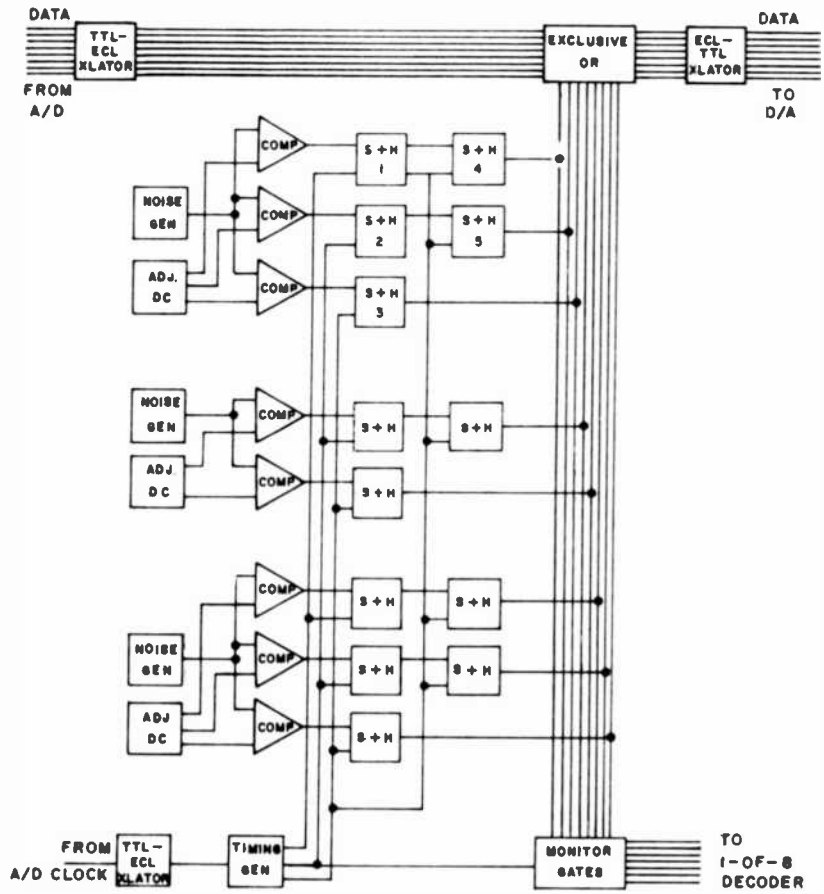


Fig. 7—Block diagram of error generator.

dc component to one side of a high-speed voltage comparator, while setting up a variable dc threshold voltage on the other comparator input. This comparator is made up of two cascaded ECL line receivers (differential amplifiers) and a Schmitt trigger circuit. The above process effectively converts the Gaussian noise to binary noise having the same

spectral characteristics. The percentage of time that the binary signal is in the "high" state is dependent on the amount of time that the difference between the random-noise signal and the adjustable threshold voltage is positive, which in turn is related to the difference between the dc component of the random noise (fixed) and the dc threshold voltage (variable).

The binary noise is next passed through several sample-and-hold circuits (D-type flip-flops) whose purpose is to generate three error streams from the original binary noise signal. This is accomplished by sampling three times during each sampling interval of the A/D converter (sample-and-hold circuits 1, 2, and 3 in Fig. 7). Since the three samples are spread equally over the 93-nanosecond data-sampling interval, three nearly uncorrelated error streams are produced (see Appendix 2 for a discussion of the correlation within, and between, the three error streams). Sample-and-hold circuits 4 and 5 are used to delay two of the error streams so that they line up exactly with the bit streams of the PCM data link. (The third error stream is already lined up.) The clocks for the sample-and-hold circuits are derived from the translated version of the A/D converter clock signal by using a pair of monostables to produce two differently delayed versions of the translated clock (timing generator). The above set of circuits is repeated twice so that a possible nine error streams can be generated—in the actual system, one error stream is not used so only eight are actually derived by the above procedure.

The outputs of the sample-and-hold circuits are finally combined with the ECL versions of the eight-bit video data streams by employing a bank of eight EXCLUSIVE-OR gates; this means that whenever an error stream is in the "high" state, the corresponding data bit will be in error (i.e., inverted). The error-filled data lines are translated back to TTL levels, and then they continue on to the D/A converter.

The error rates are monitored by a frequency counter (not shown), as explained previously, after being passed through a one-out-of-eight multiplexer (also not shown).

4. Experimental Results

The data obtained from the experiments takes the form of a proportion matrix $\tilde{P}$, which gives the observed proportion of times $\hat{P}_{ji}$ that stimulus j (column) was judged greater in picture quality than stimulus i (row). The four proportion matrices are shown in Figs. 8(a)–(d). Stimuli one through ten are, respectively, 45 dB, 10^{-7} error/bit, 10^{-6} error/bit, 10^{-5} error/bit, 38 dB, 35 dB, 10^{-4} error/bit, 31 dB, 28 dB, and 10^{-3} error/bit. (For the digital stimuli, the error rates were equal on all eight bits.) The blanks indicate missing data—either comparisons not made or propor-

$\tilde{P}_{nc}$

stimulus i \ stimulus j	1	2	3	4	5	6	7	8	9	10
1	.50000	.51471	.38235	.30882	.16176	.04412	.11765	---	---	---
2	.48529	.50000	---	---	.20588	.04412	---	---	---	---
3	.61765	---	.50000	---	.33824	---	---	---	---	---
4	.69118	---	---	.50000	.48529	.27941	---	.05882	---	---
5	.83824	.79412	.66176	.51471	.50000	.13235	.22059	---	---	---
6	.95588	.95588	---	.72059	.86765	.50000	.39706	.04412	.04412	.07353
7	.88235	---	---	---	.77941	.60294	.50000	.26471	.10294	---
8	---	---	---	.94118	---	.95588	.73529	.50000	---	.16176
9	---	---	---	---	---	.95588	.89706	---	.50000	.36765
10	---	---	---	---	---	.92647	---	.83824	.63235	.50000

Fig. 8(a)—Observed proportions $\hat{P}_{ji}$ for non-experts, couch picture ($N_{ji} = 68$).

$\tilde{P}_{ns}$

stimulus i \ stimulus j	1	2	3	4	5	6	7	8	9	10
1	.50000	.49254	.38806	.32836	.19403	.11940	.19403	---	---	---
2	.50746	.50000	---	---	.22388	.08955	---	---	---	---
3	.61194	---	.50000	---	.26866	.07463	---	---	---	---
4	.67164	---	---	.50000	.50746	.16418	---	---	---	---
5	.80597	.77612	.73134	.49254	.50000	.14925	.16418	---	---	---
6	.88060	.91045	.92537	.83582	.85075	.50000	.46269	---	---	.04478
7	.80597	---	---	---	.83582	.53731	.50000	.22388	.07463	---
8	---	---	---	---	---	---	.77612	.50000	.14925	.20896
9	---	---	---	---	---	---	.92537	.85075	.50000	.46269
10	---	---	---	---	---	.95522	---	.79104	.53731	.50000

Fig. 8(b)—Observed proportions $\hat{P}_{ji}$ for non-experts, straw picture ($N_{ji} = 67$).

$\tilde{P}_{ec}$

stimuli i \ stimulus j	1	2	3	4	5	6	7	8	9	10
1	.50000	.40000	.10000	---	---	---	---	---	---	---
2	.60000	.50000	---	---	---	---	---	---	---	---
3	.90000	---	.50000	---	.10000	---	---	---	---	---
4	---	---	---	.50000	.70000	.25000	---	---	.05000	---
5	---	---	.90000	.30000	.50000	---	---	---	---	---
6	---	---	---	.75000	---	.50000	.10000	---	---	---
7	---	---	---	---	---	.90000	.50000	.30000	.05000	---
8	---	---	---	---	---	---	.70000	.50000	---	.10000
9	---	---	---	.95000	---	---	.95000	---	.50000	.30000
10	---	---	---	---	---	---	---	.90000	.70000	.50000

Fig. 8(c)—Observed proportions $\hat{P}_{ji}$ for experts, couch picture ($N_{ji} = 20$).

$\tilde{P}_{es}$

stimulus i \ stimulus j	1	2	3	4	5	6	7	8	9	10
1	.50000	.52632	.21053	.05263	---	---	.05263	---	---	---
2	.47368	.50000	---	---	---	---	---	---	---	---
3	.78947	---	.50000	---	.05263	---	---	---	---	---
4	.94737	---	---	.50000	.57895	.26316	---	---	---	---
5	---	---	.94737	.42105	.50000	---	.05263	---	---	---
6	---	---	---	.73684	---	.50000	.36842	---	.05263	---
7	.94737	---	---	---	.94737	.63158	.50000	.31579	---	---
8	---	---	---	---	---	---	.68421	.50000	---	.10526
9	---	---	---	---	---	.94737	---	---	.50000	.63158
10	---	---	---	---	---	---	---	.89474	.36842	.50000

Fig. 8(d)—Observed proportions $\hat{P}_{ji}$ for experts, straw picture ($N_{ji} = 19$).

tions of zero or one (see footnote on p. 331). Notice that the notation for the rows and columns is the reverse of that which is normally seen, i.e., the matrix is transposed about the negative diagonal. This is done solely because of the computer program used.

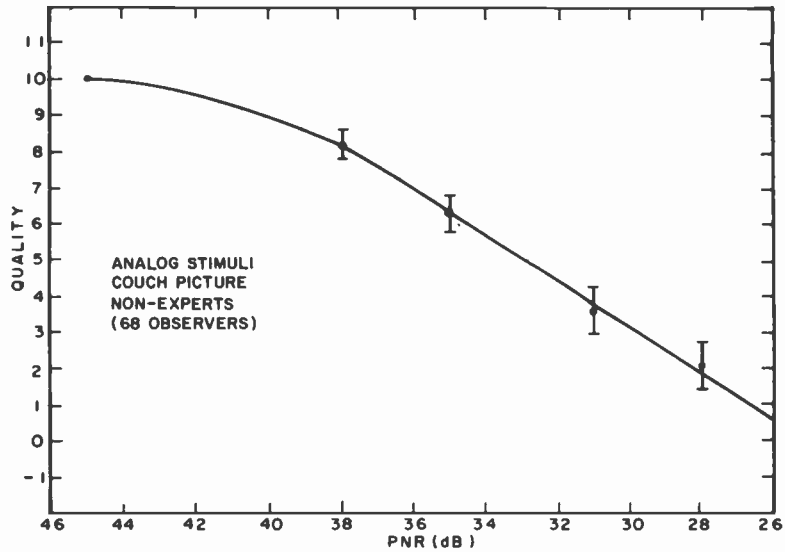
A maximum-likelihood analysis was performed using the data in the proportion matrices. In the nonexpert category, there were 68 people participating for the couch picture and 67 people participating for the straw picture. The logistic response function was chosen over the normal response function as it appeared to yield parameters that could predict the observed proportions more closely. An assumption was made that all discriminal dispersions of analog stimuli are equal and all discriminal dispersions of digital stimuli are equal.* The first and tenth stimuli were fixed at the values "ten" and "one," respectively. Figs. 9 and 10 show the affective values plotted against the stimuli values for the analog and digital stimuli, which were kept separated, for the non-experts. The error bars represent 90% confidence intervals. For the couch picture the chi-square value† is 9.464 with 12 degrees of freedom ($P = 0.663$), while for the straw picture the chi-square value is 16.788 with 11 degrees of freedom ($P = 0.114$).

In the expert category, there were 20 people participating for the couch picture and 19 people participating for the straw picture. The logistic function was again used for the same reason as above. A different assumption was made in this case about the similarity of the types of stimuli; here, it was assumed that all of the discriminal dispersions must be equal to one value.‡ The first and tenth stimuli were again fixed at the same values as above. Figs. 11 and 12 show the affective values plotted against the stimuli values for the experts; again, the error bars represent 90% confidence intervals. For the couch picture the chi-square value is 3.279 with two degrees of freedom ($P = 0.194$) while for the straw picture the chi-square value is 5.474 with 4 degrees of freedom ($P = 0.242$).

* This amounts to a partial Thurstone's Case III assumption where, instead of assuming that all dispersions can be different from each other, one assumes that only analog and digital dispersions can differ. Basically, one is assuming that all analog stimuli are similar to each other in type of impairment and that all digital stimuli are similar to each other, also. This set of assumptions gave the best goodness-of-fit in this case, while still leaving enough degrees of freedom to use the likelihood ratio test.

† One uses chi-square as a goodness-of-fit criterion often in statistics; usually one is trying to find a high chi-square value so as to be able to reject the null hypothesis. Here, however, the opposite is true: the experimenter is trying to obtain a low enough chi-square so that the null hypothesis cannot be rejected since the object of the analysis is to obtain a good fit to the original data (the P matrix). Guilford states[19] that it is probably good practice to adopt a confidence level of $P = 0.01$; i.e., anything higher than this value indicates that the null hypothesis should not be rejected, and therefore one has a good fit (discrepancies between observed and predicted proportions can be explained completely by sampling errors).

‡ This is a Thurstone's Case V assumption where one assumes that all of the stimuli are similar to each other in regard to the type of impairment, or at least the ease with which an observer can distinguish different levels of impairment. This assumption was the only one which could be made and still leave enough degrees of freedom to use the likelihood ratio chi-square test.



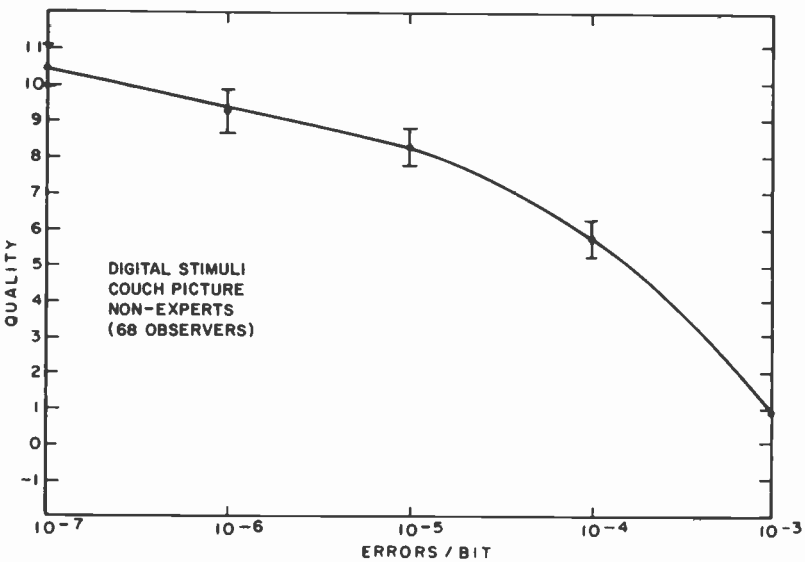
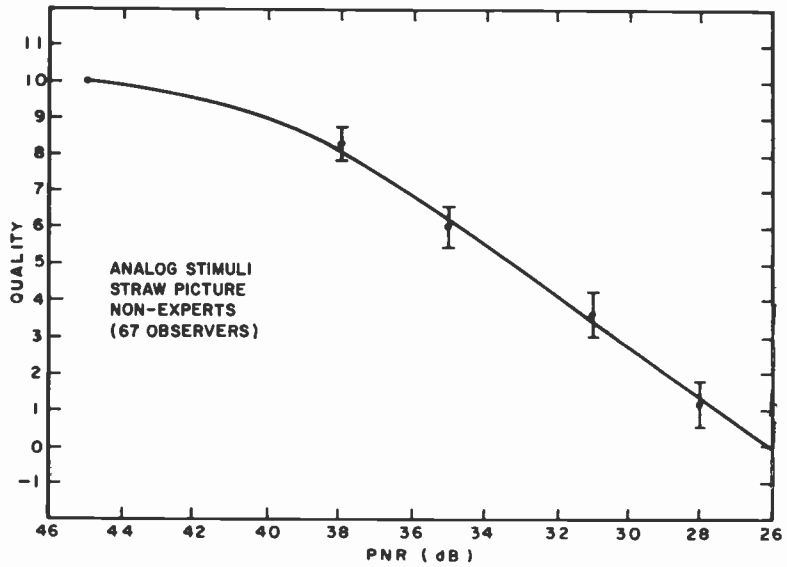


Fig. 9—Analog and digital stimuli versus quality for non-experts—couch picture.



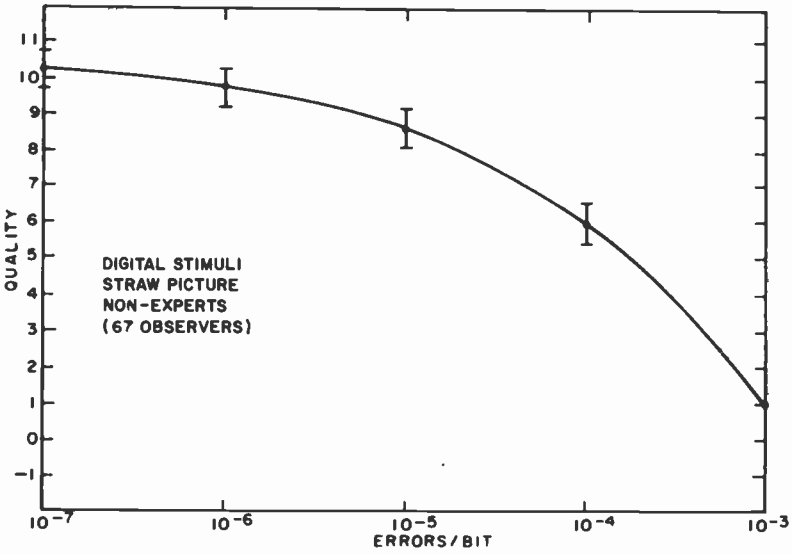
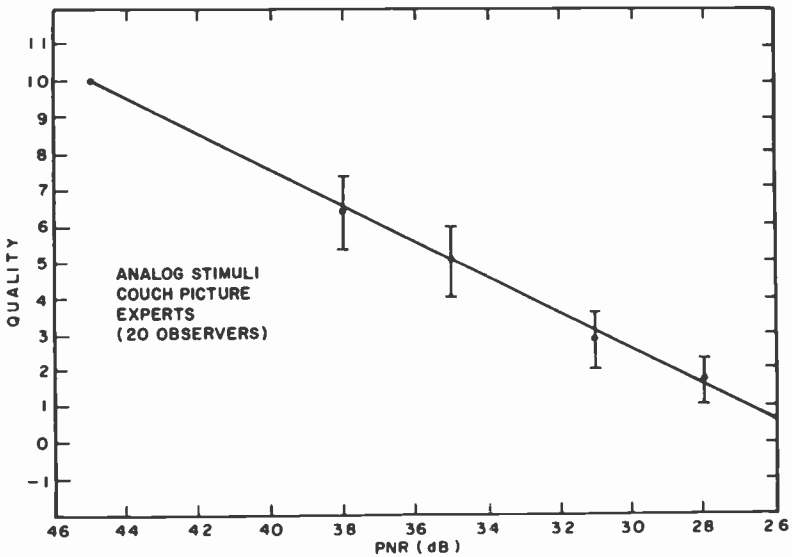


Fig. 10—Analog and digital stimuli versus quality for non-experts—straw picture.



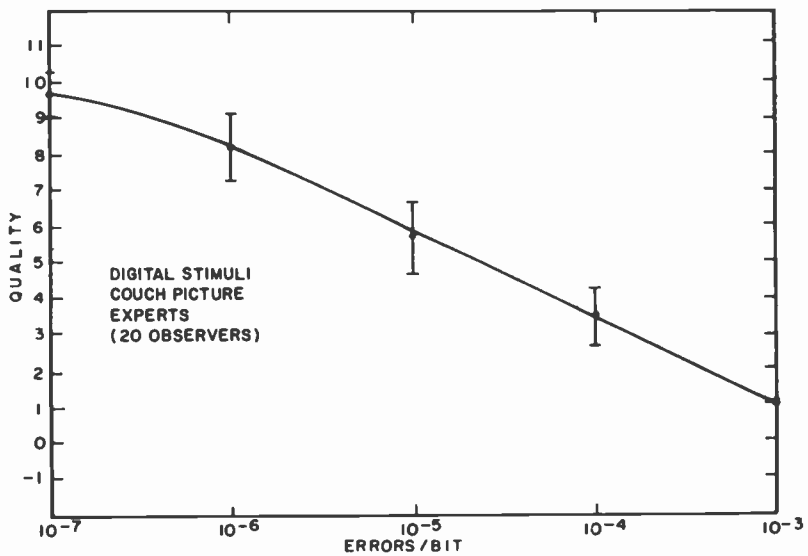
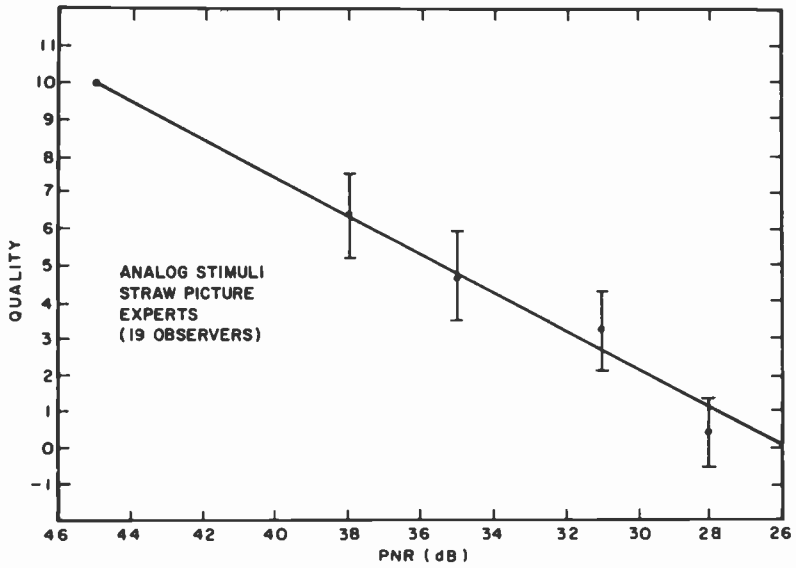


Fig. 11—Analog and digital stimuli versus quality for experts—couch picture.



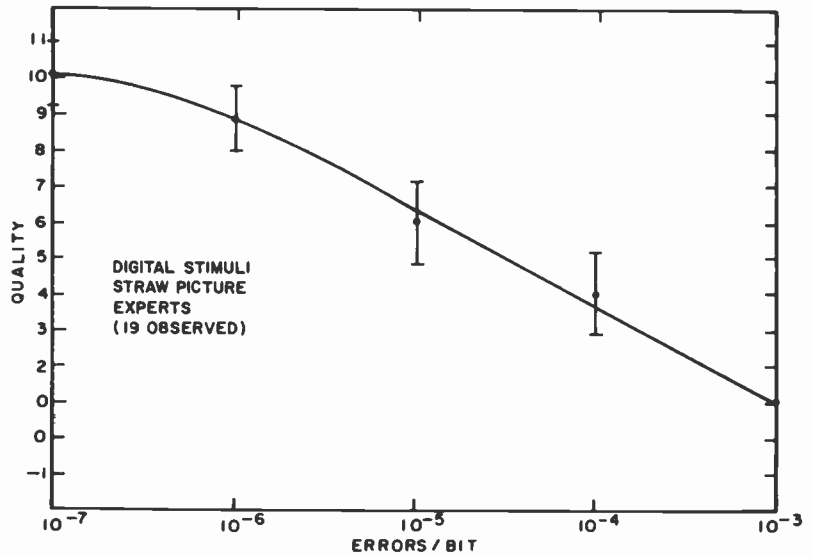
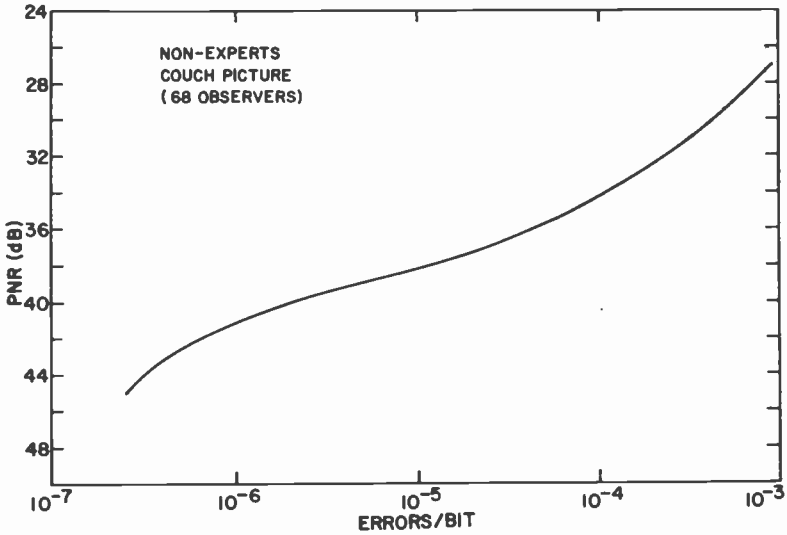


Fig. 12—Analog and digital stimuli versus quality for experts—straw picture.

By comparing the straw and couch graphs for the analog case and then the digital case, one may observe the effect that picture content and brightness have on the subjective picture quality. The first thing that one observes is that the experts and non-experts are almost identical in their responses in this respect. For the analog case, the pictures are judged to be about equally impaired at high picture-to-noise ratios, while at lower PNR's the straw picture is judged to be slightly more impaired with a given noise level. For the digital case, on the other hand, the pictures are judged to be about equally impaired at high bit error rates, while at lower rates the straw picture is judged to be slightly less impaired for a given error rate. These results appear to contradict some previous results obtained by Huang and Chikhaoui,[33] where it was found that the higher the amount of detail in a picture (which in this case was a 2.5 by 2.5 inch hard copy of a monochrome CRT display), the more objectionable the digital noise. Instead, the results obtained in the present work would seem to indicate that the bit error noise (which takes the form of multi-colored dots) is obscured more easily by a picture containing lots of detail when there are only a few scattered dots on the screen. Also, with regard to analog noise, it appears that at very low picture-to-noise ratios the analog noise is more disturbing in a highly detailed picture than in a dark, less finely detailed picture probably because it tends to blur the detail.

The graphs discussed in the preceding few paragraphs can be used to derive the relationships between the analog and digital types of noise by plotting the appropriate analog stimulus value against the digital stimulus value which yields the same level of picture quality. The approximate relationships for the four cases are shown in Figs. 13 and 14. Several relationships may be explored with these graphs. One is the comparison between non-experts and experts for each picture. The impression received, regardless of which picture was being observed, is that the experts judged the bit error noise more harshly than the non-experts when the error rate was around 10^{-4} error/bit. This might be explained by considering that experts spend a lot of time judging standard analog television pictures in terms of quality; most forms of impairment in these images would be expected to appear more like that of Gaussian white noise than of bit error noise. Thus, these experts may have become partially conditioned to seeing analog noise and would judge it more lightly than the dot noise of digital errors.

Another comparison that can be made is that between the two different pictures for each group of observers. For the non-experts it makes little difference at high noise levels which picture is observed, while at low noise levels the digital noise impairment is judged worse than analog noise impairment on the couch picture by about 2 dB. For the experts,



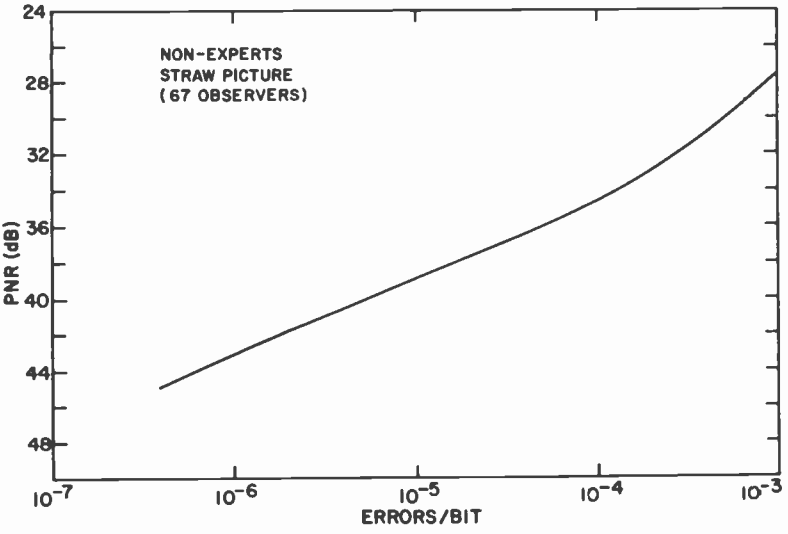
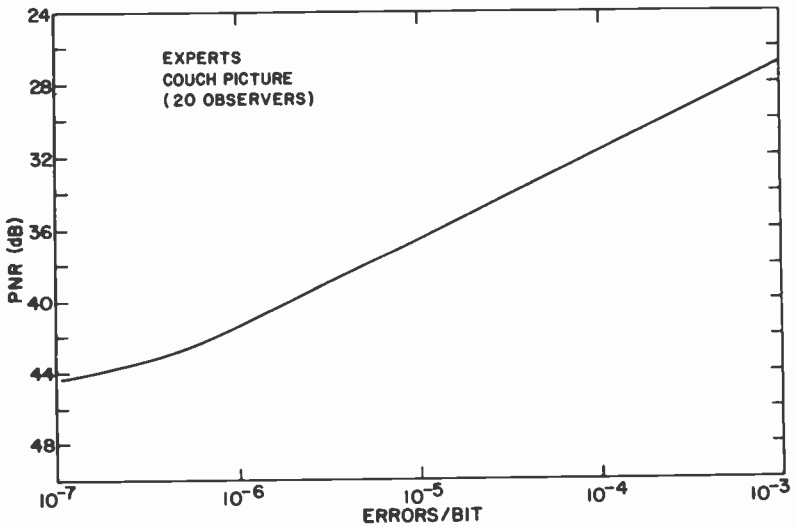


Fig. 13—Error rate versus PNR for non-experts.



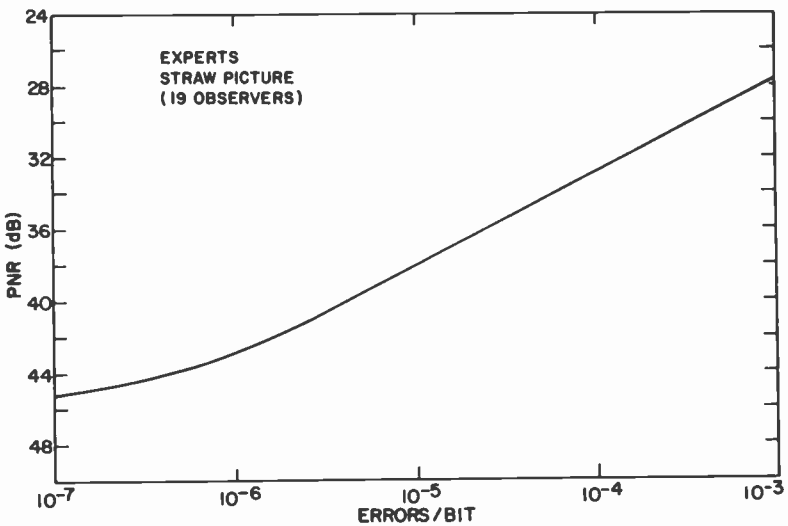


Fig. 14—Error rate versus PNR for experts.

the digital noise impairment is judged worse than the analog noise impairment on the couch picture by about 1–2 dB at all noise levels. This can be partially explained by the same reasoning as before, namely, that bit error noise will be more noticeable in a picture with little or no detail, while analog noise will become more noticeable in a highly detailed picture (at high noise levels only).

Another observation that can be made concerns the general shape of the graphs. The graphs for the non-experts show definite tendencies towards saturation, both at high and low error rates, while the graphs for the experts do not show the same trends. This may be due to the low number of observers in the latter case, which could increase the uncertainty in the extreme regions to the point where the trends might appear to reverse even if, in reality, they have not. This could only be confirmed by using more observers who, in this instance, were not available.

Finally, as was suggested in Sec. 1, the reader may wish to compare the results of this experiment with those of the TASO experiments. If one takes into account the 4.6 dB correction factor, then the error rates may be interpreted in terms of the TASO standard pictures. The results of the random noise interference tests are reproduced from Ref. [14] in Fig. 15 for purposes of comparison. One should exercise caution in this type of comparison. The TASO studies were done many years ago, and for that reason the experimental conditions could not easily be duplicated now; therefore, the conversion factor given here may not be entirely correct although it is the best estimate possible, given the circumstances.

5. Conclusions

This paper has attempted to determine, through the use of subjective testing, the way in which human observers perceive different levels of bit error noise—both by itself and in relation to analog Gaussian white noise. In addition, the effects that subject matter has on these perceptions were investigated. The results may be summed up as follows.

First, by comparing the results obtained in this study with those from the TASO studies, it would appear that a TASO grade 1 picture (PNR = 45 dB) is approximately equivalent to a bit error rate of 3×10^{-7} error/bit; a TASO grade 2 picture (PNR = 31 dB) is approximately equivalent to a bit error rate of 3.5×10^{-4} error/bit; and a TASO grade 3 picture (SNR = 26 dB) is approximately equivalent to a bit error rate of 1.2×10^{-3} error/bit.

Secondly, PCM television picture quality is more dependent on the brightness level and the amount of detail in the picture than is analog picture quality.

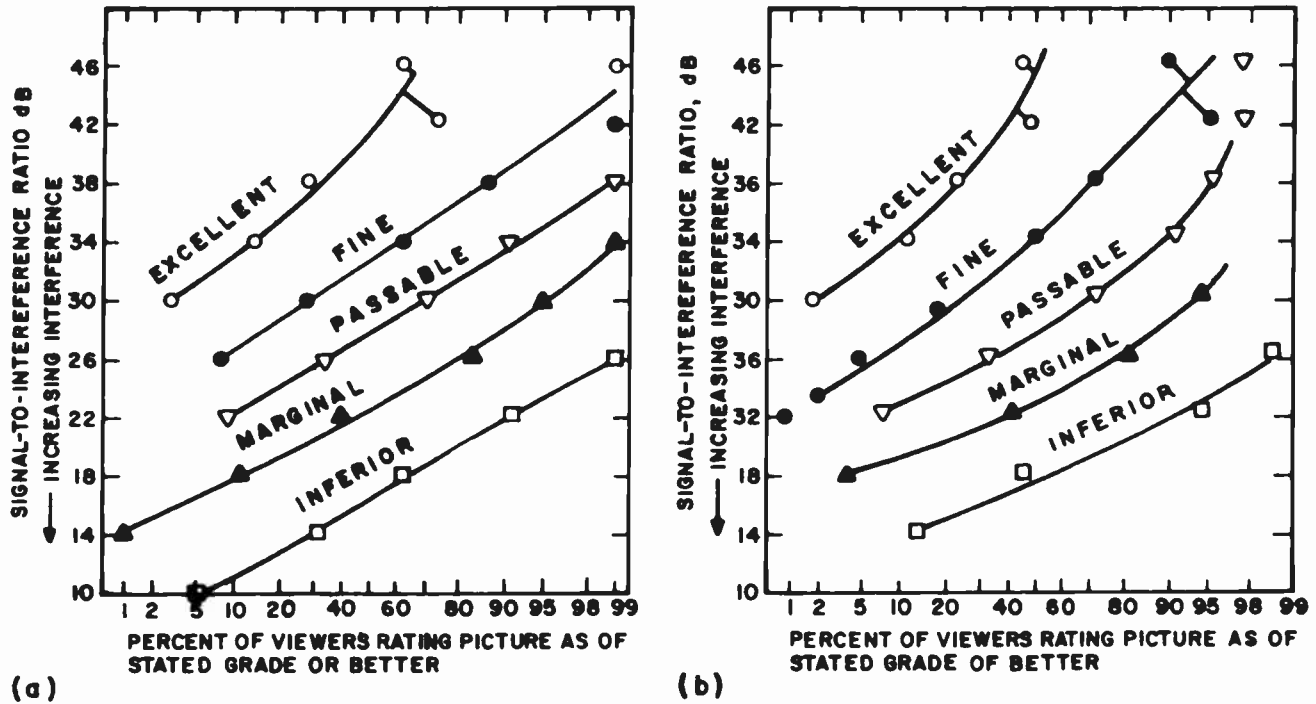


Fig. 15—Partial TASO results: (a) random-noise interference with Miss TASO picture (test 9 using 38 male and 38 female observers); (b) random noise interference average of results from seven scenes not including Miss TASO (all seven tests using 16 female observers).

Finally, apart from the actual experiment performed, the method of subjective testing and the analysis procedures used in this work are useful and can be applied to other problems of this nature.

In designing future experiments, it is helpful to keep in mind several important points. Of prime importance is that no one method of quality evaluation is perfect; all suffer from limitations and inaccuracies. The relative preference method, for example, relies on having a large number of people making a large number of comparisons; clearly, to be conducted effectively, it takes time and energy on the experimenter's part. The results from the present work should be studied carefully, for in this case, too, there are constraints on time and the test equipment that reduce to some degree the accuracy of the results and also their general applicability to a given situation.

A second point to remember is that there are many possible estimation criteria; though they almost always give similar results, it would be wise to check several criteria rather than to rely on any particular one if the time and computational aids are available. The same applies to response functions; it is wise to explore several different ones and to use some criterion such as goodness-of-fit or physical interpretability in deciding which one to use.

Finally, one has to be careful when comparing the results of more than one experiment. Differing conditions can make some types of comparison totally invalid. Only careful investigation can reveal this, but it is usually worth the trouble to check, or at least to warn colleagues of possible pitfalls.

The results described in this paper are clearly not the end of the story as far as PCM television picture quality is concerned. There is certainly a need for more investigations of this sort. Instead of having equal error rates on all bit channels, it would be useful to explore the effects on picture quality with combinations such as those which might be found in error-correcting systems. Also, the PCM system itself can be improved by any number of techniques, such as weighting and companding, and these are bound to affect the way the system responds to noise. Finally, there are other aspects, aside from bit error noise that should be explored, such as quantization noise effects and sampling rate effects. The techniques described in this paper are in fact useful for a wide variety of evaluation applications (keeping in mind that they are useful for unidimensional problems only).

Acknowledgments

I deeply appreciate the help and guidance of Professor Barry A. Blesser, my thesis supervisor, and Mr. Charles B. Oakley, my project supervisor

at RCA Laboratories. In addition, I owe a great deal of thanks to Dr. Joseph Mezrich, of RCA Laboratories, for helping to guide me through the theory and practice of subjective testing. I am also indebted to Dr. Guy W. Beakley, Dr. Hyung K. Kim, and Mr. H. Garton Lewis for their help with the equipment measurements, and testing procedures at RCA Laboratories.

Appendix 1—Independence of Forced-Choice Task from Observer Criteria

In a yes–no task, the observer is presented one of two mutually exclusive stimulus alternatives. He responds by selecting one of the two allowable response alternatives. Each trial of a yes–no task may be represented by a stimulus–response matrix such as that shown in Fig. 16. Stimulus alternatives are denoted by s, which indicates that a signal was detected,

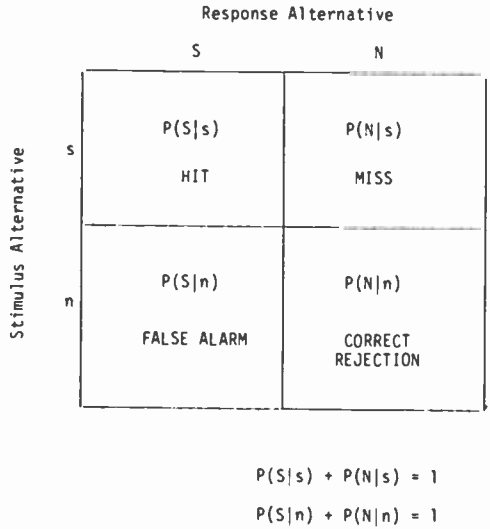


Fig. 16—Stimulus-response matrix of a yes-no trial.

and by n, which indicates the absence of a signal. The responses are denoted by S and N. A cell in the stimulus–response matrix represents the combination of a stimulus and a response, of which there are four possible. Note that the entries in the cells are conditional probabilities, e.g., $P(N|s)$ is the probability that a given stimulus event s will evoke a response N. Since only two numbers can be freely entered in the matrix, the matrix has only two degrees of freedom. Conditional probabilities

are used so that the number of times that an alternative is presented does not overtly affect the data and mask the observer's behavior, which is our main concern.

Because there are only two degrees of freedom in the stimulus–response matrix, all of the information in the matrix can be represented as a point on a two-dimensional graph, which is commonly called the "receiver-operating-characteristic" graph, or the ROC graph. A typical ROC graph is shown in Fig. 17. To see how this graph is determined, let's assume that the estimate of the probability of an observer responding S when n was presented was 0.20, and the corresponding estimate of the probability of the observer responding S when s was presented was 0.40.

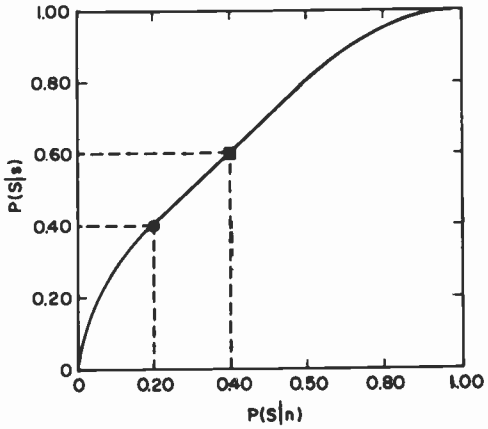


Fig. 17—An ROC graph.

This point is plotted as a circle on the ROC graph. Now suppose that the observer is induced to change his decision criterion in some way (no physical parameter of the stimulus situation is changed in any way). Say, for example, that he was indiscriminately rewarded for saying "yes." In this case we would expect both coordinate values of the circle to increase; i.e., the observer might increase his hit probability $P(S|s)$ from 0.40 to 0.60, but his false alarm probability $P(S|n)$ would also increase, say, from 0.20 to 0.40. The resulting new point is denoted by the square in the graph. These points can be connected by a smooth curve that represents the various modes of observing that the observer might adopt under a particular set of fixed stimulus conditions.

By calculations, which need not be gone into here, it can be determined that the percentage of correct responses in a two-alternative forced-

choice task (which differs from a yes-no task in that a signal occurs in one of two spatial or temporal intervals and the observer is forced to choose one of them) is given by the area under the ROC curve. Now, the area under the curve provides a convenient and simple index of the detectability of the signal because no assumption is made about the form or character of the underlying observer criteria when calculating the area. Hence, the percentage correct estimated for a forced-choice task is not dependent on the observer's criteria for choosing one signal over another. Since the response probabilities P_{ab} fall in this category, they also are not dependent on the observer's criteria for choosing stimuli.

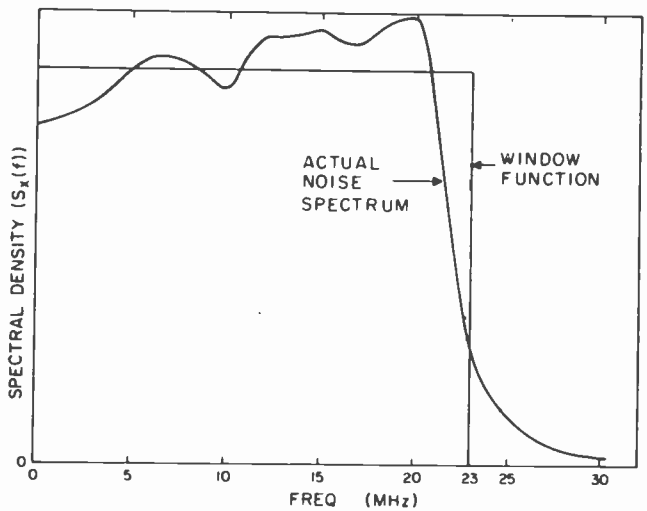


Fig. 18—Spectral density of noise generator output.

Appendix 2—Discussion of Correlation in Error Streams

The spectral density of the random noise generators used in the error generator is shown in Fig. 18. This spectral density can, for all practical purposes, be approximated by the window function, which is superimposed on it. For sampled band-limited white noise (which is what this window function represents), the correlation between samples may be found in the following manner.[34]

Given band-limited white noise (see Fig. 19(a)), the autocorrelation function may be found by using the inverse Fourier transform;

$$R_x(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_x(\omega)e^{j\omega\tau}\,d\omega \qquad [15]$$

$$= \frac{1}{2\pi}\int_{-2\pi W}^{2\pi W} S_0 e^{j\omega\tau}\,d\omega = \frac{S_0 e^{j\omega\tau}}{2\pi j\tau}\Bigg]_{-2\pi W}^{2\pi W}$$

$$= \frac{S_0}{2\pi}\frac{e^{j2\pi W\tau} - e^{-j2\pi W\tau}}{j\tau} = \frac{S_0}{2\pi}\sin 2\pi W\tau$$

$$= 2WS_0\left[\frac{\sin 2\pi W\tau}{2\pi W\tau}\right] \qquad [16]$$

This function is shown in Fig. 19(b). In the present case, $W = 23$ MHz and $S_0 = \frac{1}{2}W$. Thus, the peak value of $R_x(\tau)$ is one and $\frac{1}{2}W = 2.17 \times 10^{-8}$ Hz^{-1}. For $\tau = 31$ ns, which is the time between errors from one channel to an adjacent channel being driven by the same noise generator,

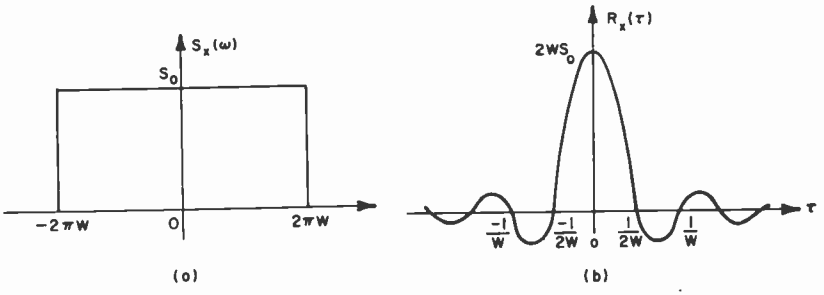


Fig. 19—Bandlimited white noise: (a) spectral density and (b) autocorrelation function.

$R_x(\tau) = -0.217$. For $\tau = 62$ ns, which is the time between every other error, $R_x(\tau) = 0.050$. For $\tau = 93$ ns, which is the time between errors on a single noise channel, $R_x(\tau) = 0.057$. This indicates that the errors are almost uncorrelated within streams and are slightly correlated with certain adjacent error streams, though probably not noticeably.

The previous discussion points out one more thing. The correlation within, and between, error streams can be made increasingly smaller by increasing the time between samples, i.e., increasing the delays in the sample-and-hold circuits from a fraction of one data sampling interval to many intervals. This may be accomplished, with some increase in

expense and complexity, by using shift registers to add the desired extra amount of delay. The amount of correlation in this case is found exactly the same way as above, except that the values of τ will be much larger, giving much smaller correlations (by putting us further out on the tails in Fig. 19(b)). In the present case, it was felt that this was not necessary as the correlation did not appear to be perceivable.

References:

[1] A. K. Bhushan, Master's Thesis, M.I.T., 1967.

[2] B. M. Oliver, J. R. Pierce and C. E. Shannon, "The Philosophy of PCM," *Proc. IRE*, **36**, p. 1324, Nov. 1948.

[3] E. M. Deloraine and A. H. Reeves, "The 25th Anniversary of Pulse-code Modulation," *IEEE Spectrum*, **2**, p 56, May 1965.

[4] "IEEE Recommended Practice for Speech Quality Measurements," *IEEE Trans. on Audio and Electroacoustics*, **AU-17**, No. 3, Sept. 1969.

[5] M. H. L. Hecker and N. Guttman, "A Survey of Methods for Measuring Speech Quality," *J. Audio Eng. Soc.*, **15**, p. 400 (1967).

[6] R. W. Donaldson and D. Chan, "Analysis and Subjective Evaluation of Differential Pulse-code Modulation Voice Communication Systems," *IEEE Trans. on Comm. Tech.*, **COM-17**, No. 1, Feb. 1969.

[7] R. W. Donaldson and R. J. Douville, "Analysis, Subjective Evaluation, Optimization, and Comparison of the Performance Capabilities of PCM, DPCM, ΔM, AM, and PM Voice Communication Systems," *IEEE Trans. on Comm. Tech.*, **COM-17**, No. 4, Aug. 1969.

[8] J. M. Barstow and H. N. Christopher, "The Measurement of Random Monochrome Video Interference," *Trans. AIEE*, **73**, Pt. 1, p. 735, Jan. 1954.

[9] R. C. Brainard, F. W. Kammerer and E. G. Kimme, "Estimation of the Subjective Effects of Noise in Low-Resolution Television Systems," *IRE Trans. on Info. Theory*, **IT-8**, p. 99, Feb. 1962.

[10] J. M. Barstow and H. N. Christopher, "The Measurement of Random Video Interference to Monochrome and Color Television Pictures," *Trans. AIEE*, **81**, Pt. 1, p. 313, Nov. 1962.

[11] W. D. Voiers, M. F. Cohen, and J. Michunas, "Evaluation of Speech Processing Devices I, Intelligibility, Quality, Speaker Recognizability," AFCRL-65-826, 1965.

[12] *Levels of Picture Quality*, TASO Rept. to the FCC, Panel 6, January, 1959.

[13] G. L. Fredendall and W. L. Behrend, "Picture Quality—Procedures for Evaluating Subjective Effects of Interference," *Proc. IRE*, p. 1030, June 1960.

[14] C. E. Dean, "Measurements of the Subjective Effects of Interference in Television Reception," *Proc. IRE*, p. 1035, June, 1960.

[15] J. R. Cavanaugh, "A Single Weighting Characteristic for Random Noise in Monochrome and NTSC Color Television," *J. SMPTE*, **79**, p. 105, Feb. 1970.

[16] C. B. Grether and R. W. Stroh, "Subjective Evaluation of Differential Pulse-code Modulation Using the Speech 'Goodness' Rating Scale," *IEEE Trans. on Audio and Electroacoustics*, **AU-21**, No. 3, June 1973.

[17] J. R. Cavanaugh and A. M. Lessman, "The Subjective Effect of Random Noise Spectra on 525-Line NTSC Color Television," *J. SMPTE*, **83**, p. 829, Oct. 1974.

[18] L. L. Thurstone, "A Law of Comparative Judgment," *Psychol. Rev.*, **34**, p. 273 (1927).

[19] J. P. Guilford, *Psychometric Methods*, McGraw-Hill, New York, 1954.

[20] W. S. Torgerson, *Theory and Methods of Scaling*, Wiley, New York, 1958.

[21] R. D. Bock and L. V. Jones, *The Measurement and Prediction of Judgment and Choice*, Holden-Day, San Francisco, 1968.

[22] J. L. Arbuckle and J. H. Nugent, "A General Procedure for Parameter Estimation for the Law of Comparative Judgment," *Brit. J. Math. and Stat. Psychol.*, **26**, p. 240 (1973).

[23] H. Gulliksen, "A Least-Squares Solution for Paired Comparisons with Incomplete Data," *Psychometrika*, **21**, p. 125 (1956).

[24] L. R. Tucker, "A Maximum-Likelihood Solution for Paired Comparisons by Thurstone's Case V," University of Illinois Technical Report (1964).

[25] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Wiley, New York, 1966.

[26] A. W. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill, New York, 1967.

[27] J. L. Arbuckle and J. H. Nugent, "A General Program for Parameter Estimation in Thurstone's Law of Comparative Judgment," *Behavioral Science*, **18**, p. 228 (1973).

[28] H. Gulliksen, personal communication.

[29] J. J. Gibson, personal communication.

[30] "Applications of Pseudo-Random Sequences," Tau-Tron, Inc. Technology Note, No. 1M 7211-01.

[31] "A Tutorial on the Properties of Pseudo-Random Sequences," Tau-Tron, Inc., Technology Note, No. 1M 7212-01.

[32] W. W. Peterson, *Error-Correcting Codes*, M.I.T. Press, Cambridge and Wiley, New York, pp. 251–270, 1961.

[33] T. S. Huang and M. T. Chikhaoui, "The Effect of BSC on PCM Picture Quality," *IEEE Trans. on Info. Theory*, **IT-13**, No. 2, April 1967.

[34] G. R. Cooper and C. D. McGillem, *Methods of Signal and System Analysis*, Holt, Reinhart and Winston, New York, 1967.

[35] G. W. Beakley, personal communication.

Al_2O_3 as a Radiation-Tolerant CMOS Dielectric*

K. M. Schlesier, J. M. Shaw, and C. W. Benyon, Jr.

RCA Laboratories, Princeton, N.J. 08540

Abstract—Complementary-symmetry MOS circuits with their high noise immunity and low power consumption are excellent candidates for many stringent military and aerospace applications. However, their sensitivity to high-energy ionizing radiation often restricts their use. By employing Al_2O_3 as the gate dielectric, CMOS circuits that withstand up to 10^8 rads (Si) of ionizing radiation can be made. Al_2O_3 has been applied to bulk CMOS circuits with aluminum gates and to CMOS in silicon-on-sapphire with both aluminum gates and self-aligned polysilicon gates. In this paper the fabrication, characterization, and limitations of Al_2O_3 as a gate dielectric are presented. We also present an evaluation of Al_2O_3 as a radiation-hard dielectric as compared to other hardening approaches, which are based on improving the standard SiO_2 gate dielectric.

1. Introduction

In the early nineteen sixties it was believed that MOS transistors would be inherently "hard" to high-energy radiation since they are majority-carrier devices and do not rely on high lifetime for successful operation. Lifetime degradation due to neutron bombardment was known to be a major cause of bipolar transistor failure in a radiation environment. It has been proved that MOS transistors are, in fact, very insensitive to neutron damage.[1] However, it was soon discovered that ionizing radiation (x-rays, gamma rays, and high-energy charged particles) produced substantial threshold shifts and gain degradation in MOS transistors.

* Work was supported by Air Force Cambridge Research Laboratory under Contract F19628-73-C-0146, by Air Force Avionics Laboratory under Contracts F33615-73-C-1100 and F33615-74-C-1175, and by Office of Naval Research under Contract N00014-73-C-0277.

Ionizing radiation creates mobile carriers in the gate dielectric; and motion and semipermanent trapping of these carriers contributes both to space charge in the gate insulator and to interface states at the silicon/gate dielectric interface.

Radiation-induced threshold shifts are generally more severe when there is a positive gate bias during irradiation. In 1965 Zaininger proposed that ionizing radiation produces free hole–electron pairs in the gate dielectric that can move from the point of creation due to an applied bias.[2] The electrons are readily swept out of the insulator, while the holes tend to be trapped near the most negative electrode. With positive bias the holes are trapped close to the silicon, where they have the strongest effect on the silicon. The basic features of this model have been substantiated through a variety of experiments.[3–7]

Programs were initiated to either improve the SiO_2 to minimize hole trapping or to replace it with a dielectric that did not have this property. Both of these approaches have resulted in substantially radiation-hardened CMOS circuits. The interested reader will find some of the more fruitful approaches to SiO_2 hardening described in Ref. [8]–[14].

In 1969 Zaininger and Waxman discovered that MOS devices made with Al_2O_3 gate dielectric had much less radiation-induced charge than then available devices with SiO_2 dielectric.[15] They formed the Al_2O_3 by the complete plasma anodization of a thin Al film that was evaporated onto silicon. Al_2O_3 has also been formed by chemical vapor deposition using both aluminum isopropoxide[16] and aluminum chloride[17] as carrier sources. These forms of Al_2O_3 also showed the radiation hardness of the plasma-anodized material. All of the work described in this paper deals with Al_2O_3 formed by chemical vapor deposition from aluminum chloride, since this form was found to be more suitable for making CMOS devices.

In addition to the relatively permanent radiation effects in the dielectric, there are transient radiation effects that occur in all integrated circuits. These transient effects are observed when there is a short high-intensity burst of ionizing radiation even when the total radiation dose is too small to cause threshold shifts in transistors. Radiation bursts cause photocurrents to flow in p-n junctions; the currents are generally proportional to the instantaneous radiation dose rate. Transient photocurrents can cause logic upset, circuit latch-up, and even circuit burnout. The most widely accepted technique for suppressing transient photocurrents is to provide dielectric isolation between transistors. In CMOS circuits this can be conveniently done by building the circuits in a thin film of silicon on a sapphire substrate (SOS). The transistors are constructed on individual silicon islands that are completely isolated from surrounding devices. Within the transistors, the relatively poor

lifetime of SOS material also serves to decrease photocurrents. In SOS there are finite photocurrents that flow through the sapphire, but these are short-range currents and provide a conductive path between source and drain of an individual transistor but not between adjacent transistors. By careful circuit design these shunting paths can be accounted for and the transient radiation upset level can be made as high as 10^{10} rads (Si)/sec compared to 10^{8} rads (Si)/sec for conventional CMOS circuits.[18]

SOS construction provides an additional silicon/insulator interface where the epitaxial silicon contacts the sapphire substrate. This interface is subject to charge build-up and interface state creation as is the interface between silicon and gate dielectric.[12,19,20] The problem occurs primarily in n-channel devices where positive charge build-up creates an inversion layer at the silicon/sapphire interface. This "back channel" effect dramatically increases the apparent leakage current of SOS devices and is a major problem in making radiation hard CMOS/SOS devices. The problem will not be discussed in this paper but is included as a caveat for the reader.

The approach to making CMOS circuits that are tolerant of both total dose and transient radiation effects is to make them in silicon on sapphire (for transient hardness) with a hard gate dielectric. The use of SOS provides additional benefits in terms of speed and speed–power product. To enhance these properties the circuits should be made with self-aligned silicon gates to achieve maximum speed benefit. Si gate/Al_2O_3/CMOS/SOS circuits have been successfully fabricated in both simple inverters and complex LSI arrays. The remainder of this paper deals exclusively with considerations involved with using Al_2O_3 as the gate dielectric for CMOS/SOS circuits.

2. Deposition of Al_2O_3

2.1 Basic Process

The Al_2O_3 films investigated here have been formed by the hydrolysis of aluminum chloride vapors at a heated surface by the reaction[21]

$$2AlCl_3 + 3H_2 + 3CO_2 \rightarrow Al_2O_3 + 6HCl + 3CO. \qquad [1]$$

The formation of Al_2O_3 occurs in two steps; CO_2 and H_2 react to form H_2O and CO and the H_2O then reacts with $AlCl_3$ to form Al_2O_3 and HCl. The bulk of the gas flow is H_2, and the reaction is limited by the availability of CO_2 and $AlCl_3$. Therefore, the gaseous by-products consist mostly of H_2, with small amounts of CO and HCl. The H_2O formation occurs at temperatures above 750°C, so that the deposition takes place

only in the immediate vicinity of the rf-heated susceptor block that holds the wafers. The reaction of H_2O and $AlCl_3$ can occur at room temperatures, where it can produce a hydrolyzed form of Al_2O_3. It is important to keep the gas lines bearing $AlCl_3$ free of H_2O contamination to avoid clogging the system.

At room temperature, $AlCl_3$ is a solid with a relatively high vapor pressure. To transport $AlCl_3$ as a gas it is heated in a sublimator and a small amount of H_2 is passed over its surface. These gasses are transported to the reactor in a 180°C heated line to prevent condensation of

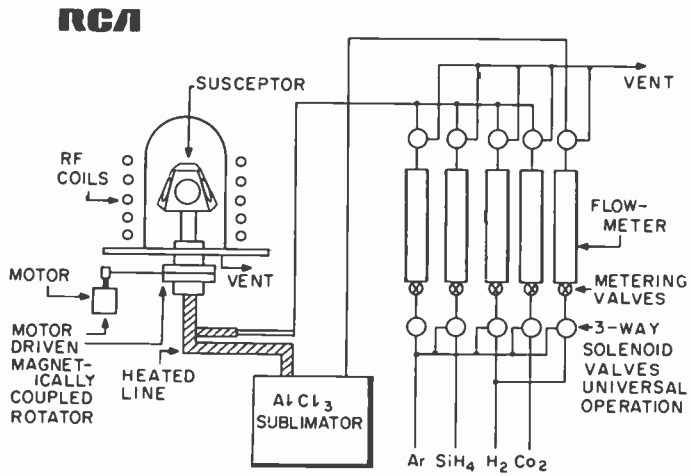


Fig. 1—Schematic of the Al_2O_3 deposition system.

$AlCl_3$. The sublimator temperature is maintained at 125°C where the $AlCl_3$ vapor pressure is approximately 12 Torr. This is at the approximate "knee" of the $AlCl_3$ vapor-pressure curve. Temperatures slightly above 125°C liberate excessive amounts of gaseous $AlCl_3$, while temperatures below 125°C severely limit the $AlCl_3$ vapor supply. A constant sublimator temperature is essential for maintaining process control.

The main flow of the H_2 and CO_2 is brought to the reactor by a separate line. This line also carries the silane mixture when the reactor is used for the deposition of polycrystalline silicon layers.

A schematic of the system is shown in Fig. 1. The depositions are carried out in an epitaxial type reactor fitted with carbon susceptor cut in a barrel configuration and capable of holding five 2¼-inch-diameter substrates. Susceptor rotation is provided by magnetically coupling the pedestal mount to a variable-speed a-c motor. A heavy wall quartz bell

jar is used to enclose the susceptor. To ensure that no $AlCl_3$ vapors condense on the walls of this jar, no artificial means of cooling such as a water jacket are provided. During a routine deposition, the walls of the jar reach temperatures in excess of 300°C with no resultant condensation. Power is supplied by a 15 kW rf generator.

2.2 Al_2O_3 for CMOS Applications

For CMOS applications a gate insulator should have a low charge density, low surface-state density, low pinhole count, and high breakdown voltage. Initial work with the $AlCl_3$ process resulted in Al_2O_3 with a high negative fixed charge and large numbers of surface states. The deposition parameters were refined to achieve virtually no fast interface states and only $3 \times 10^{11}/cm^2$ negative charges. This density may appear high but causes only a small transistor threshold shift due to the high dielectric constant (8.5) of Al_2O_3.

Most important in producing suitable Al_2O_3 films are the deposition rate, the $AlCl_3/CO_2$ ratio, and post-deposition anneals. A high growth rate tends to produce more negative charge; rates of 50 to 100 Å/minute are most acceptable. The deposition rate should be limited by the amount of CO_2 available since excessive CO_2 results in more negative charge. Fast growth rates and excessive CO_2 can both lead to a partially hydrolyzed form of Al_2O_3 that may be the cause of negative charge. On the other hand, it is important not to have excessive amounts of $AlCl_3$ present. This invariably results in large numbers of surface states and hysteresis in final Al_2O_3 devices. The best results are obtained with a near stochiometric ratio of $AlCl_3/CO_2$ but slightly favoring the $AlCl_3$ supply. After deposition it is necessary to "soak" the Al_2O_3 films in H_2 for about 15 minutes to reduce negative charge.

Deposition temperature is important in the sense that increased temperature speeds up the growth rate. Initially much of the Al_2O_3 work was done with aluminum gates, where a relatively low deposition temperature of 850°C can be used. This temperature is suitable since after deposition, no more high-temperature steps are utilized in the process.

The silicon gate process, however, requires subsequent processing at 1050°C. We have found that a better silicon gate circuit results with a 975°C deposition temperature and reduced $AlCl_3$-CO_2 gas flows to maintain a low growth rate.[22] Even so, the Al_2O_3 must be annealed at 1050°C to permit structural changes to occur prior to the application of silicon gate material. At deposition temperatures above 1000°C the H_2-CO_2 reaction can occur at some distance from the susceptor, so that a gas-phase reaction occurs wherein Al_2O_3 falls in a "snow" onto the

wafers. Suitable Al_2O_3 growths can be performed at temperatures from 850° to 975°C.

3. CMOS/SOS/Al_2O_3 Processing

A number of processing sequences have been developed to produce semiconductor devices with Al_2O_3 as the gate dielectric. We present the processing steps used to make silicon-gate Al_2O_3 CMOS/SOS circuits. In general most of the processing follows the standards used in SiO_2 dielectric processing with a few deviations in anneal temperatures and

Table 1—CMOS/SOS/Al_2O_3/Al Gate Process Flow Chart

Silicon Island Deposition
1. Deposit P-Film
2. Deposit Silane Oxide
3. Photoresist
4. Etch
5. Deposit N-Film
6. Deposit Silane Oxide
7. Photoresist
8. Etch (Silicon Oxide & Silicon)

Channel Definition
1. Oxidize
2. Photoresist
3. Etch
4. Deposit Phos Glass
5. Photoresist
6. Etch
7. Deposit Boron Glass
8. Diffuse
9. Etch (Remove Glasses)

Step Oxide Deposition
1. Deposit Oxide
2. Steam Oxide
3. Photoresist
4. Etch

Al_2O_3 Deposition
1. Deposit Al_2O_3
2. Anneal Al_2O_3 (Slow Cool)

Al_2O_3 Etch
1. Deposit Oxide
2. Photoresist
3. Etch Oxide
4. Etch Al_2O_3
5. Etch (Remove Oxides)

Metallization
1. Deposit Aluminum
2. Photoresist
3. Etch
4. Alloy

Overcoat Oxide
1. Deposit Oxide
2. Photoresist
3. Etch

TEST

etching. Deposition of the Al_2O_3 replaces the thermal growth of SiO_2, and Al_2O_3 must be etched in hot phosphoric acid instead of buffered HF.

Tables 1 and 2 and Figs. 2 and 3 show the processing steps used to form Al-gate and Si-gate devices, respectively. In both processes the n- and p-type silicon islands are formed by two epitaxial growths and pattern definition steps. A single epitaxial growth combined with ion implantation could also be used to form the silicon islands. Al_2O_3 has a negative

oxide charge (Q_{ss}). In order to achieve p-channel transistors with a negative threshold voltage the n-type island should have a doping of $10^{17}/cm^3$ to counter-balance the negative charge. The doping of $2 \times 10^{16}/cm^3$ is used for p-type islands in the n-channel device.

3.1 Al Gate Process

In the Al gate processes the wafers are processed through diffusions to provide source–drain regions. Some of the islands are degenerately doped

Table 2—CMOS/SOS/Al_2O_3/Si Gate Process Flow Chart

Silicon Island Deposition
1. Deposit P-Film
2. Deposit Silane Oxide
3. Photoresist
4. Etch
5. Deposit N-Film
6. Deposit Silane Oxide
7. Photoresist
8. Etch (Silicon & Silicon Oxide)

Al_2O_3/Poly Deposition
1. Deposit Al_2O_3
2. Anneal Al_2O_3
3. Deposit Poly Silicon

Poly Silicon Doping
1. Deposit Doping Glass
2. Diffuse into Poly
3. Etch (Remove Glass)
4. Deposit Oxide
5. Densify Oxide
6. Photoresist
7. Etch (Oxide, Silicon & Al_2O_3)

Drain-Source Doping
1. Deposit Phos Glass
2. Photoresist
3. Etch Phos Glass
4. Deposit Boron Glass
5. Diffuse (1050°C Inert Gas)
6. Etch (Remove Glasses)

Crossover Oxide
1. Deposit Oxide 6000 A
2. Steam Oxide 940°C 20 Min

Al_2O_3 Anneal
1. Soak 950°C Hydrogen
2. Slow Cool
3. Photoresist
4. Etch

Metallization
1. Deposit Aluminum
2. Photoresist
3. Etch
4. Alloy

Overcoat
1. Deposit Oxide
2. Photoresist
3. Etch

TEST

over their entire area. The wafer is coated with a 6000 Å layer of SiO_2 that is then removed from all islands where a transistor is to be formed. The SiO_2 is left on the degenerately doped islands to allow them to be used as crossunder interconnects in the final circuit. This brings us to the point

shown in Fig. 2a. Only the transistor islands are shown in this figure. The Al_2O_3 is deposited and annealed in the Al_2O_3 deposition system (Fig. 2b). Immediately upon removal the wafers are coated with an undoped SiO_2 layer. The contact window pattern is etched into this layer, which then acts as an etching mask to etch Al_2O_3. Al_2O_3 etches in 180°C phosphoric acid at about 100 Å/minute. After Al_2O_3 etching, the overcoat SiO_2 layer is removed in buffered HF (Fig. 2c). During this etch, holes on the Al_2O_3 permit HF to open contact windows to the degenerately doped silicon islands through the 6000 Å crossover oxide. Aluminum is evaporated and defined to complete the circuit (Fig. 2d). An overcoat scratch protection oxide may be applied to protect the final circuit.

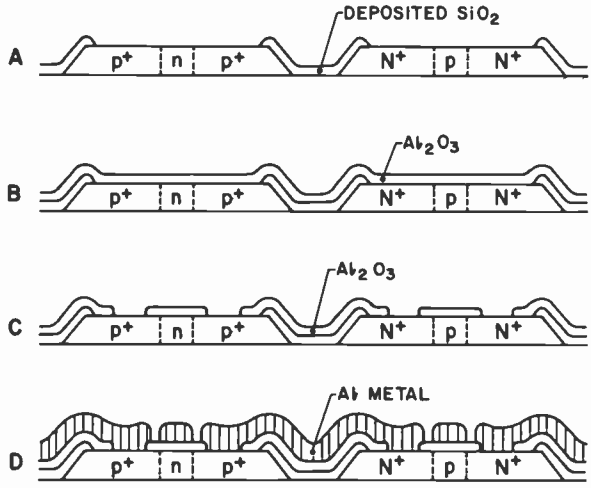


Fig. 2—Some of the process steps used in forming aluminum gate Al_2O_3/CMOS/SOS circuits.

3.2 Si Gate Process

In the silicon-gate process the Al_2O_3 is deposited directly on the silicon islands before source–drain diffusions are formed. Al_2O_3 undergoes irreversible changes when subjected to temperatures higher than the deposition temperature. In order to form source–drain diffusions the wafers must eventually be subjected to 1050°C. If the polysilicon is in place during these diffusions the quality of the Al_2O_3 is irreversible degraded. However, when the Al_2O_3 is subjected to high temperatures without polysilicon, the structural changes do not degrade the Al_2O_3 quality.

Therefore, immediately after Al_2O_3 deposition the wafers are annealed at 1050°C to permit all structural changes to occur prior to polysilicon deposition. The polysilicon layer is then deposited bringing the wafer to the point of Fig. 3a.

The polysilicon layer is doped by diffusion to obtain the correct conductivity (Fig. 3b). An undoped SiO_2 layer is then deposited on the

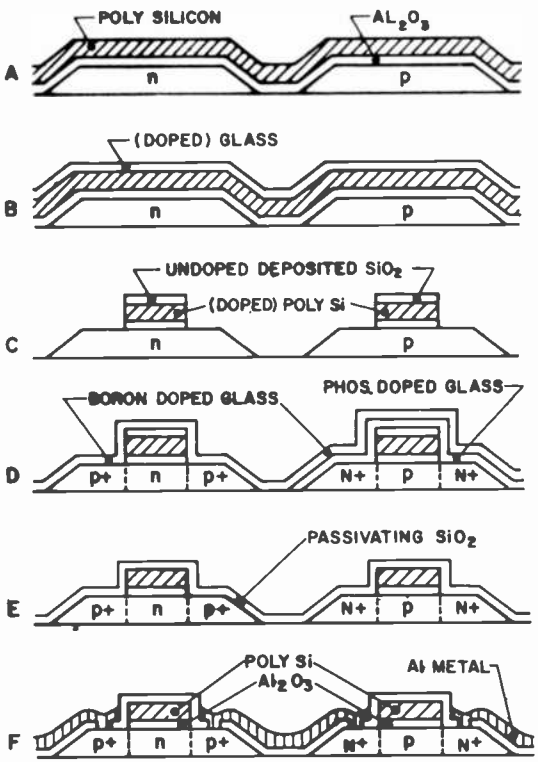


Fig. 3—Some of the process steps used in forming silicon gate Al_2O_3/CMOS/SOS circuits.

poly-silicon and a pattern is formed in this SiO_2 masking layer using buffered HF etchant. The SiO_2 acts as a mask for the subsequent etching of the polysilicon in KOH and the Al_2O_3 in hot phosphoric acid (Fig. 3c). A phosphorus glass layer is deposited and defined to leave it covering only the n-channel devices. A boron glass is deposited and left undefined. The n and p-type source–drain regions are diffused from these glasses (Fig. 3d). After diffusion all doping glasses and caps are removed in HF.

This can be done with the Al_2O_3 process since the channel dielectric is relatively inert to HF etching. A deposition of 6000 Å of undoped silane oxide (SiO_2) is applied to the wafers, which are then oxidized in HCl-steam for 20 minutes. The steam oxide improves the dielectric strength of the gate oxide by passivating the edges of the channel regions. The composite of deposited and grown oxide acts as a crossover insulator for overlapped metal and polysilicon electrodes. Other polysilicon electrodes on the sapphire itself act as crossunder interconnects. Windows are opened to permit aluminum metalization to contact to the source–drain and polysilicon regions (Fig. 3e). The metal is evaporated and defined, and an overcoat oxide may be provided (Fig. 3f).

3.3 Yield Considerations

A useful integrated-circuit process should have the highest possible yield of functional devices. Overall the yield of Al_2O_3 devices is lower than that of comparable SiO_2 devices, but there are several cases where an individual wafer yield compares quite favorably with that of SiO_2 wafers. In particular, several wafers of simple CD 4007 type circuits with aluminum gates on bulk silicon substrates had yields in the range of 70%, which is quite acceptable. In part, lower yields result from working on a laboratory basis, where a great deal of time elapses between processing of one wafer and the next. Yield improvements could be expected if Al_2O_3 wafers were run on a more continuous basis in a well-regulated production facility.

There are, however, some yield problems that are unique to the Al_2O_3 process and these are addressed here. In Al gate circuits on either bulk or SOS, a major yield loss often occurs due to pinholes in the Al_2O_3 film causing gate shorts. By comparing yields on circuits of various complexities and geometries it is apparent that the pinholes are distributed randomly across the area of the film and can be as dense as $2000/cm^2$. Much lower pinhole counts are achieved when Al_2O_3 is deposited on blank silicon wafers. We attribute the pinholes in Al gate circuits to the processing steps used between Al_2O_3 deposition and metallization. In device processing the Al_2O_3 is coated with SiO_2 which is then defined using photoresist and HF and then used as an etch mask to open contact windows in Al_2O_3 with hot phosphoric acid. This SiO_2 mask is stripped in HF prior to metallization. Pinholes in Al_2O_3 could arise from defects in photoresist, defects in the SiO_2 film, and weaknesses in the original Al_2O_3 film that are transformed into pinholes by the deposition and removal of SiO_2. The latter factor is most likely the cause of most Al_2O_3 pinholes and may be related to the cleaning of the wafers and the Al_2O_3 deposition system.

In silicon gate circuits a yield problem arises from nonuniform Al_2O_3 thickness. A circuit structure with dramatically different island sizes will have relatively thin Al_2O_3 on small islands due to their larger periphery/area ratio compared to thicker Al_2O_3 on larger islands. To ensure complete contact opening in the large islands the thin Al_2O_3 on small islands must be severely overetched. In the silicon gate process this will cause undercutting beneath the polysilicon gate material, promoting short circuits. A possible remedy to this problem is the use of ion implantation through the unetched Al_2O_3 to provide source–drain dopant combined with Al_2O_3 etching at the time of contact window opening through the thick SiO_2 crossover oxide. Overetching of Al_2O_3 will then occur at a point further from the gate electrode and thereby decrease the chance for a gate short.

4. Radiation Behavior of Al_2O_3

4.1 Basic Radiation Effects in Al_2O_3

The radiation effects in Al_2O_3 are in some respects similar to those in SiO_2. Ionizing radiation produces free hole–electron pairs that can move under the influence of an applied field and become trapped. There is a marked tendency to trap more holes than electrons so that radiation produces a positive trapped charge (negative threshold voltage shifts). There are four factors that can act to make the net effect of radiation damage appear to be less severe in Al_2O_3 than in SiO_2:

(1) The tendency to trap electrons is greater in Al_2O_3 than in SiO_2 so that both polarities of charge are trapped, thus leading to compensation effects.
(2) The Al_2O_3 dielectric constant is larger so that more net trapped charge is required to produce a given threshold voltage shift.
(3) Under the action of a large electric field, electrons can be injected into the Al_2O_3 from the silicon or gate electrodes and thereby compensate or eliminate the effects of radiation-induced positive trapped charge.
(4) There appears to be little interface state creation and there is no field-effect mobility degradation even after a large radiation dose.

Because of the last feature the radiation damage effects in Al_2O_3 are primarily due to charges trapped within the insulator, although these charges may reside close to the interface.

In one sense there is no permanent radiation damage in Al_2O_3 films. Under continuous radiation with a fixed applied bias, the charges and

fields within the insulator arrange themselves to a steady-state distribution. Usually it takes about 10^6 rads (Si) of continuous irradiation to achieve a steady-state situation. When the radiation is terminated (but bias maintained) the charges will redistribute, and usually a relaxation or decrease in charge is observed. The steady-state distribution can be reachieved by resuming radiation.

The nature of the steady-state distribution (quantity and location of trapped charge) depends critically on the applied bias. If the bias is changed during irradiation, it will take an additional 10^6 rads (Si) to achieve a new steady-state distribution. For example, the steady-state charge distribution corresponding to 10 V bias can be achieved by irra-

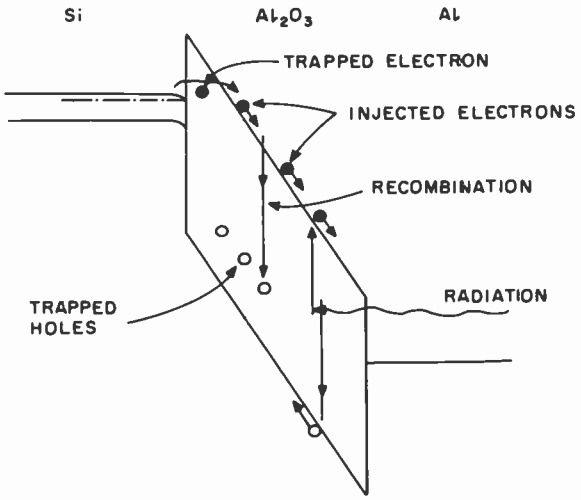


Fig. 4—Band model of Al_2O_3 showing electronic processes that occur during irradiation.

diating a fresh sample to 10^6 rads (Si), or by using a sample that has previously been irradiated up to 10^8 rads (Si) at other biases and irradiating it for an additional 10^6 rads (Si) with 10 V bias. No "permanent" radiation damage is observed up to 10^8 rads (Si) (largest dose tested).

Some of the more important physical processes that occur during irradiation are illustrated in Fig. 4. Ionizing radiation creates free electrons and holes, and some of these become trapped. At present it is not certain whether the holes are uniformly trapped in the bulk of the material or reside in traps that are unique to the interfaces. Because of the electric field at the interface with the negative electrode (silicon in Fig. 4), electronic charges can be injected into the Al_2O_3. The injection mecha-

nism is not clearly understood but is believed to be a combination of tunneling into traps within the Al_2O_3 forbidden gap and of thermal excitation. The injection current density is a strong (nearly exponential) function of the interface field.[23,24] As incoming radiation produces more trapped holes, the interface field is increased. Under some conditions the steady-state situation occurs when the rate at which new trapped holes are induced equals the rate at which they are annihilated by injected electrons.

The charge-injection mechanism can occur in the absence of irradiation by the application of a large bias. In this case the injection mechanism leads to an instability because a number of the injected electrons become trapped near the injecting electrode, producing a negative oxide charge. If this trapping occurs during irradiation, it leads to the possibility that the steady-state condition is actually a case of an ever-increasing negative charge near the injecting electrode compensating an ever-increasing positive charge induced by radiation and hole-trapping. This situation is unlikely because of the large doses, 10^8 rads (Si), through which the steady-state situation prevails. It is more likely that injected electrons recombine with free holes as they are created or annihilate trapped holes.

4.2 Experimental Techniques for Radiation Tests

Al_2O_3 devices were irradiated with 1 MeV electrons produced by the Van de Graaff accelerator at RCA Laboratories, Princeton. This apparatus can supply beam current densities from 5×10^{-9} to 5×10^{-7} A/cm^2 corresponding to dose rates of 10^3 to 10^5 rads (Si)/sec (3×10^7 1 MeV electron/cm^2 are equivalent to 1 rad (Si)). The beam currents were monitored by a Faraday cup, and after the desired beam current was achieved and stabilized, the samples were rotated into the beam path by means of a wheel. The total dose was determined by the time of exposure, and most irradiations were done at 10^4 rads (Si)/sec.

The samples were electrically attached to equipment located outside the radiation chamber. With this equipment the electrical properties of the devices such as CMOS inverter characteristics of MOS C-V plot could be measured without removing the samples from the irradiation chamber. In fact, the devices could be measured even when in the beam path. The radiation produced photocurrents in the silicon but at a dose rate of 10^4 rads (Si)/sec these photocurrents did not significantly perturb the measurements. In this manner the radiation effects could be measured *in situ* without any time between the accumulation of a given dose

and the measurement. This is important in accurately assessing relaxation effects.

To accomplish these measurements an automatic sweep supply was connected to the gate of the device. During most of the radiation period the supply held a constant voltage on the gate. When triggered, the supply would rapidly sweep over a voltage range and return to the holding voltage. During the sweep some characteristic of the device such as transistor current or MOS capacitance was displayed as a function

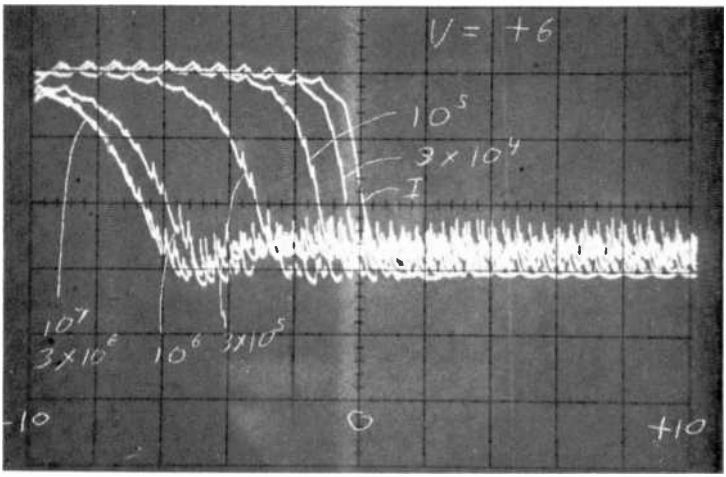


Fig. 5—Capacitance-voltage curves of an Al_2O_3 MOS capacitor at various doses of radiation from 0 to 10^7 rads (Si). The curves were displayed without removing the sample from the radiation beam.

of gate voltage on an X-Y storage scope. The sweep supply was triggered at several radiation levels; and after the desired total dose was achieved, a photograph was made of all curves on the oscilloscope screen. The device could then be rotated out of the beam and its characteristics displayed at various times to study relaxation effects.

4.3 Results of Radiation Experiments

Fig. 5 shows a photograph of the storage oscilloscope after a typical radiation run. These curves show the C-V characteristics of a p-type MOS capacitor with 700 Å of oxide at various radiation levels between 0 and 10^7 rads (Si). In this experiment the radiation rate was 10^4 rads (Si)/sec and the holding potential was 6 V. The noise on the curves is a combi-

nation of 60-Hz pickup and, in depletion, the effects of photocurrents produced by the electron beam as it sweeps across the sample. In this case the radiation effects saturate after 10^6 rads (Si). Irradiation up to 10^7 rads (Si) produces no further change in the C-V characteristic. Several samples were irradiated, each with a different holding bias. The results are summarized in Fig. 6, where the flatband voltage is plotted

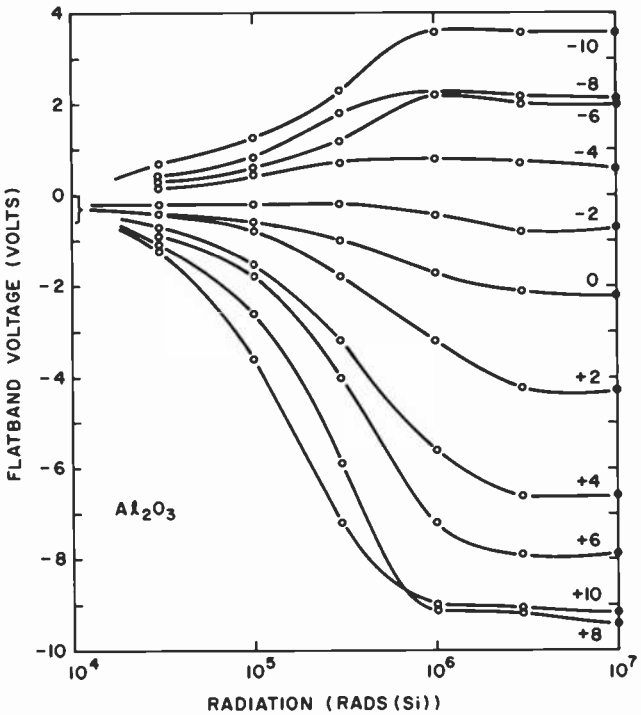


Fig. 6—Flat-band voltage of an Al_2O_3 MOS capacitor as a function of radiation dose with bias during irradiation as a parameter. A fresh sample was used to generate each curve.

as a function of radiation dose with bias as a parameter. For this set of devices the radiation effects were fairly well saturated by 10^6 rads (Si). It should be noted that with negative holding voltage the radiation produced a significant distortion in the C-V curves. The positive shifts in inversion voltage (bottom of C-V curve) were larger than the flat-band voltages shown in Fig. 7. The observation of a positive flat-band voltage

shift for negative voltage during irradiation indicates a net negative charge in the oxide after irradiation.

The saturated flat-band voltage after 10^7 rads (Si) is plotted as a function of irradiation bias in Fig. 7. Data from Fig. 6 as well as from a similar set of curves for a 400-Å sample are shown. This figure illustrates another unique feature of Al_2O_3 radiation behavior. On the 400-Å oxide the radiation-induced shift becomes smaller for biases larger than 4 V. This is attributed to the ability of the silicon electrode to inject electrons into the oxide.

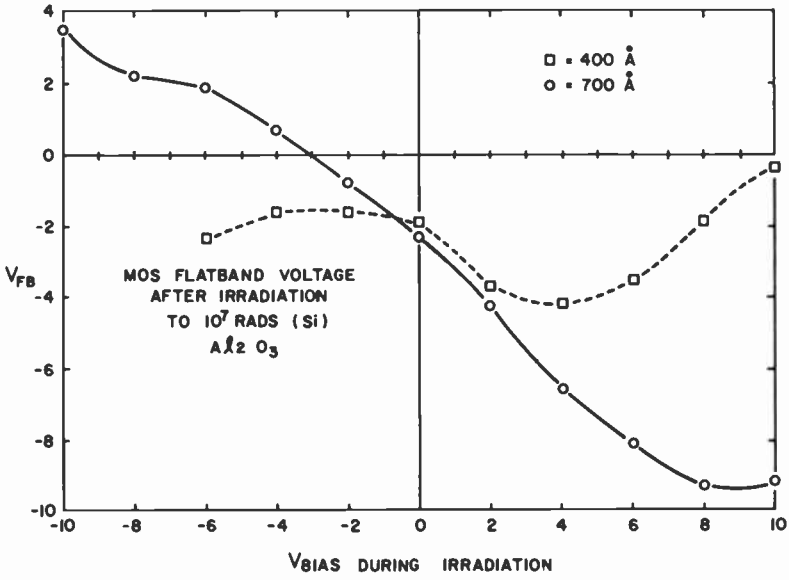


Fig. 7—Saturated flat-band voltage after 10^7 rads (Si) as a function of radiation bais. Each point represents a different MOS capacitor.

Additional data showing the same effect is given in Fig. 8, where data for both MOS capacitors and NMOS/SOS transistors with various oxide thicknesses are shown. These curves were generated by achieving a steady-state situation with 1-MeV electrons at a rate of 10^4 rads (Si)/sec. All curves show similar qualitative results—a small shift under negative bias and a maximum shift under intermediate positive bias. Some of the curves were constructed using new devices to generate each point, while others utilized only one device, re-irradiated at each bias level.

The bias dependence illustrated in Figs. 7 and 8 can be broken into three regimes: (1) negative bias, (2) small positive bias where the applied voltage enhances charge buildup and (3) large positive bias where the applied bias suppresses the charge build-up.

1. Under negative bias the charge build-up is relatively small and sometimes indicates a net negative charge. There is a question as to whether there may be a large positive charge build-up near the gate

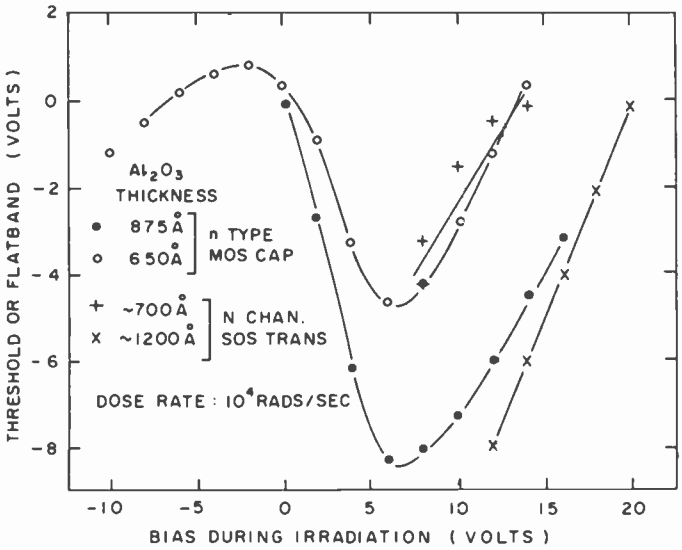


Fig. 8—Steady-state radiation-induced flat-band voltage as a function of gate bias. Each point represents the flat-band or threshold voltage achieved for continuous radiation at a particular bias for more than 10^6 rads (Si). The curves are not significantly altered for radiation doses up to 10^8 rads (Si).

electrode during negative bias. The charge injection mechanism is similar for injection from either the silicon or aluminum gate electrode. The radiation effects may also be similar for each electrode and depend only on the bias. Other than these considerations, little attention has been given to the negative-bias case. This in part arises from the fact that the radiation-induced shifts are acceptably small for negative bias.

2. The effects with small positive gate voltages are similar to those commonly observed in SiO_2. Increasing the bias enhances separation of created hole–electron pairs and/or forces the positive charges to be trapped closer to the silicon interface. These positive charges increase the field at the silicon interface which is given by

$$E_i = \frac{V_{Applied} - V_{FB}}{L}, \qquad [2]$$

and therefore decrease the electric field in the bulk and near the gate electrode. In Eq. [2] and following equations, L is the oxide thickness. In this mode the steady-state situation is believed to occur when the field in the bulk of the insulator becomes too small to cause efficient separation of hole–electron pairs as they are created. Therefore, the radiation-induced effects are limited by recombination in the low-field region of the insulator.

3. As the bias is increased, the steady-state interface field increases to the point where significant amounts of electronic charge injection from the silicon can influence the steady-state distribution. The rate of charge injection is approximately exponential with the interface electric field.[23,24] Therefore, a small increase in electric field can create a tremendous increase of injected electrons, which can recombine with positive charge. Because of this, the interface electric field cannot increase indefinitely, and approaches a maximum value. Therefore, the steady-state flat-band voltage must follow the applied voltage for larger bias levels according to Eq. [2], where E_i is limited to a maximum value. This leads to the decrease in radiation-induced shifts for large positive bias observed in Figs. 7 and 8. In Fig. 9, the steady-state-interface electric field is plotted as a function of bias voltage for two of the MOS capacitors shown in Fig. 8. As the bias is increased, the field approaches a maximum value of about 2.1×10^6 V/cm. From independent measurements of the injected current density in the absence of radiation,[23,24] this field corresponds to an injection current of 6.3×10^{11} electrons/cm^2-sec.

The injection-limited steady-state situation under large positive bias is actually a dynamic balance between the rate of positive-charge creation by irradiation and negative-charge injection due to interface field. Therefore, both a relaxation of radiation effects after radiation is terminated and a dose-rate dependence of the steady-state charge should be observed. These effects are illustrated in Fig. 10 for an n-channel SOS transistor. The upper left-hand portion shows the threshold voltage as a function of irradiation dose for a dose rate of 10^4 rads (Si)/sec. A steady-state threshold voltage, which depends on bias, is reached by 10^6 rads (Si). When radiation is terminated, there is a logarithmic relaxation of positive charge (and/or injection and trapping of negative charge), as shown in the upper right portion. The logarithmic dependence arises from the exponential relation between injection current and interface field. When the bias is kept constant but the dose rate changes, the steady-state threshold voltage shows the dependence illustrated at the bottom of Fig. 10. Again, the logarithmic dependence results from the

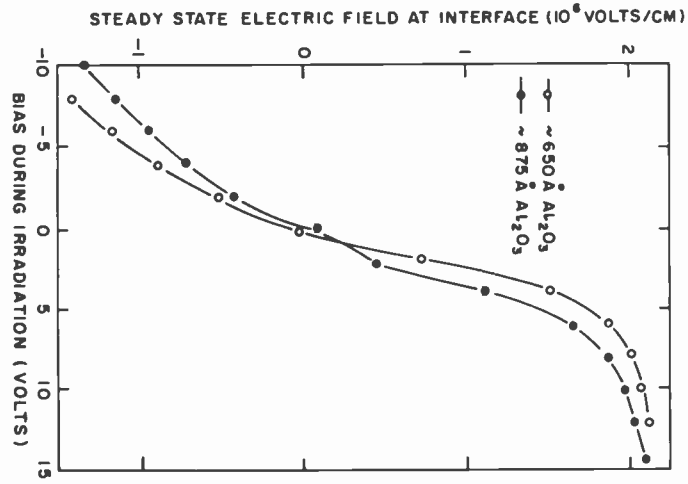


Fig. 9—Steady-state electric field at the Al_2O_3/Si interface as a function of gate bias during irradiation. Note the field limitation under positive bias which occurs for lower voltage on the thinner oxide.

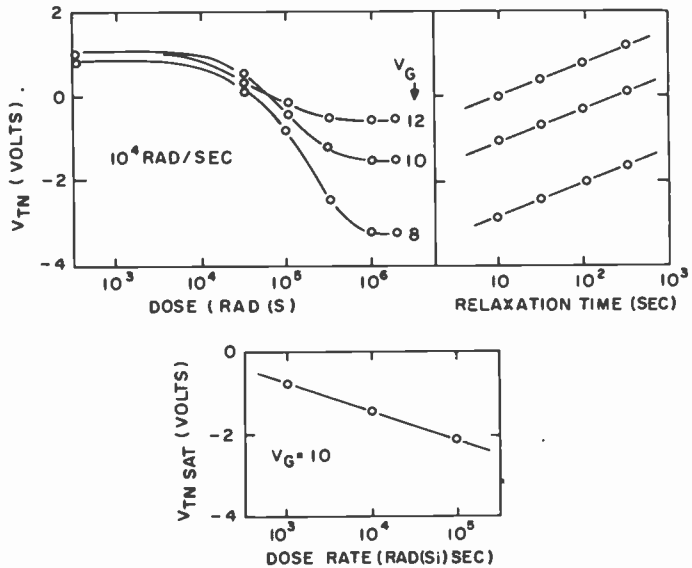


Fig. 10—Radiation, relaxation, and dose-rate data for an Al_2O_3 n-channel SOS device. There is a relaxation effect after radiation is terminated and a dose-rate dependence of the steady-state threshold voltage for a given gate bias.

exponential nature of the injection field characteristic. These results illustrate the proportionality between the charge-injection rate and the radiation dose rate in the injection-limited steady-state case.

4.4 Practical Radiation Limitations for Al_2O_3/CMOS/SOS Devices

A feature of Al_2O_3 that has made it attractive for use with CMOS circuits is that it was thought to be tolerant of both negative and positive bias. This is not exactly true as illustrated in Figs. 7 and 8. Al_2O_3 devices show small threshold shifts under zero and negative bias, but for positive biases

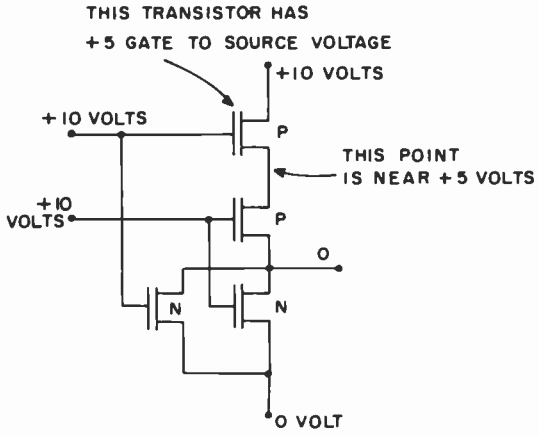


Fig. 11—Circuit for a CMOS/SOS two-input NOR gate showing the bias conditions that result in a +5 volt gate-source voltage with a supply voltage of 10 volts.

the shift is small only when the injection-limiting situation occurs. There are large threshold shifts for intermediate values of positive bias. By choosing the proper oxide thickness (about 650 Å) the devices can be made radiation-tolerant at 10 V. However, unless special precautions are used, a CMOS/SOS circuit operating with a 10-V supply may have a +5-V gate bias on some p-channel devices. This is illustrated in the NOR gate of Fig. 11. When both inputs are high (+10), the series p-channel devices will be cut off and the output will be low (0). Because of the superlinear leakage characteristics of SOS transistors the node between the devices will assume a voltage close to +5 V. Therefore, the upper p-channel device will have a 5-V gate-to-substrate voltage in this case.

The steady-state threshold voltage was measured as a function of bias for both p and n-channel SOS transistors. The results shown in Fig. 12 illustrate the radiation hardness of Al_2O_3 for all bias values except +4 to +10 volts. In this range the upper p-channel device of Fig. 11 could have a post radiation threshold voltage as high as −7.6 V. While this would not result in a static logic failure, it would greatly reduce the response time of the NOR gate.

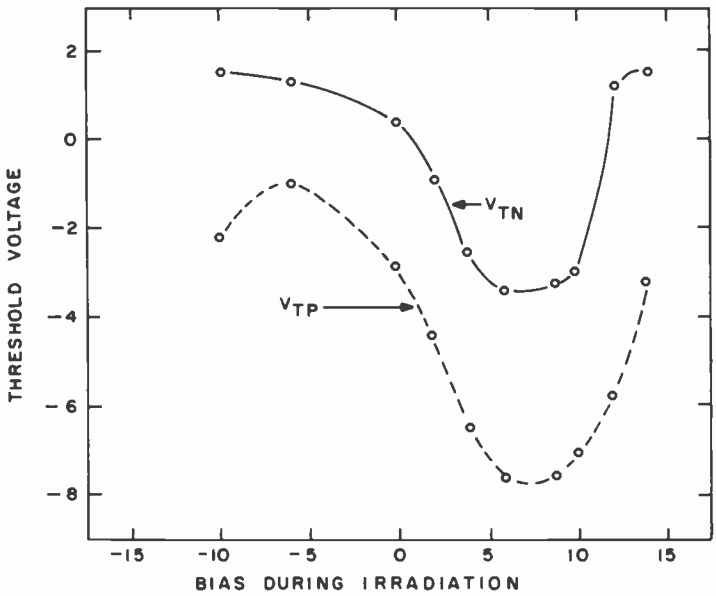


Fig. 12—Steady-state radiation-induced threshold voltages for n- and p-channel Al_2O_3/SOS transistors as a function of gate to source bias.

At present we believe that series-connected p-channel devices are the only place where an intermediate positive bias can occur. This problem could be alleviated by using p-channel substrate contacts that are tied to ground or by eliminating series-connected p-channel strings from the circuit design.

5. Charge Injection and Stability of Al_2O_3 Devices

Aluminum oxide films on silicon exhibit a phenomenon known as charge injection. When a sufficiently high voltage is applied to the gate, electrons can be injected into the oxide from the negative electrode. This affects Al_2O_3 devices in two ways:

(1) The ability to inject charge plays a role in the radiation behavior of the device.
(2) If the injected charges are trapped in the oxide, they can cause a shift in the threshold voltage of MOS transistors.

The first effect is discussed extensively in Sec. 4, and it is clear that from a radiation hardness standpoint, charge injection is a desirable feature of Al_2O_3. In this section we discuss the effects of charge injection on the stability of Al_2O_3 devices in the absence of radiation.

Charge injection into Al_2O_3 can occur from either the gate electrode or from the silicon substrate whenever a sufficient field is applied to draw electrons into the Al_2O_3. Those electrons that remain trapped in the dielectric form a negative space charge causing a positive threshold voltage shift. The injection problem is aggravated by operation in higher temperature ambients. In addition, short high-voltage transients can produce sufficient injection to promote undesirable threshold shifts. These shifts can occur at voltages that are much smaller than those required to induce catastrophic breakdown. Consequently, final Al_2O_3 circuits may require transient protection devices that are more critical than in SiO_2 circuits. In this section of the report the discussion will be limited to the long-term stability effects encountered during normal operation. Since most injected charge is trapped near the injecting (negative) electrode, only the positive gate bias condition will be discussed in detail.

Before discussing the basic charge-injection mechanism it is worthwhile to illustrate the effects that it causes. A typical result of flat-band voltage shift due to charge injection is shown in Fig. 13. The shift in flat-band voltage of an MOS capacitor after 500 seconds of bias application is plotted as a function of bias. For each point on the graph a fresh (uninjected) capacitor was used. It can be seen that for negative bias about twice as much voltage is required to produce a given shift as for positive bias. Most likely this occurs because, for negative bias, most of the trapped charge resides near the gate electrode.

5.1 Charge-Injection Mechanism

Powell has studied the charge-injection and trapping mechanisms using a high-field pulsing technique.[23,24] The results of his study can be summarized as follows:

(1) The electron current injected into the insulator depends only on the field at the injecting electrode, and not on the amount of previously injected charge.

(2) The injection current density has a nearly exponential dependence on interface field.
(3) The amount of flat-band shift depends only on the total charge injected into the sample and not on the rate at which it is injected.
(4) For small amounts of injected charge, Q_I, all of the charge is trapped near the injecting electrode. For large Q_I either the charge is trapped in the bulk of Al_2O_3 and/or the efficiency of trapping near the interface becomes less than one.

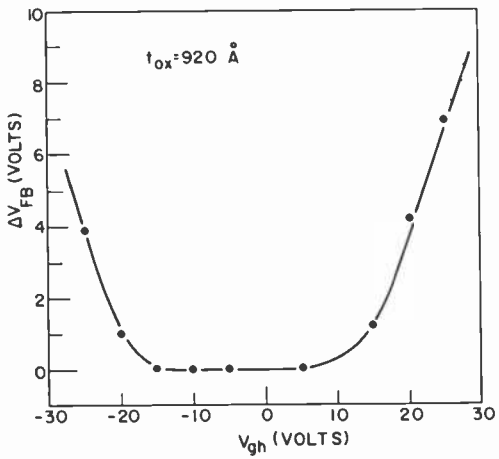


Fig. 13—Flat-band voltage shift induced in Al_2O_3 MOS capacitors by application of bias for 500 seconds.

The nearly exponential current–field relationship leads to a logarithmic time dependence of flat-band or threshold voltage shift with time during constant voltage operation. The injection current density can be empirically written as

$$J_I = J_0 \exp\{\alpha E_i\} = J_0 \exp\left\{\alpha\left(\frac{V - V_{FB}}{L}\right)\right., \qquad [3]$$

where J_0(A/cm^2) and α(V/cm)$^{-1}$ are parameters that may be temperature-dependent. The term $(V - V_{FB})/L$ is the interface field. The effective amount of trapped charge depends on the total amount of injected charge so that

$$Q_T = \int_0^t kJ_I dt, \qquad [4]$$

where k describes the efficiency of trapping and has the value of unity

for small amounts of injection. The trapped charge results in a flat-band voltage given by

$$V_{FB} = V_{FBi} + \frac{L}{\epsilon} Q_T, \qquad [5]$$

where V_{FBi} is the initial flat-band voltage, ϵ is the Al_2O_3 dielectric constant, and all of the charge is trapped very close to the silicon electrode. If charge trapping occurs in the bulk of the insulator, this can be accounted for by a decrease in the parameter k described above. When Eqs. [3], [4], and [5] are combined with a constant applied voltage V, the result is

$$V_{FB} = V_{FBi} + \frac{L}{\alpha} \ln \left(\frac{t}{t_0} + 1 \right), \qquad [6]$$

where

$$\frac{1}{t_0} = \frac{\alpha k}{\epsilon} J_0 \exp \left\{ \frac{\alpha}{L} (V - V_{FBi}) \right\}. \qquad [7]$$

These equations assume that the efficiency of trapping near the interface k is a constant and, therefore, should be valid for short times.

According to Eqs. [6] and [7], the flat-band voltage shift should be characterized by a time, t_0, where shifting begins and a slope (on a log plot) given by L/α. This logarithmic time dependence is encountered during relaxation of radiation-induced charge as described in Sec 4. It is interesting to note that the logarithmic slope depends only on the sample thickness and is smaller for thinner samples. On the other hand, the t_0 point depends primarily on the initial electric field and, at a fixed bias, is shorter for thinner samples.

Fig. 14 shows the time dependence of threshold voltage for various positive biases applied to the gate of an Al_2O_3 n-channel transistor. In agreement with the analysis, the threshold voltage has a logarithmic time dependence and has a t_0 value that decreases as the applied bias increases. However, the logarithmic slope rate shows a slight dependence on bias voltage, being larger for higher bias. This is probably due to inaccuracies in the current field relation of Eq. [3]. The deviations from the logarithmic rate for the large 16-V bias case probably are due to a decreasing value for k after large amounts of injection.

Fig. 15 shows a plot of logarithmic slope as a function of oxide thickness. The three data points for the thin (480-Å) sample correspond to the three samples shown in Fig. 14. In general, there is good agreement between Fig. 15 and the predictions of Eq. [6]; the logarithmic shift rate depends linearly on sample thickness.

The practical application of this result is that improved stability can be achieved by using thinner films of Al_2O_3 and by specifying a lower operating voltage. By reducing both the voltage and thickness proportionally, the operating field would be kept constant so that the time constant t_0 would be unchanged and the injection-limited radiation behavior discussion in Sec. 4 would still apply at the operating voltage. This conclusion is a complete reversal of previous suggestions that improved stability would result from the use of thicker films with a given voltage. The latter idea may effect an increase in the time t_0 from the

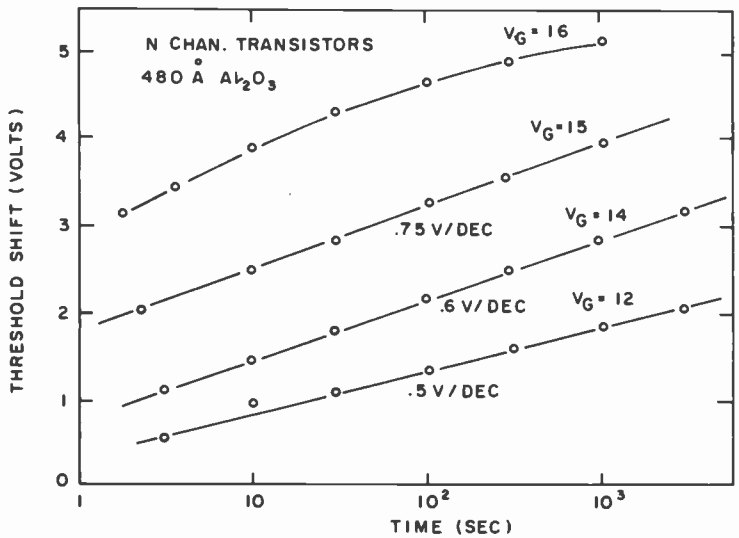


Fig. 14—Charge injection induced threshold shifts as a function of time for various applied biases.

order of seconds to several hundred seconds, but both of these times are short compared with the desired operating life for the devices (several thousand hours). In addition, the use of thinner films results in smaller amounts of radiation-induced threshold shift as discussed in Sec. 4. A drawback to the use of thinner films is that the initial threshold voltages become lower (especially for the n-channel device) so that less radiation-induced shift can be tolerated. Furthermore, a lower operating voltage would restrict the amount of positive n-channel threshold shift that could be tolerated before circuit performance was adversely affected. The stability of the devices during high transient voltages would also

be adversely affected. More study and experience is necessary before an optimum design for a given application can be achieved.

5.2 Life Test on Al_2O_3/CMOS/SOS Devices

Long-term life tests were performed on CMOS/SOS inverters at 125°C. In these tests devices were placed in an oven and an operating voltage of $V_{in} = V_{DD} = +10$ V was applied. Periodically, the samples were removed from the oven for threshold-voltage measurement. In these tests

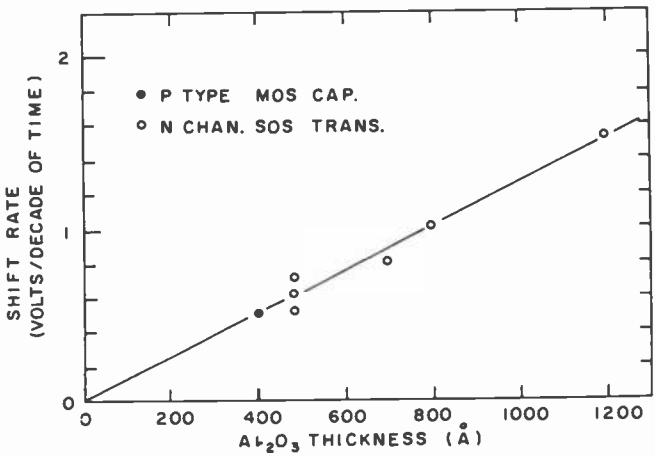


Fig. 15—Logarithmic shift rate due to charge injection as a function of Al_2O_3 thickness. The three values at 480 Å correspond to the rates indicated in Fig. 14.

the bias was removed prior to taking the samples out of the oven, so that some relaxation of effects might have occurred during the measurement procedure.

The results of these tests are shown in Fig. 16 for CMOS/SOS devices with Al_2O_3 thicknesses of 750 and 1200 Å. The threshold shifts are nearly logarithmic in time but there is no large difference in the slope of the two curves as was observed for room-temperature testing. The reason for similar slopes on oxides with widely differing thicknesses is not understood at present; however, it may be due to relaxation of charge between the time that the bias is removed and the samples are cooled to room temperature. On other samples it was observed that high-temperature annealing of injected charge has a very rapid component occurring in less than a minute followed by a much slower component. The amount of flat-band voltage reduction during the first minute depends on the

peak temperature applied to the sample. A more standard testing procedure may be required to resolve the different results obtained with high-temperature moderate-field injection and room-temperature high-field injection.

In a more comprehensive reliability study it was observed that many p-channel devices operated with a gate-to-source voltage of −10 V had enough charge injection to shift the threshold into depletion. Even though threshold shifts are less for negative gate voltage, this represents a serious failure mechanism since the standby power is dramatically increased.

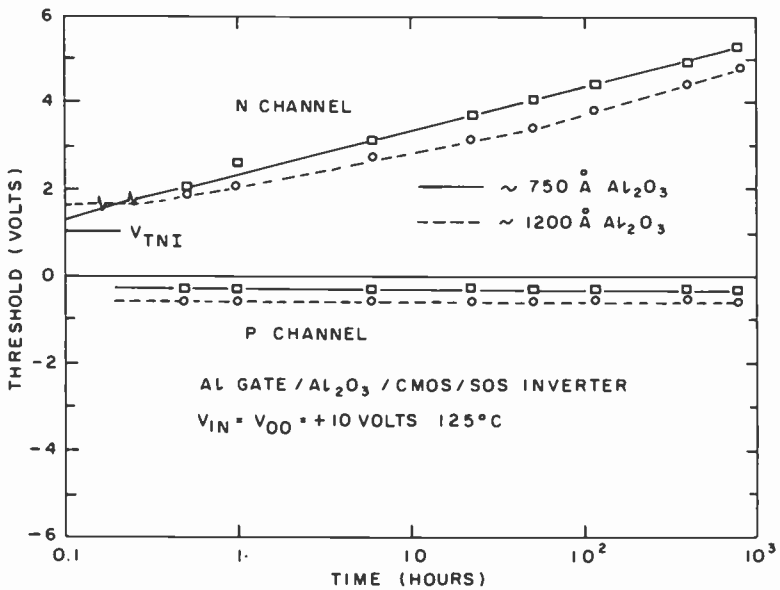


Fig. 16—Threshold voltage for CMOS inverters as a function of time that the devices were biased at $V_{in} = V_{DD} = +10$ V at 125°C. For each point on the curve the bias was removed and the samples cooled to room temperature prior to measurement. Relaxation of charge during the cool-down cycle may account for the similar slope for both 750-Å and 1200-Å samples.

5.3 Conclusions on Charge Injection

Charge injection plays an important role in Al_2O_3 devices and appears to enhance radiation hardness under certain bias conditions; however, the trapping of injected electrons leads to a threshold voltage instability in the absence of radiation. Investigations performed to date have not

resulted in Al_2O_3 films with improved stability; i.e., all Al_2O_3 films seem to have the same injection properties independent of the method of fabrication. However, a better understanding of the charge-injection mechanism suggests that improved device stability may be achieved with thinner films and lower operating voltages. In addition, the studies have brought us closer to the goal of predicting Al_2O_3 device behavior in particular environment. The ability to predict the device behavior may be a valid approach to reliability as opposed to the achievement of absolute stability. Since radiation-hard circuit design must be able to tolerate radiation-induced threshold changes, it should also be able to withstand threshold variations due to charge injection.

6. Conclusions

The impetus behind Al_2O_3 gate insulator development is to achieve a CMOS circuit technology that is tolerant to high levels of ionizing radiation. Although this is not the only use of Al_2O_3 in integrated circuits, it is the sole purpose of the developments described here. In this section we summarize the properties of Al_2O_3 with respect to its usefulness as a CMOS gate insulator, and place the Al_2O_3 technology in perspective compared with other techniques for hardening CMOS circuits.

6.1 Al_2O_3 as a Radiation Hard CMOS Gate Insulator

Presently employed commercial CMOS gate insulator technology is not sufficiently radiation tolerant for some military and space requirements. It is clear that a change in insulator technology is required to achieve the desired goals. In making this change the improved gate insulator technology should have the following characteristics:

(1) Radiation Tolerance With Bias from −15 to +15 Volts

Al_2O_3 is radiation tolerant under zero and negative bias, but only under a sufficient positive bias to produce some charge injection during irradiation. Results indicate that by making thinner Al_2O_3 films, hardness could be achieved over the range −5 volts to +5 volts.

(2) Manufacturability

The Al_2O_3 technology uses a deposited rather than thermally grown process. Improvements in insulator deposition are continuously being made, and it is expected that such a process could be implemented in a production line once the proper techniques and controls were established.

(3) *Minimum Effect on Device Characteristics*

Al_2O_3 has a higher dielectric constant and gives lower carrier mobility than thermally grown SiO_2 resulting in reduced on-chip speed. In addition, there is a negative interface charge which results in low p-channel thresholds. The major importance of these changes, if any, is not clear.

(4) *Minimum Impact on Device Yield*

The present status of Al_2O_3 deposition has not resulted in high yields of devices. As techniques and controls for deposition are developed, it is expected that improved yields will occur.

(5) *Meet Reliability Specifications*

Charge-injection instabilities cause major threshold shifts during elevated temperature operation. It is clear that charge injection is desirable to achieve radiation hardness in Al_2O_3 gate insulators. It is also apparent that any radiation-tolerant CMOS circuit design must be able to tolerate some threshold shift due to radiation damage. Therefore, a shift during non-radiation operation may be tolerable also. However, the present formulation of reliability standards (e.g., Mil-M-38510) equates reliability with stability. Requirements on stability would have to be relaxed in order for Al_2O_3/CMOS circuits to be considered reliable. Threshold shifts and circuit changes due to charge injection in Al_2O_3 are predictable. Al_2O_3 circuits could be considered reliable if reliability concepts were changed to equate reliability with predictability rather than with stability.

In summary Al_2O_3 does not fully meet requirements 1 and 5 for hardness and reliability. Improvements in manufacturability and yield would be expected with further development of the Al_2O_3 process.

6.2 Comparison of Radiation Hardening with Al_2O_3 and SiO_2 Gate Insulators

At one time it was believed that thermally grown SiO_2 was inherently sensitive to ionizing radiation. It has since been clearly demonstrated that this is not true. The radiation tolerance of thermally grown SiO_2 is strongly related to the detailed process used during manufacture. There appears to be more than one way to form radiation-tolerant SiO_2, and there are several detailed problems associated with each of these

approaches involving both manufacturability and the parameters of the final circuit. It is not our purpose to discuss these problems; instead we cover the more general features of the comparison of Al_2O_3 and SiO_2 dielectrics for radiation hardening.

It appears that hardness of SiO_2 is directly related to the nature of the Si/SiO_2 interface. When this interface is modified by changing the growth process, large differences in radiation tolerance can be observed without major changes in any other electrical parameter such as field-effect mobility, charge concentration, or stability. Radiation hardness in Al_2O_3 is in part related to its charge-injection properties. As a result, the Al_2O_3 hardness is more dependent on bias voltage, while hardened SiO_2 is less sensitive to bias. Due to charge injection, Al_2O_3 is inherently unstable, whereas hard SiO_2 appears to have the same stability that is associated with unhardened SiO_2. Because of these features, the hard SiO_2 approach is much closer to achieving the requirements 1 and 5 in Sec. 6.1 for hardness and reliability.

The impact of the hard SiO_2 approach on manufacturability, device parameters, and yield has been established for a product capable of 10^5 rads (Si). Megarad hard CMOS is currently being evaluated under these same conditions. For these reasons most people associated with CMOS manufacture are more likely to favor a modification of the SiO_2 growth as opposed to a substitution of a deposited insulator for the SiO_2.

An undesirable feature of hard SiO_2 is that radiation damage is cumulative and irreversible. An SiO_2 device that can tolerate 10^6 rads (Si) will not necessarily tolerate 10^7 rads (Si). With Al_2O_3 a steady state with respect to irradiation is achieved after 10^6 rads (Si). Further irradiation does not cause further degradation in Al_2O_3 up to at least 10^8 rads (Si). Most present-day radiation-hardness requirements call for a tolerance of a maximum of 10^6 rads (Si), which appears to be within the capabilities of hard SiO_2.

Based on these considerations it is apparent that hard SiO_2 is more likely to succeed in meeting the megarad requirements for radiation-hard CMOS circuits. The use of Al_2O_3 for radiation hardening is dependent on requirements for a technology that can tolerate extremely high doses of radiation (10^8 rads (Si) or more), which cannot be satisfied with SiO_2 gate insulators.

References:

[1] K. G. Aubuchon and J. Bereisa, "Performance of Hardened P-MOS Devices in Severe Neutron Environments," *IEEE Trans. Nuc., Sci.*, **NS-19**, p. 299, Dec. 1972.

[2] K. H. Zaininger, "Electron Bombardment of MOS Capacitors," *Appl. Phys. Lett.*, **8**, p. 140 (1966).

[3] R. J. Powell and G. F. Derbenwick, "Vacuum Ultraviolet Radiation Effects in SiO_2," *IEEE Trans. Nucl. Sci.*, **NS-18**, p. 99 (1971).

[4] O. L. Curtis, J. R. Srour, and K. Y. Chiu, "Hole and Electron Transport in SiO_2 Films, *J. Appl. Phys.*, **45**, p. 4506 (1974).

[5] R. J. Powell, "Radiation Induced Hole Transport and Electron Tunnel Injection in SiO_2 Films," *IEEE Trans. Nucl. Sci.*, **NS-22**, p. 2240, Dec. 1975.

[6] H. E. Boesch, Jr., F. B. McLean, J. M. McGarrity, and G. A. Ausman, Jr., "Hole Transport and Charge Relaxation in Irradiated SiO_2 MOS Capacitors," *IEEE Trans. Nucl. Sci.*, **NS-22**, p. 2163 (1975).

[7] R. C. Hughes, E. P. Eer Nisse, and H. J. Stern, "Hole Transport in MOS Oxides," *IEEE Trans. Nucl. Sci.*, **NS-22**, p. 2227 (1975).

[8] R. A. Kjar, J. L. Peel, and C. Y. Wrigley, "Effects of Metallic Doping on Ionization Damage in MOSFET's," *IEEE Trans. Nucl. Sci.*, **NS-16**, p. 207 (1969).

[9] K. G. Aubuchon, "Radiation Hardening of PMOS Devices by Optimization of the Thermal SiO_2 Gate Insulator," *IEEE Trans. Nucl. Sci.*, **NS-18**, p. 117 (1971).

[10] J. L. Peel and G. Kinoshita, "Radiation-Hardened Complementary MOS Using SiO_2 Gate Insulators," *IEEE Trans. Nucl. Sci.*, **NS-19**, p. 271 (1972).

[11] K. Schlesier and P. E. Norris, "CMOS Hardening Techniques," *IEEE Trans. Nucl. Sci.*, **NS-19**, p. 275 (1972).

[12] K. M. Schlesier, "Radiation Hardening of CMOS/SOS integrated Circuits," *IEEE Trans. Nucl. Sci.*, **NS-21**, p. 152 (1974).

[13] G. F. Derbenwick and B. L. Gregory, "Process Optimization of Radiation Hardened CMOS Integrated Circuits," *IEEE Trans. Nucl. Sci.*, **NS-22**, p. 2151 (1975).

[14] K. G. Aubuchon and E. Harrari, "Radiation Hardened CMOS/SOS," *IEEE Trans. Nucl. Sci.*, **NS-22**, p. 2181 (1975).

[15] K. H. Zaininger and A. S. Waxman, "Radiation Resistance of Al_2O_3 MOS Devices," *IEEE Trans. Electron Dev.*, **ED-16**, p. 333 (1969).

[16] M. T. Duffy and W. Kern, "Chemical Vapor Deposition of Aluminum Oxide Films from Organo-Aluminum Compounds," *RCA Review*, **31**, p. 754 (1970).

[17] S. K. Tung and R. E. Caffrey, "The Deposition of Oxide on Silicon by the Reaction of a Metal Halide with a Hydrogen-Carbon Dioxide Mixture," *Trans. Met. Soc. AIME*, **233**, p. 572 (1968).

[18] D. H. Phillips, "Silicon-on-Sapphire Device Photocurrent Predictions," *IEEE Trans. Nucl. Sci.*, **NS-21**, p. 217 (1974).

[19] R. A. Kjar and J. L. Peel, "Radiation Induced Leakage Current in N-Channel SOS Transistors," *IEEE Trans. Nucl. Sci.*, **NS-21**, p. 208 (1974).

[20] D. Neaman, B. Buchanan, and W. Shedd, "Ionizing Radiation Effects in SOS Structures," *IEEE Trans. Nucl. Sci.*, **NS-22**, p. 2197 (1975).

[21] Ç. F. Powell, J. H. Oxley, and J. M. Blocher, Jr., *Vapor Deposition*, John Wiley and Sons, Inc., New York (1966A).

[22] M. J. Markulec, B. R. Parson, and J. M. Shaw, "Design of the Computation and Processing Unit for the Survivable Metal Oxide Semiconductor Array Computer," Vol. I, AF Contract F33615-73-C-1100.

[23] R. J. Powell, "Radiation and Charge Transport Phenomena in Improved Hardened Insulators," Final Report Contract N00014-73-C-0277, Office of Naval Research (Jan. 1974).

[24] R. J. Powell and G. W. Hughes, "Charge Injection and Trapping in Al_2O_3 Gate Insulators," *IEEE Trans. Nucl. Sci.*, **NS-21**, p. 179 (Dec. 1974).

Microsonic Pulse Filters—Replacements for Traditional Butterworth Designs

J. H. McCusker, S. S. Perlman, and H. S. Veloric

RCA Laboratories, Princeton, N. J. 08540

Abstract—Microsonic filter response characteristics are analyzed under cw, pulsed, and transient input conditions. Previous cw formulations had to be extended for pulsed and transient response calculations because the acoustic waves generated are not, in general, spatially long enough to encompass the entire transducer pattern. A synthesis procedure is also developed and used for the successful design of six pulse filters employed in a practical airborne collision-avoidance system, SECANT VECAS GM/A. The microsonic filters replace six standard lumped-element five-pole Butterworth filters required to distinguish various 1-μs pulses received from nearby aircraft. Each microsonic filter (on ST-cut quartz) comprises one uniform interdigital transducer with 35 to 39 lambda/4 fingers and one Gaussian apodized transducer with 221 to 257 fingers, depending on the synchronous frequencies (53 to 64 MHz) that match the frequencies contained in the different i-f input pulse envelopes.

1. Introduction

Analysis of the response characteristics of a standard two-transducer configured microsonic filter is presented here. Equations are developed for calculating the pulse and transient response characteristics of the devices, and a synthesis procedure is applied to the design of a set of six filters needed to replace a corresponding set of five-pole Butterworth filter designs. Fabrication procedures and measurements on the filters are described.

The microsonic pulse filters have been incorporated in prototype hardware for SECANT VECAS, an aircraft anti-collision system under

development since 1969.[1,2] The acronym SECANT is derived from "Separation and Control of Aircraft by Non-Synchronous Techniques," while VECAS is from "Vertical Escape Collision Avoidance System."

SECANT is a cooperative system requiring electronic equipment on each aircraft in flight but not requiring any ground-support equipment. Each aircraft is capable of receiving, responding, and interrogating with pulse signals. The SECANT equipment distinguishes these pulse signals and calculates appropriate vertical escape maneuvers when necessary. The microsonic pulse filters were designed for this purpose and the filters were employed in the third generation prototypes of hardware, SECANT VECAS GA/M (General Aviation/Military), which is smaller, lighter and less expensive than the previous generations.[2] Excellent in-flight test results, conducted by the Naval Air Development Center, Warminster, Pennsylvania, have proven that SECANT is a viable system for the prevention of mid-air collisions.[2]

2. Filter Specifications

Each aircraft in the SECANT system contains both an interrogator and a transponder. It must periodically send out pulses ("probing" or interrogating) and receive replies from other nearby aircraft. Also, it must listen for probes from other aircraft and send back replies ("remitting" or transponding).[1]

The SECANT system operates at L-band (1.595 to 1.620 GHz), with 1-μs probes, and replies on 24 different frequencies (1 MHz apart) with pulse repetition rates up to 1000/sec. Each SECANT equipped aircraft selects a set of 12 frequencies depending on its altitude (low band $<$ 10,000 ft. or high band $>$ 10,000 ft.). The 12 frequencies selected are divided equally into two antenna fields for the top and bottom of the aircraft. Each antenna field is composed of two probe frequencies (P and Q) and a corresponding set of four reply frequencies (P^+, P^- and Q^+, Q^-). Probe frequencies are randomly selected between P and Q in order to establish a correlation procedure for distinguishing between desired and undesired replies.[1]

Signal processing in the SECANT receiver starts with a mixer generating a common i-f between 54 to 64 MHz for each of the antenna fields. Pulses pass through a broadband limiter to reduce the dynamic range from 60 to 9 dB without introducing significant distortion into the essentially rectangular 1-μs pulse envelopes. A bank of six microsonic filters separates the input pulses into channels (P, P^+, P^-, Q, Q^+, Q^-) depending on the frequency contained within the envelope. At i-f, the six channels are centered at 55 (P), 56 (Q), 59 (P^+), 60 (P^-), 63 (Q^-), and

64 (Q^+) MHz. The response of each channel filter has to match the following list of specifications for the input of the 1-μs probes and replies.

(1) Resonant pulse response shall not be significantly stretched. Maximum 0.5 amplitude (−6 dB from peak response) envelope pulse width shall not exceed 1.35 μs, while the maximum 0.22 amplitude (−13 dB from peak response) shall not exceed 2.0 μs.
(2) Adjacent channel response (at ± 1 MHz intervals) shall not exceed 0.25 (−12 dB from peak resonant response).
(3) Resonant pulse response delay, from leading edge of the input pulse at half amplitude to leading edge of output pulse at half amplitude, shall be 1.80 ± 0.05 μs.
(4) Continuous-wave adjacent channel and out-of-band response shall be less than 0.03 (−30 dB from peak resonant response).
(5) Satisfactory performance is required over a temperature range from 0 to 80°C.

In the first two generations of SECANT hardware, the specifications were reasonably matched by employing five-pole Butterworth lumped-element filter designs.[1] Maintaining the equal time delay ((3) above) for all the channels was one of the most difficult design problems to overcome. Compensation for delay variations in the filter designs was required. Adjustable video delay lines had to be inserted in each channel. Specification (2) relating to adjacent channel response was only marginally matched. Furthermore, the Butterworth filters were nonreproducible, bulky, and expensive. These disadvantages combined to make the filters unacceptable for the third-generation hardware, SECANT VECAS GM/A.[2] Microsonic filters, on the other hand, were found to be much more acceptable in response characteristics, reproducibility, size, and cost.

3. Microsonic Filter Design

Microsonic bandpass filters are currently under investigation in various laboratories. A significant effort has been aimed at the design of bandpass (continuous wave) filters. However, the synthesis routines that have been developed for such filters are not adequate for designing pulse-response filters, mainly because the input signal may be too short to generate an acoustic wave spatially long enough to encompass the entire length of the output transducer. Consequently, the response of the transducer to such a pulse has to be calculated as the sum of the responses of the finger pattern overlaps that are actually activated as the

acoustic pulse travels through the transducer. The same is true for the evaluation of the transient response of a microsonic bandpass filter. Effects of rapid changes in the frequency and/or amplitude content of the input signals cannot be calculated by standard cw formulations.

The pulse response, or time response, of a specific filter pattern can be determined with the aid of Fourier Transforms [3]

$$f(t)_{response} = F^{-1}\,(f(f)_{pulse} \times f(f)_{sender} \times f(f)_{receiver}). \qquad [1]$$

In principle, if any combination of three of these variables can be specified, then the remaining variable can be evaluated. For the SECANT filter, $f(t)_{response}$, $f(f)_{pulse}$, and $f(f)_{sender}$ can be specified provided the following is required: (a) only specification (1) regarding output pulse shape need be satisfied, (b) only responses for input envelopes containing the synchronous frequency modulation need be considered, and (c) only one pre-selected uniform sending (input) transducer is used. Under these conditions, $f(f)_{receiver}$ can be evaluated and its inverse transform can be used to specify the apodization function for the finger overlap pattern of the receiver transducer. Unfortunately, the remaining specifications also have to be satisfied. For example, specification (2) (adjacent channel response) may not be simultaneously satisfied with the previous conditions because $f(f)_{pulse}$ in Eq. [1] is now different for the different frequency input pulses (corresponding to the various adjacent channel frequencies), and the response maximum will probably exceed the allowable −12 dB. In general, specification (2) can be achieved only at the expense of specification (1) and a compromise has to be established. The adjacent-channel response maximum can be lowered by stretching out the response time. This can be accomplished by changing the overlap pattern and increasing the time length of the transducer. The resonant response, however, is also stretched by this change, and specification (1) may be violated. Furthermore, specification (4) (cw response) may play a role, because the cw response is also determined by the transducer patterns. The cw response $f(f)_{response}$ can be evaluated by taking the inverse transform of Eq. [1] and eliminating $f(f)_{pulse}$ by setting it equal to a constant, which is the case when the reponse is calculated during the time in the center of a very long input pulse. Specification (4) cannot be satisfied if the receiving transducer is much less than 2 μs, because the trap bandwidth and out-of-band rejection is generally inversely proportional to the transducer length.

The above discussion clearly indicates the necessity of a trial and error synthesis technique where various tradeoffs in the specifications have to be evaluated. Our procedure was to select a pattern for the receiver transducer and evaluate Eq. [1] for the various responses corresponding

to specifications (1), (2), and (4). The best compromise filter design was then selected.

A computer program had to be written. Mathematically, the program calculates Eq. [1] completely in the time domain by evaluating the time convolution integrals for input pulse and the sender transducer, and the result is convoluted with the receiver transducer.[3]

$$f(t)_{response} = (f(t)_{pulse} * f(t)_{sender}) * f(t)_{receiver} \qquad [2]$$

or

$$v(t) = \Bigg[\exp(+i\omega_x t) \times \Bigg(\sum_{N_S=0}^{N_{S(max)}} \sum_{N_{PS}=0}^{N_{PS(max)}} \mathrm{OLAP}_S(N_S) \exp(-i\omega_x N_S/2F_0)\, \delta(N_T - N_{PS} - N_S) \Bigg) \times \Bigg(\sum_{N_R=0}^{N_{R(max)}} \sum_{N_{PR}=0}^{N_{PR(max)}} \mathrm{OLAP}_R(N_R) \exp(-i\omega_x N_R/2F_0)\, \delta(N_T - N_{PR} - N_R) \Bigg) \Bigg] \qquad [2a]$$

where

$t = [N_T + (2L/\lambda_0)]\ 1/2F_0$ = digitized time,
L = center-to-center distance between the transducers,
F_0 = synchronous frequency of the transducer patterns,
λ_0 = synchronous frequency wavelength,
N_T = running integer to digitized time,
$N_{S(max)}$ = NFS-1 = number of sender fingers-1,
$N_{R(max)}$ = NFR-1 = number of receiver fingers-1,
$\mathrm{OLAP}_S(N_S) = (-1)^{N_S}$ = overlap values for the uniform sender,
$\mathrm{OLAP}_R(N_R)$ = arbitrary overlap weighting function for the receiver,
$\omega_x = 2\pi F_x$ for the frequency F_x contained in the pulse envelope,
T_x = time of rectangular input pulse,
$N_{PS(max)} = 2F_0 T_x$ = digitizing integer for pulse width at the sender, and
$N_{PR(max)} = N_{PS}$ + (NFS-1) = digitizing integer for expanded pulse width at the receiver.

The time response in Eq. [2a] was digitized into integer (N_T) multiples of time $1/(2F_0)$ corresponding to a travel distance of the acoustic wave

equal to $\lambda_0/2$, which is the distance between adjacent overlaps in the sender and receiver patterns assuming that both have the same F_0.

A Gaussian apodization function was found to yield an acceptable compromise for the filter specifications:

$$\mathrm{OLAP}_R(N_R) = (-1)^{N_R} \exp[-A((\mathrm{NFR}\text{-}1)/2 - N_R)^2] \qquad [3]$$

where $N_R = 1$ is the first overlap value nearest the sender and N_R = NFR-1 is the last overlap value. The value of A was chosen to make OLAP_R (1) = 0.1 times the tallest overlap, which is taken as unity. In general, a larger OLAP_R (1) improves the adjacent channel rejection but also stretches the resonant pulse width.

The sending transducer should have as few fingers as possible in order to minimize the stretch of the input pulse. Radiation impedance of the device, however, is inversely proportional to the square of the transducer length and a compromise has to be established again. Specification (5) relating to temperature effects requires the use of ST-cut quartz as a device substrate because it has the best temperature coefficient behavior, leading to an essentially zero coefficient for the resonant frequency of the transducers. Unfortunately, the radiation impedance of a uniform transducer on quartz is very high unless the number of fingers is at least 30. The compromise was to select the time length of the input transducer at about 0.31 μs. This corresponds to uniform transducers with 35 fingers for the 55 and 56 MHz filters, 37 for the 59 and 60 MHz filters, and 39 for the remaining two filters. The acoustic waves generated by these transducers ($f(t)_{input} * f(t)_{sender}$) are stretched by the time length of the transducer so that the pulse length, from the start to finish, is about 1.31 μs for each SECANT filter instead of the 1 μs specified for the electrical input pulse. This corresponds to an acoustic pulse entering the receiver with a half amplitude width of about 1.15 μs.

The choice of the Gaussian apodization function for the output transducer was determined through trial and error techniques. In order to match specification (2), the sum of the adjacent channel responses due to the overlaps covered by the acoustic pulse as it penetrates the transducer has to alternately increase and decrease such that the peak amplitude remains low. This requires that the apodization function build up gradually. An endless variety of apodization functions matching this condition can be employed. Each design will result in slightly different responses which will compromise the specifications. Table 1 demonstrates this effect for the design of the 55 MHz filter employing a uniform sender with 35 fingers and the various apodized receivers listed. Filter (1) represents the Gaussian distribution actually selected for the SECANT design. Specifications (1), (2), and (4) (relating to shape, adjacent

channel response, and cw) are satisfied. Specifications (3) and (5) (delay time and temperature) will be equally satisfied by all of the filter designs. Improvement in specification (2) can be obtained, but only at the expense of specification (1). In general, the longer the filter, the better specification (2) will be matched. This is demonstrated by the next two filters in the table, which employ stair-step apodization functions of 2- and 3-μs lengths. The adjacent channel response improves from -15.6 to -19.2 dB, but the -6 dB width of the output pulse increases from 1.48 to 2.05 μs. A compromise stair-step design leading to acceptable specifications (1) and (2) responses is demonstrated by filter (4) which is 1.6 μs long. The problem with this filter, however, is its relatively large cw responses for signals outside the $\pm$ 1 MHz band. The last filter shown differs from all the others in that its design is based on a cw procedure. The apodization is a Hammond weighted cosine function. In this case, the adjacent channel pulse response is too large.

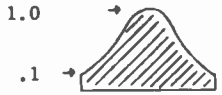
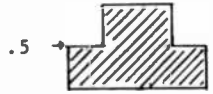
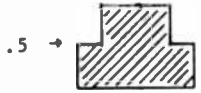
Another filter configuration worth mentioning is one comprising only uniform transducers. The best choice for the SECANT filters is for both the sender and receiver transducers to be the same and 1 μs long (NFS = NFR = $2F_0(1\ \mu s) + 1$). The pulse response of these filters will satisfy the specifications (-6 dB width = 1.3 μs, -13 dB width = 1.85 μs, and adjacent channel response = -13.5 dB) but, unfortunately, the cw out-of-band response cannot be matched (maximum response is about -27 dB) even though the adjacent channel cw responses are zero.

The Gaussian function, filter (1) in Table 1, obviously represents the best compromise to the filter specifications. Each of the SECANT filters were designed with this function, as expressed by Eq. [3]. The time length of the filters are all set at approximately 2 μs which corresponds to 221 for the 55 MHz, 225 for the 56 MHz, 237 for the 59 MHz, 241 for the 60 MHz, 253 for the 63 MHz, and 257 for the 64 MHz filter.

Matching the time delay specification, specification (3), for a microsonic filter is relatively easy. The time delay is directly related to the center-to-center distance between the sending and receiving transducers. A delay of 1.75 μs was selected for the time as measured at the half amplitude points of the pulse. The delay was intentionally selected short by 50 ns in order to accommodate an expected delay, which will be present in the input and output series tuned circuits. A center-to-center distance of 1.9 μs for the transducers yielded the desired delay.

The impedance of the transducers is controlled, of course, by the finger pattern and the aperture of the transducer. Since the pattern is defined, only the aperture can be varied. A 200 mil (0.5 cm) aperture was selected, which results in series equivalent radiation resistance and capacitance of about 16 ohms and 4.5 pF for the uniform and 20 ohms and 16 pF for the apodized transducers. An intentional mismatch, to the 50 ohm

Table 1—Microsonic Filter Responses Calculated for Different Receiver Patterns

Filter No.	Apodization *	Receiver Length		Pulse Response			CW Response	
		NFR	μs	Width at −6 dB (μs)	Width at −13 dB (μs)	at ±1 MHz (dB)	at ±1 MHz (dB)	Maximum (dB)
1		221	2	1.3	1.85	−13.4	<−50	−36.5
2		221	2	1.48	2.2	−15.6	<−50	−23.2
3		331	3	2.05	2.9	−19.2	<−50	−30.2
4		177	1.6	1.3	1.9	−14.0	<−50	−21.6
5		187	1.7	1.15	1.6	−11.5	<−50	-45

* Sketch is for receiver transducer. All filters are 55 MHz with uniform sender, NFS = 35.

generator and load impedance, is employed as a means to suppress triple transit responses characteristic of microsonic filters. Power insertion loss of the filters at synchronous frequency in a series tuned input and output configuration is expected to be about 17 dB (5 + 7 + 5 dB for the input circuit, transducer, and output circuit losses, respectively) for cw operation and about 19 dB for operation with the specified synchronous input pulse (the transducer loss increases to about 9 dB because the acoustic pulse is too short to ever completely excite the receiver pattern).

The overall dimensions of the complete filter pattern are approximately equal for all of the SECANT filters. The time length is about 1.9 μs center-to-center distance plus half the sum of the transducer lengths:

$$\text{Time Length} \cong 1.9 + (0.31 + 2.0)/2 = 3.15\ \mu s,$$

which equals about 400 mils (1 cm) for the length and approximately 300 mils (0.75 cm) for the height, 200 mils (0.5 cm) of aperture, and 50 mils (1.25 mm) each for the contacts. The quartz crystal size required, however, is about 600 mils (1.5 cm) long by 400 mils (1 cm) wide. The extra length is used for an acoustical absorber material to help eliminate the possibility of reflections from the crystal ends and the extra height for spare room.

4. Microsonic Filter Fabrication

Microsonic filters for SECANT employ interdigital patterns involving about 275 fingers of aluminum approximately 13 μm wide and 0.15 μm thick on 13 μm centers. Fabrication requires a series of specific operations that had to be developed for the polished (3-μm finish) ST-cut quartz substrates employed.* These included substrate preparation, aluminum metalization, photoresist processing, chemical etching, mounting, and wire bonding.

Substrate preparation is important, since the presence of any residue dust on the substrate surface will degrade the line definition. Fortunately, quartz is not ferroelectric and has a low dielectric constant, which reduces the chances for electrostatic attraction of particles. In addition, the material is relatively inert and can be cleaned in acids. Satisfactory preparation is achieved by placing the wafers in a teflon fixture and boiling them in J-100 resist strip† at 125°C. Sequential rinsing in methanol, acetone, and propanol completed the cleaning process.

* Valpey Fisher Corporation, Hopkinton, Massachusetts.

† IRCL Laboratories, Richardson, Texas.

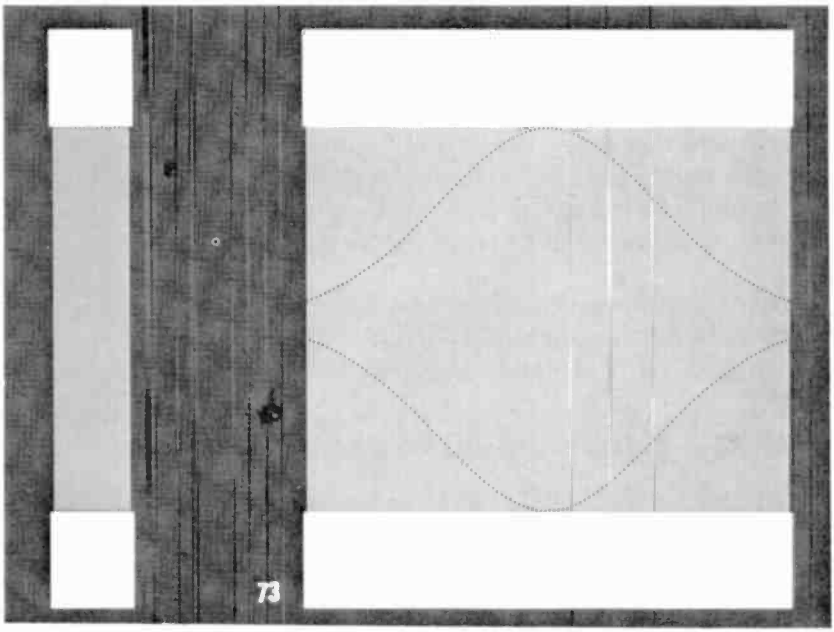


Fig. 1—Photomicrograph of the transducer patterns for the 55-MHz SECANT filter.

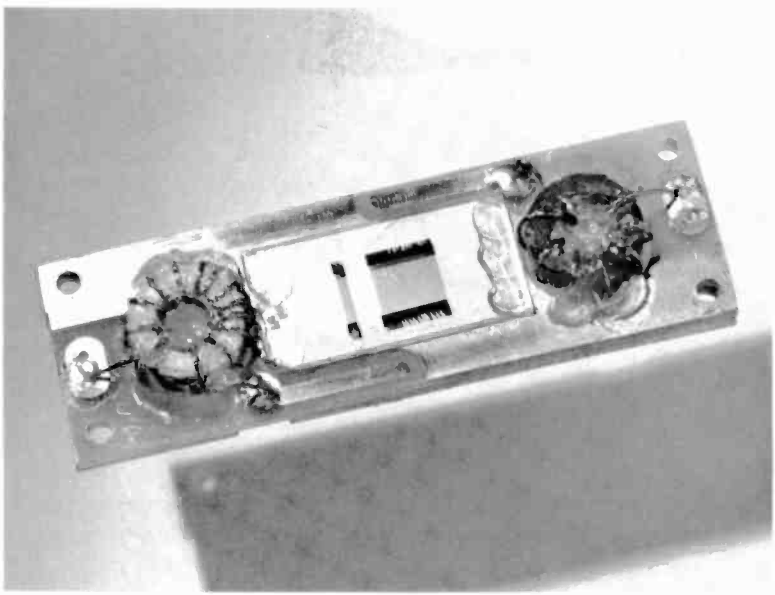


Fig. 2—Complete SECANT filter assembly including series tuned input and output coils.

Metalization is accomplished by a slow electron-gun evaporation of aluminum from a molten pool contained in a carbon boat. Substrate temperature is maintained at 150°C during the evaporation in a vacuum of less than 3×10^{-6} torr. A film thickness of 0.15 μm is deposited in about 1–2 minutes.

Shipley 1350B diluted 80/20 with thinner is used as the photoresist material. The mixture is filtered to remove any trace of solid impurities and dispensed directly from a syringe to a 5000 rpm spinning substrate to obtain a thickness of about 1 μm. A 75°C bake follows for about 15 minutes. An emulsion mask of the device pattern is set in place with a Kasper mask aligner and exposed for about 7 seconds with a Preco Hg-arc light source. The pattern is then developed with Shipley developer diluted 1:1 with H_2O. Optimum reproduction is achieved by beaker developing for 15 seconds, followed by spray developing until the entire pattern is clear. The wafer is then immersed in DI water and baked at 75°C for 15 minutes.

Beaker etching at 40°C with a mixture of HNO_3, H_3PO_4, and acetic acid is used to define the pattern. Chemical etching, however, often leaves shorts in the pattern due to masking of H_2 bubbles, which adhere to the Al in the viscous etch. This is alleviated by numerous short etches and rinsing to remove the trapped gas. The finished wafer is carefully water-rinsed to remove any traces of acid. The photoresist is removed with acetone and the wafers inspected for electrical shorts.

A photomicrograph of the pattern for the 55-MHz SECANT filter is shown in Fig. 1. A standard lambda/4 device with dummy fingers is used. Each finger and spacing is 14.35 μm. The other higher frequency filters are similar except that the number of fingers increases and the wavelength decreases.

The filters are individually cemented with General Electric RTV108 to a precut epoxy-glass board equipped with gold-flashed circuit paths and recesses for the tuning inductors. Aluminum 25-μm wire is ultrasonically bonded to the pads and contact areas of the device. About five wires are attached to the contacts of the uniform transducer and 16 to those of the weighted transducer. Bonding wires are kept as short (about 0.5 cm) as possible. Toroidal inductors with powered iron cores (Micrometals T44-6) are set within the recess areas and connected in series with the transducers. Each was trimmed by turn-spreading to obtain zero phase angle for the impedance of the circuit at the center frequency of the filter. The RTV108 is applied to cement the inductors in place and to secure the inductor setting. An operating Q at center frequency of approximately 140 was obtained. The cement is also used as the acoustic absorber material at the ends of substrate. A photograph of a completed filter assembly is shown in Fig. 2.

Table 2—Data for Set of Microsonic Filters Employed in SECANT VECAS GA/M Prototype

Filter Frequency (MHz)	Delay (μs)	Resonant Response*		Adjacent Response*		CW Response	
		Width at −6 dB (μs)	Width at −13 dB (μs)	at −1 MHz (dB)	at +1 MHz (dB)	at ±1 MHz (dB)	Maximum (dB)
55	1.76	1.35	1.9	−12.5	−12.6	<−40	<−35
55	1.76	1.35	1.9	−12.6	−12.6	<−40	<−35
56	1.76	1.35	1.9	−11.9	−13.0	<−40	<−35
56	1.74	1.35	1.9	−12.0	−13.1	<−40	<−35
59	1.76	1.35	1.9	−11.9	−12.8	<−40	<−35
60	1.72	1.35	1.9	−11.3	−12.5	<−40	<−35
63	1.72	1.35	1.9	−11.5	−12.7	<−40	<−35
64	1.73	1.35	1.9	−11.6	−12.8	<−40	<−35

*Data for 1 μs pulse envelopes with 50 ns rise and fall times.

5. Microsonic Filter Performance

A complete SECANT detector module requires eight filter assemblies, two each at 55 and 56 MHz and one each at 59, 60, 63, and 64 MHz. The extra two filters are used in the two different antenna fields for the top and the bottom of the aircraft. The outputs of the filter assemblies are connected directly to a 50-ohm detector IC with a maximum wire length

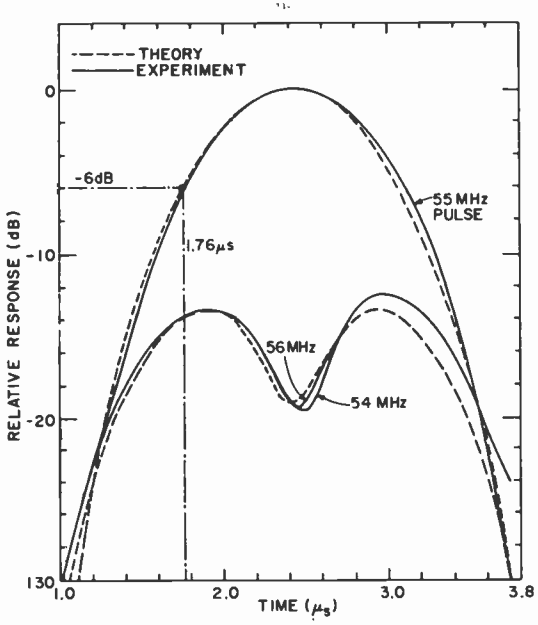


Fig. 3—Resonant and adjacent channel response characteristics for the 55-MHz SECANT filter.

of about 1.5 cm. The input of the filter assembly is paired with another assembly by using two quarter-wavelength RG178B coax cables connected in parallel at the input signal source.[2]

Each filter assembly was measured and recorded photographically using an attenuation-substitution technique. A Tektronix 454 oscilloscope was used at a constant gain setting of 5 mv/div. The performance of the microsonic filters came close to the theoretical predictions. Table 2 gives response data for the six different frequency filters. A comparison with the filter specification list indicates that all of the filters matched the specifications, with the possible exception of 12-dB adjacent-channel rejection. This specification had to be compromised with the specification relating to the ½ width spread of the pulse. Figs. 3 and 4 show the

experimental pulse response data for the 55 and 64 MHz filters compared to the theoretical predictions calculated using Eqs. [2] and [3]. The match is reasonably close except towards the end of the pulse where higher than anticipated responses are obtained. These responses are due to internal reflections of the acoustic pulse as it passes through the transducer. The effect can be eliminated by changing the design to a lambda/8 pattern

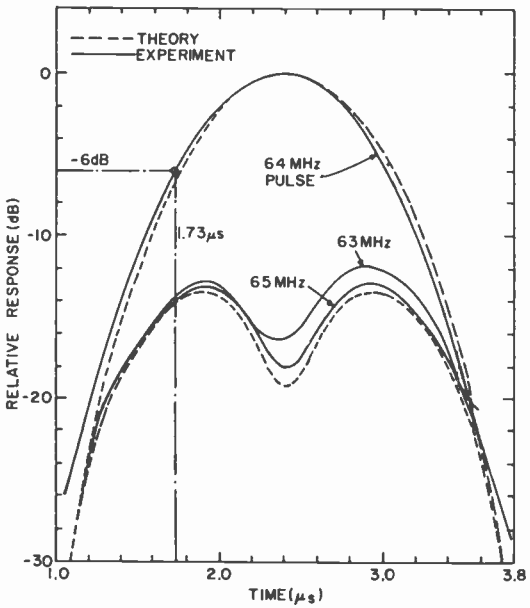


Fig. 4—Resonant and adjacent channel response characteristics for the 64-MHz SECANT filter.

(each lambda/4 finger is replaced by two lambda/8 fingers spaced lambda/4 apart). This solution was intentionally avoided because it complicates the fabrication of the filters. The masks needed for the filters are photoreductions of enlarged patterns generated by a mechanical (digitally positioned) pattern-making machine. These instruments necessarily have a minimum incremental step distance, which limits the allowable finger pattern periodicity that can be generated. The periodicity of the filter pattern has to be accurate to within about 0.05% in order to match the specification relating to adjacent-channel response. The resonant frequency of the filters has to be maintained to within ± 25 KHz.

6. Conclusions

An analytical method to evaluate the response characteristics of microsonic filters under pulse or transient excitation is presented. A synthesis procedure is established and used for the design of microsonic pulse response filters to replace lumped element five-pole Butterworth filters currently employed in the second generation hardware for SECANT. A set of six such filters, at different frequencies, has been successfully incorporated in the third-generation hardware, SECANT VECAS GM/A. Prototype systems of this hardware have been delivered to the U. S. Government for evaluation.[2] Measurements on the systems indicate that the performance of the filters is excellent. Furthermore, the filters proved to be reliable, relatively inexpensive, extremely compact, and very reproducible.

Acknowledgments

The authors wish to express appreciation to F. Wozniak for help in fabricating the microsonic filters, to members of the team working on the SECANT Project and, in particular, to J. Hall for suggesting the use of microsonic filters and L. Anderson for providing the measurement data on the filters installed in the SECANT prototype hardware.

References:

[1] J. L. Parsons, "SECANT: A Solution to the Problem of Mid-Air Collisions," Reprint RE-18-6-4, RCA Tech. Communications, Cherry Hill, N. J. 08101.

[2] L. Anderson, et al., "SECANT VECAS GA/M Flight Test and Evaluation Equipment," Final Report, TP2234, Naval Air Development Center (November 1975).

[3] A. Papoulis, *The Fourier Integral and Its Applications*, pp. 1–35 (McGraw-Hill Book Company, New York, 1962).

A New Generation of MOS/Bipolar Operational Amplifiers

Otto H. Schade, Jr.

RCA Solid State Division, Somerville, N. J.

Abstract—This paper discusses differences among the various emerging FET/bipolar technologies. It is argued that MOS/bipolar technology provides extended performance for low-cost high-yield linear IC's and that it will, therefore, coexist with "standard" bipolar processes.

Introduction

Until recently, the development of the monolithic operational amplifier (op-amp) could broadly be characterized as the use of new circuit techniques to achieve improved performance. The classical op-amp characteristics evolved from the circuit developments introduced by industry types 101, 108, and 741, and designs were then diversified for special applications such as high slew rate, programmability, ground-referenced sources, instrumentation use, micropower operation, and the like. Recent literature has comprehensively reviewed the elements of monolithic op-amp design.[1] Performance improvements have also been realized with super-beta transistors, which brought input-bias current into the nanoampere range, but the most dramatic changes have been brought about by recent technology advances. At the 1974 International Solid-State Circuits Conference, several manufacturers described[2–4] efforts in FET/bipolar circuit design, and new JFET/bipolar and MOSFET/bipolar op-amps were forecast.

This paper discusses the performance characteristics of several new MOS/bipolar amplifiers, how they compare with present monolithic

bipolar and JFET amplifiers, and what their future role appears to be. Three MOS/bipolar (BiMOS) monolithic circuits will serve to illustrate varied applications of bipolar and MOS devices in op-amp design:

(1) The "3100" is a wideband op-amp using PMOS enhancement devices for level shifting and output functions; it attains a 38-MHz unity gain crossing in contrast to the 1-MHZ figure of a lateral p-n-p transistor.

(2) The "3130" features a MOSFET input stage made possible by a new CMOS/bipolar process.[5] It combines high impedance, low-current MOSFET input, and low cost.[6,7] The amplifier has the ability to operate single supply with ground-referenced sources, and the CMOS inverter output provides a rail-to-rail swing for loads such as CMOS logic.

(3) The "3140", which is self-compensated, offers the input characteristics of the 3130 coupled with a broadband bipolar class AB output capable of sinking loads to the negative rail. Its internal bias circuit permits operation from 4 to 44 volts, single or dual supply, maintaining essentially constant speed, gain, and bandwidth over that range.

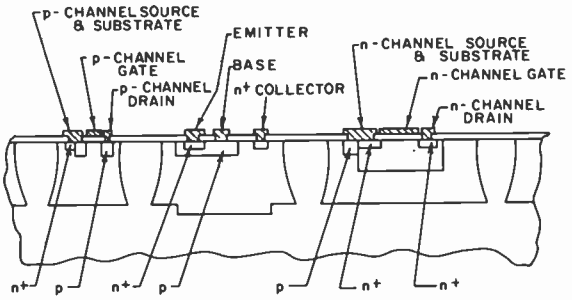


Fig. 1—Structural cross-sections of MOS/bipolar devices.

2. FET/Bipolar Monolithic Amplifiers

Fig. 1 shows structural cross sections of bipolar and MOS devices. The PMOS substrate is an N-epitaxial isolation "boat"; the p-type source and drain are formed from the 200 ohm-per-square bipolar base/resistor diffusion. The NMOS device is also built in isolated n-epi, with a p-substrate ion implant containing n+ emitter diffusions for source and drain. Gate oxide thickness is 1000 Å, thresholds are 1.7 volts, and a typical channel length is about 0.3 mil. A device having 1000 micromhos

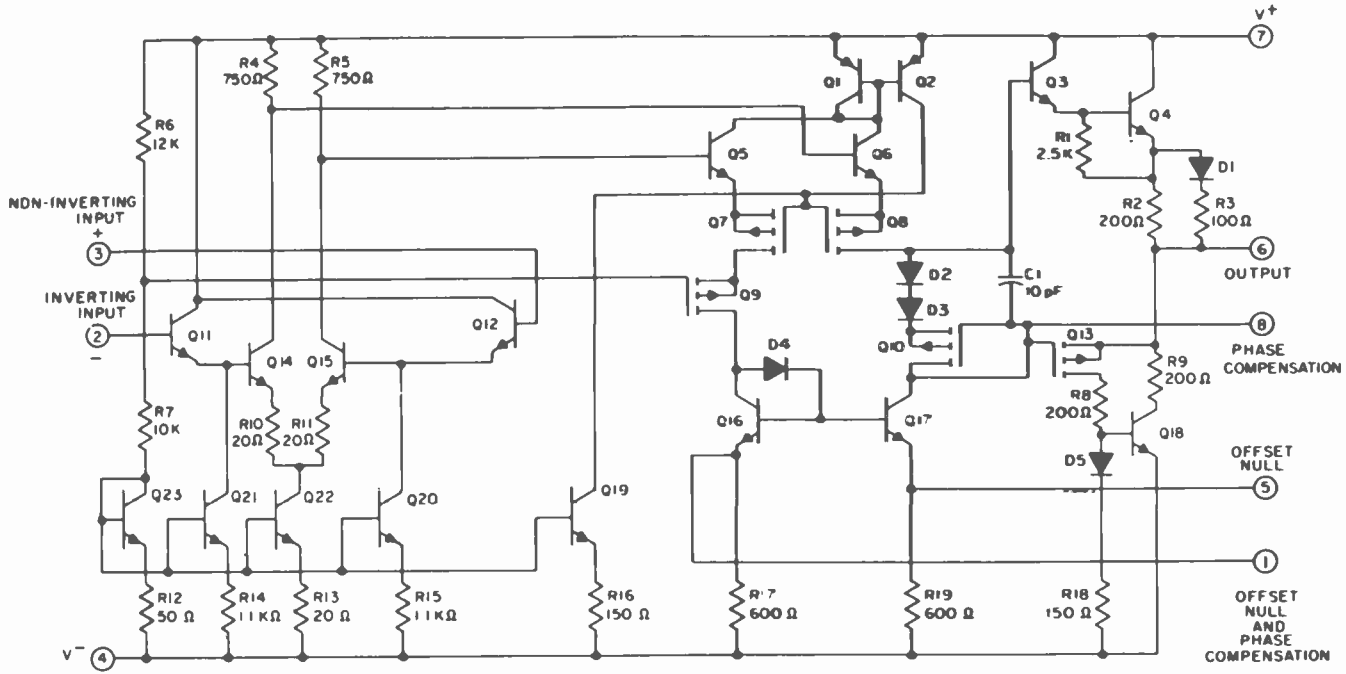


Fig. 2—CA3100 wideband operational amplifier schematic.

transconductance at 2 mA is readily built in a 7 × 10-mil isolation boat.

The 3100 amplifier schematic is shown in Fig. 2. The npn input stage is followed by a buffered grounded-gate PMOS pair that terminates in a bipolar mirror. Level shifting is therefore accomplished without the phase-shift limitations inherent with pnp lateral devices. The output stage employs a PMOS/npn-mirror composite as a p-type source fol-

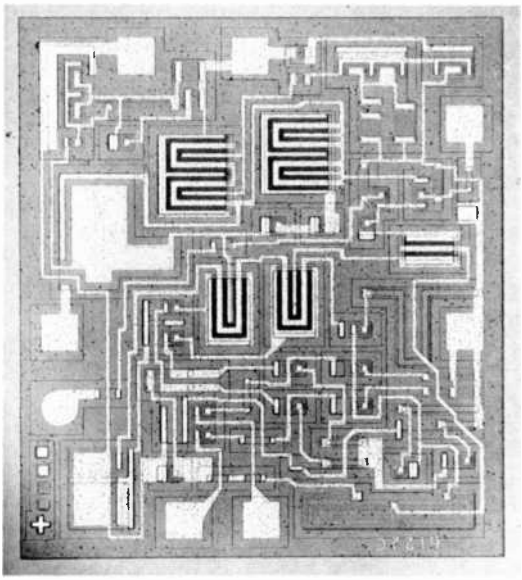


Fig. 3—The CA3100 chip.

lower, idling at 2 mA with 30 mA peak capability. Emitter and mirror ballasting provide for a stable response through a 50 MHz bandwidth. Fig. 3 shows the 53 × 60-mil chip.

The 3130 schematic is seen in Fig. 4 to be very simple for a three-stage operational amplifier. Division of input PMOS channel currents is established by an npn mirror that has been degenerated to improve balance and noise. The second bipolar gain stage provides for Miller compensation, and the output is a CMOS inverter. The NMOS imposes a breakdown limitation; about 18 volts V_{DS} for the 3130, although subsequent structural improvements would permit a 30-volt rating. However, power dissipation in such an inverter would be excessive, due to the high idling currents. The 59 × 69-mil 3130 chip is shown in Fig. 5. The cross-coupled input stage is composed of six separate PMOS's; the npn mirror (center of chip) and degenerating resistors are also cross-

coupled to remove first-order temperature gradient effects. The inverter NMOS is the lower of the two MOS's at the far right.

Particular attention must be paid to input- and output-stage design in order to obtain extended performance in an FET/bipolar amplifier. The low transconductance of an FET introduces potential compromises

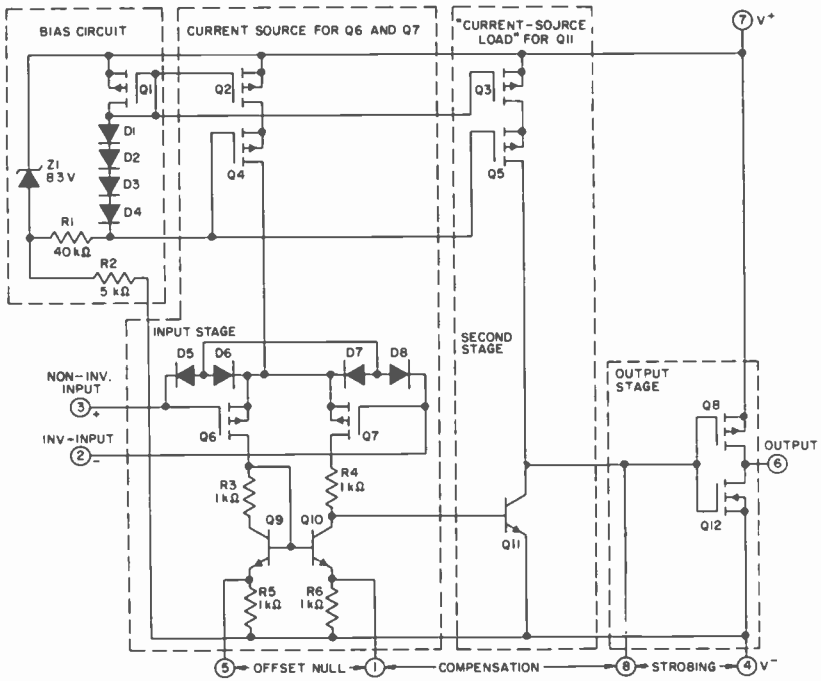


Fig. 4—CA3130 MOS/bipolar operational amplifier.

in power dissipation, offset voltage, temperature drift, and bandwidth. Intermediate amplifier stages and bias circuitry employing enhancement FET's can be designed with familiar bipolar techniques. JFET devices may require greater design efforts. For example, a current-mirror amplifier is not readily designed using a depletion device, since the gate potential of the standard configuration falls "outside" the supply rails to which sources are connected. A MOSFET gate may be overdriven in both positive and negative directions with respect to the source, without introducing a forward-biased junction (or any junction, for that matter). The use of dc feedback in JFET circuits supports the contention that the present state of the processing art permits closer control of MOSFET threshold (a few tenths of a volt) than JFET pinch-off voltage, thus

simplifying device biasing, level shifting, and common-mode range control.

In an amplifier having a simple input stage design, a rough measure of FET transfer-characteristic control is reflected in common-mode voltage and open-loop gain specifications. A design employing dc feed-

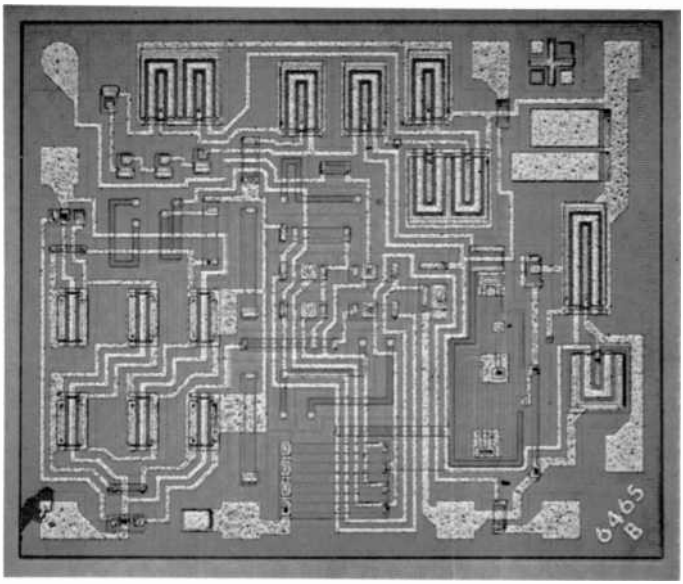


Fig. 5—CA3130 chip (input pair at lower left).

back (such as the LF356) can correct for pinch-off voltage spread; the potential impressed across a JFET current generator (sink) being compared to a series combination of bipolar base-emitter junctions and emitter resistor. However, the "error" due to pinch-off variations is now translated to changes of input-stage current. In addition, the error takes on the temperature dependence of the loop components. Evidence of such behavior is suggested by the published data for open-loop gain characteristics (Fig. 6). In contrast, the 3140 MOSFET input stage is quite stable, being operated at a current level where the transfer-characteristic temperature coefficient is essentially zero. The selection of op-amp compensation net-works is therefore straight-forward, and the negative-rail common-mode range is readily defined.

3. Differential FET Inputs

In general, the application of MOSFET pairs as monolithic op-amp in-

puts has demanded relatively large-area devices in order to obtain good transfer-characteristics match (without trimming) and acceptable transconductance. The relatively low transconductance of any monolithic FET is a disadvantage because, by definition, the input difference signal required to achieve current "balance" tends to be large. For example, a transconductance of 400 micromhos at 100 μA drain current

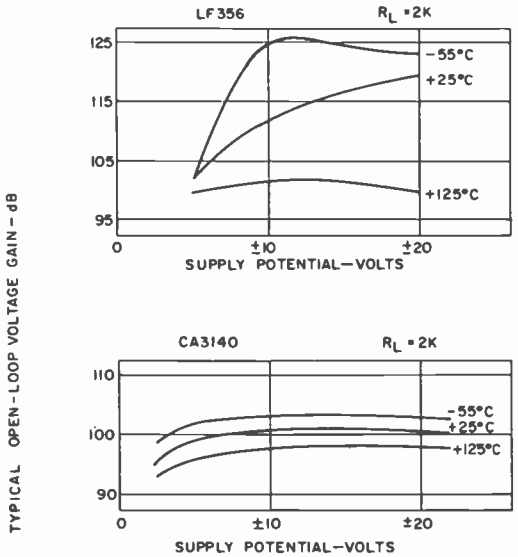


Fig. 6—Gain variations in JFET and MOSFET amplifiers.

requires 1 mV of signal to respond to a change of only 0.4 μA; that is, 0.4% I_D/mV. In contrast, a bipolar device responds with 4% I_C/mV. Not only must the FET pair be better matched than its bipolar counterpart, but the circuit means for establishing the drain currents must also be closely controlled. The criteria may be extended to include temperature drift and long-term stability as well.

Current practice in both MOSFET and JFET technologies requires large devices for low offset temperature drift. Noise performance is also improved, although the JFET's used in recent monolithic amplifiers are clearly superior to both their predecessors and the MOSFET in this respect. Disadvantages of large device pairs are higher parasitic capacitances and, in the case of JFET's, increased junction leakages. Although the MOSFET requires input protection from excessive voltage transients, the leakage associated with monolithic protective networks can be small in comparison and is unrelated to FET size.

MOS/bipolar technology provides means for building protective diode networks superior to those made with the standard CMOS process. In essence, the diode and its series resistance are placed in shunt with the gate/channel oxide. To be effective, a diode must turn on early in the static pulse, and its series resistance must be low enough that the potential impressed across it does not exceed the dielectric strength of the oxide. Bipolar emitter/base diffusions produce relatively fast diodes having several tens of ohms resistance rather than several hundred (as occurs in a standard CMOS process) and protect against current pulses

Table 1—Comparison of Input Characteristics of Bipolar and FET-Input Amplifiers

Type	Input Offset (mV)		Drift Typical (μV/°C)	Temperature Range (°C)
	Typical	Max		
	Bipolar Amplifiers			
318	4	10	—	0 to +70
324	2	7	7	0 to +70
308	2	7.5	6	0 to +70
741C	2	6	—	0 to +70
3100	1	5	—	−55 to +125
	FET-Input Amplifiers			
8007M/8007C	10/20	20/50	75 (max)	0 to +70
AD540J	—	50	20	0 to +70
CA3130	8	15	10	−55 to +125
CA3140	5	15	10	−55 to +125
CA3140A	2	5	10	−55 to +125
LF356	3	10	5	0 to +70

up to an ampere or more. Laboratory tests simulating static body discharges have shown MOS/bipolar-protected MOSFET's and JFET's to be similarly resistant and the 3130 is excellent in this respect. The test results are in general agreement with the observations of Kirk et al on static-induced damage of various semiconductor devices.[8]

A comparison of some "commercial grade" bipolar- and FET-input amplifier offsets, drifts, and temperature ranges is given in Table 1. It is seen that recent FET designs offer a performance approaching that of bipolar op-amps. Figs. 7 through 10 show laboratory measurements from which one may deduce offset drifts of the 356, 540J, 3130, and 3140 types. They appear to be representative of that characteristic, although the spread of voltage should not be construed as typical because of the small sample size. Published input bias currents are compared in Fig. 11, showing the expected advantage for the FET's. The translation from the published data for the 356 (which gives case temperature rather than common ambient temperature) was corroborated by spot-checks in the laboratory. It could be argued that in order to achieve good offset spec-

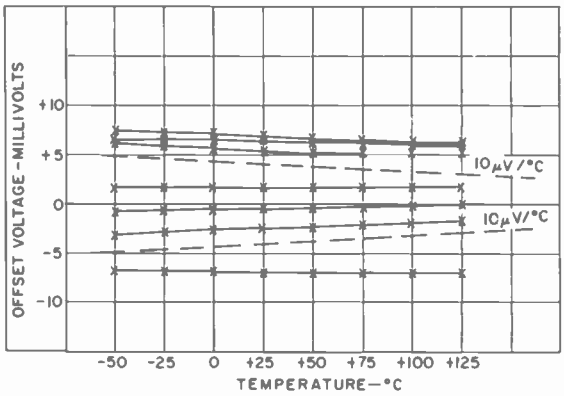


Fig. 7—LF356 input offset drift (laboratory data).

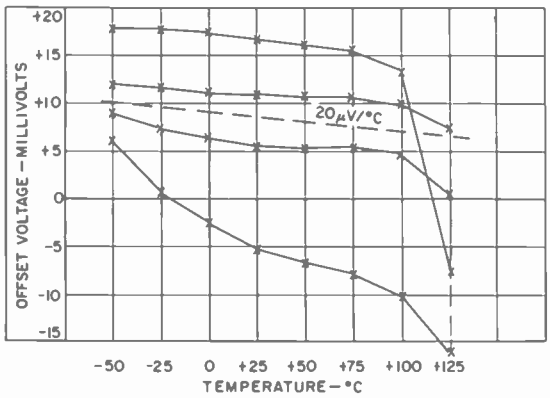


Fig. 8—AD540J input offset drift (laboratory data).

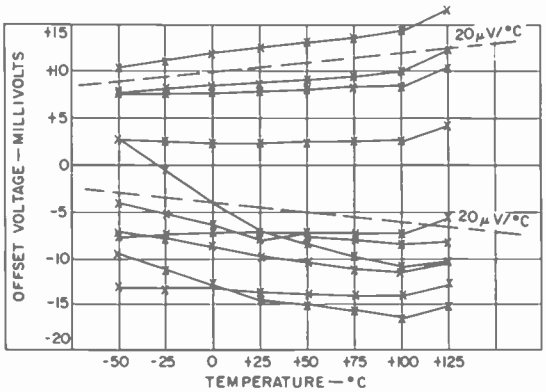


Fig. 9—CA3130 input offset drift (laboratory data).

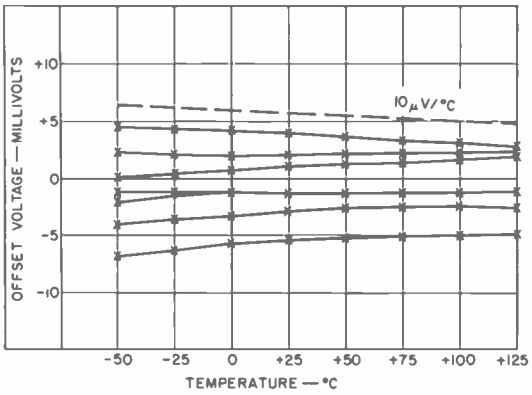


Fig. 10—CA3140 input offset drift (laboratory data).

ifications, the large-area JFET has sacrificed an order of magnitude increase in input bias.

A practical consideration for cost-sensitive applications is common-mode input range, particularly for single-supply operation. Fig. 12 shows the permissible operating range of the bipolar and FET-input amplifiers. It is seen that even though a gate/source bias potential may be large

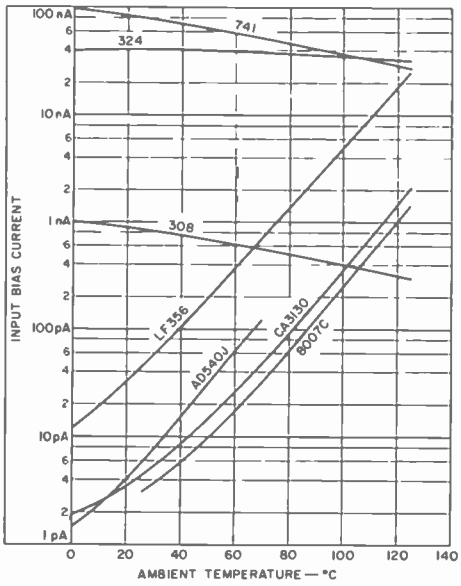


Fig. 11—Typical input bias currents (see text).

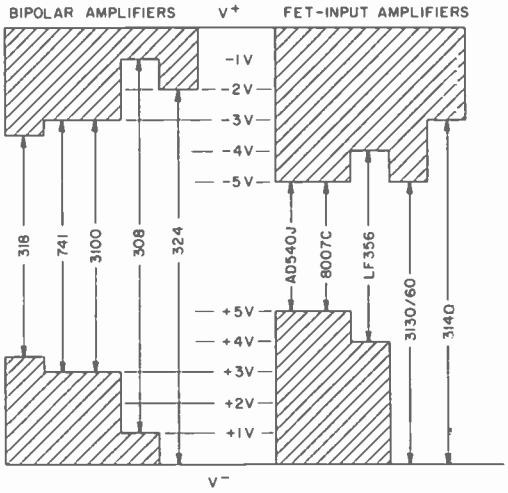


Fig. 12—Guaranteed common-mode input range.

compared to a base/emitter offset, satisfactory input ranges can still be achieved. The enhancement PMOS, in particular, readily provides the negative-rail capability of the 324, and in addition maintains output signal sense during overdrive.

Published equivalent input noise voltage with 100-ohm source resistance is shown in Fig. 13. The 3140 MOSFET pair is larger in active area than the 3130 and does exhibit slightly lower noise, although both types

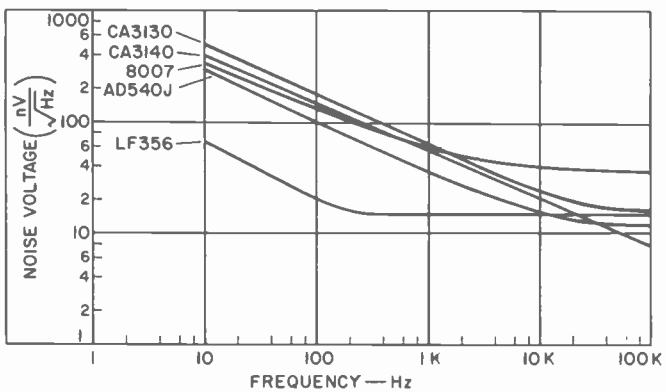


Fig. 13—Typical equivalent-input noise voltages.

are poorer than the 8007 and 540 JFET's. The LF356 shows markedly lower noise than any of the other types. For source resistances around 1 megohm or more, the MOSFET no longer suffers a disadvantage. Although it would be possible to reduce 3140 noise by further increases in channel area, the concurrent increase of input capacitance permits only modest reductions before speed and bandwidth are affected.

Stability of input offset during life is an important consideration for an FET-input amplifier, but little data has been published by manufacturers. MOSFET life drift, in particular, is known to be sensitive to impurities at the gate-oxide interface (a mechanism not found in the JFET); good performance demands a well-controlled manufacturing environment. As a general observation, inputs run at the low (0-volt) differential stresses that prevail during a linear operating mode usually show 1 to 2 mV drifts in 1000 hours at 125°C, and less than 1 mV drift

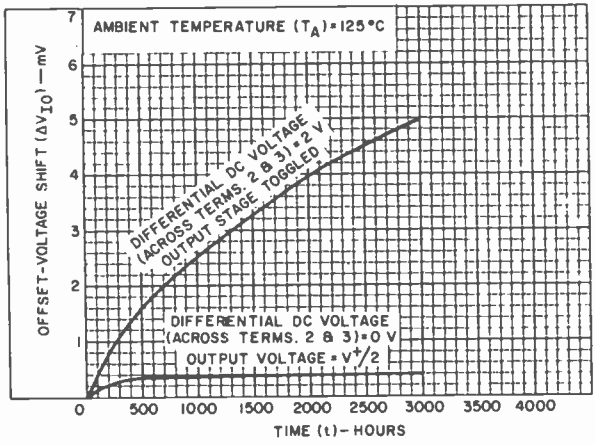


Fig. 14—Typical CA3130 offset voltage life-drift.

with ambients below 85°C. A large differential stress, such as encountered in comparator operation, can cause substantially greater drifts at 125°C (see Fig. 14). Additional tests show that stress-induced shifts anneal out when the voltage is removed. Published data on JFET-input characteristics[9] indicates the AD503 may drift a few mV in several thousand hours, although temperature and voltage were not specified. In comparison, bipolar offset drifts are commonly less than a few tenths of a mV in several thousand hours.

4. Wideband MOSFET Outputs

Simplified equivalents for the output stages of types 741, 3100, 3130, and 3140 are shown in Fig. 15. The Class AB emitter follower (type 741) has an idling current regulated by the current-source level and some form of temperature-tracking network Z between the transistor bases. The phase-shift limitations of the monolithic pnp are avoided with a composite PMOS/npn output device in the 3100, so that current sinking may be accomplished without excessive gate–source driving voltage. Stable performance is attained through 50 MHz bandwidth by judicious choice

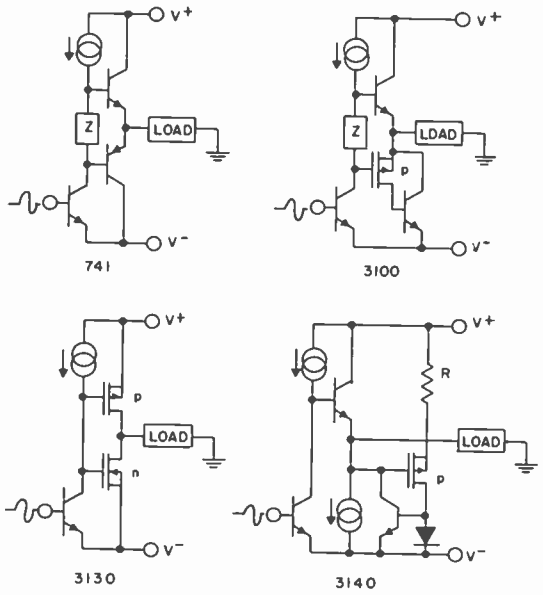


Fig. 15—Simplified output stage configurations.

of bias currents in both n and p legs, but voltage swing to the negative rail is inferior to that of a high-transconductance bipolar device. This limitation is not so important in an amplifier for video and similar signal-processing applications, but becomes a poor compliment to an input stage having negative-rail capability. The 3130 employs a CMOS inverter stage, so that moderate loads may be handled with voltage excursions close to both rails. The Class A biasing (derived by merely tying the gates together) depends upon the potential between supply rails, and is not practical for a large range of supply voltages because of power dissipation.

In addition, large voltage excursions are quite nonlinear and poorly suited for fast transient response. Advantages are the accuracy potential in follower operation,[10] the simplicity of power boosting by paralleling 3600 inverter-array stages, and ease of driving CMOS logic.

The 3140 employs a broadband output stage having unique properties. In the equivalent circuit shown, a constant current generator (sink) establishes the idling current, and an additional "dynamic" sink responsive to signal voltage level provides for load current pull down. The theoretically more-desirable conventional Class B circuits which are responsive to the current level of the load introduce a transistor "sense" junction that precludes the ability to pull down within logic levels of the rail. Resistor R is in effect a low-power dummy load, providing an increasing load current capability as load potential falls. The circuit may be designed with R returned to a reference potential intermediate between the rails, in which case for symmetrical dual-supply operation with ground-returned load, efficient Class AB operation is obtained. The 3140 dynamic sink pulls a 2000-ohm load within saturation potential of the negative rail from 4- to 44-volt supply operation. Sink demands in excess of the load requirement are supplied by the npn emitter follower. The penalty for load-voltage sensing is a pre-ordained sink-current characteristic; the mirror demands certain current levels whether the load resistor is large or small. For this reason, the approach is generally not suitable for "micropower" applications when the supply voltages are high. The output swing of several amplifiers is shown in Fig. 16, where the ability to drive a 2000-ohm load has become a common specification for an op-amp. Similar to the input common-mode comparison, the MOSFET amplifier is seen to be competitive with, if not superior to, bipolar and JFET amplifiers, a concept that was contested only a few years ago.

5. Extended-Voltage-Range Op-AMP

The 3140 represents an application of the 3130 input and second stages in an extended-voltage-range amplifier (see Fig. 17). Such an approach maintains negative-rail input capability, although it puts a burden on the input stage to achieve additional voltage gain. Increased transconductance-to-capacitance ratio is provided by the input pair (Fig. 18, lower left) by further layout refinements. Common-mode input range is improved by using a bipolar current source; actually cascoded pnp's. The 3140 bias supply, far left in the schematic, is a mirrored loop that stabilizes at a current determined by the emitter resistor and PMOS gate–source potential. The latter provides a 2.8-volt "reference", so that rel-

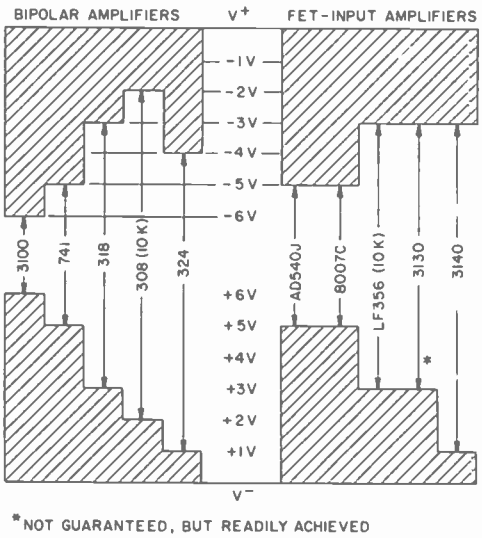


Fig. 16—Guaranteed 2000-ohm output voltage swing.

atively constant bias currents, gain, and speed are maintained down to about 4 volts between supply rails.

The output stage, as previously described, provides an efficient Class AB operation for dual supplies and typical op-amp loads. Typical 3130 and 3140 output swings are compared in Fig. 19. In addition, the new amplifier drives considerably greater load capacitance than the 3130

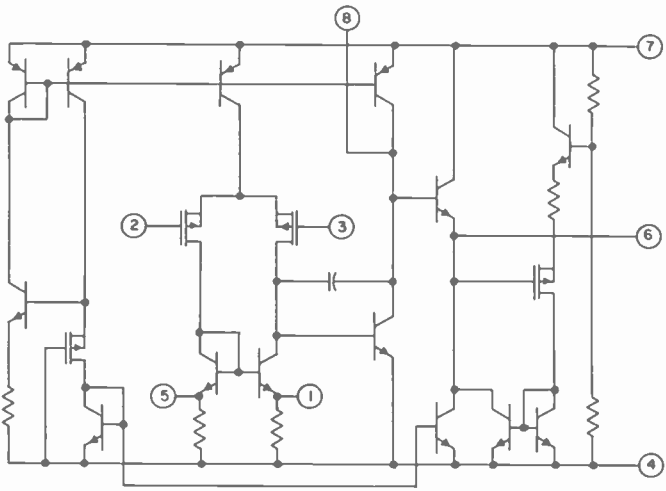


Fig. 17—CA3140 simplified schematic.

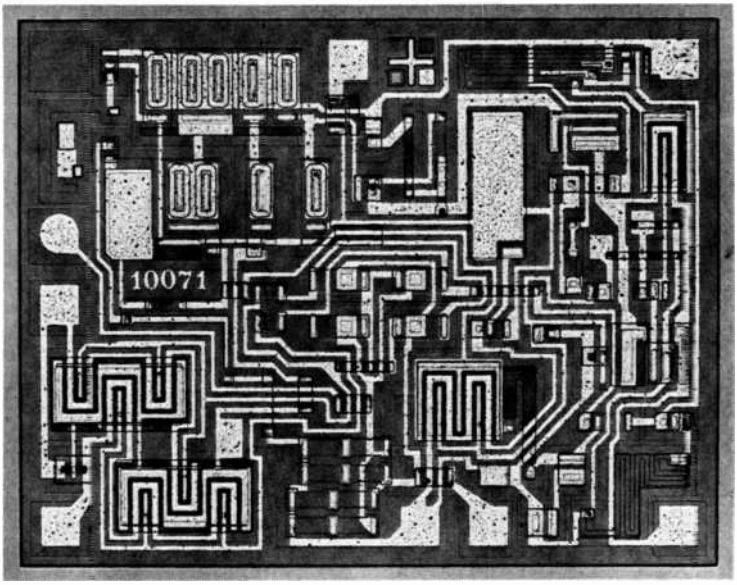


Fig. 18—CA3140 chip (input pair at lower left).

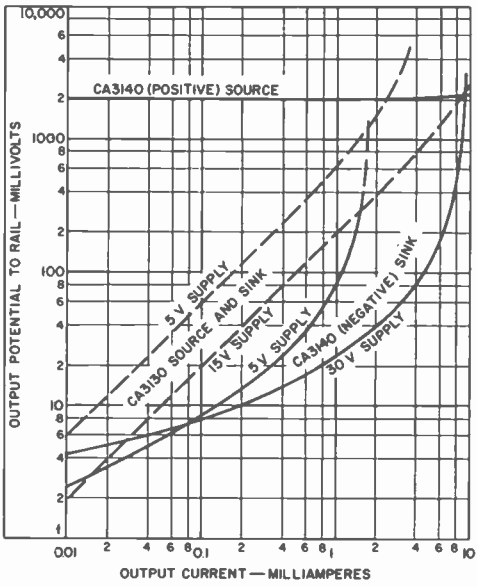


Fig. 19—CA3130 and CA3140 typical output swings.

without excessive ringing or oscillation. The laboratory measurements in Fig. 20 show that the LF356 and 3140 inverter and follower settling times with 2000 ohm/100 pF load are roughly comparable. For small signals, the 3140 provides a modest advantage.

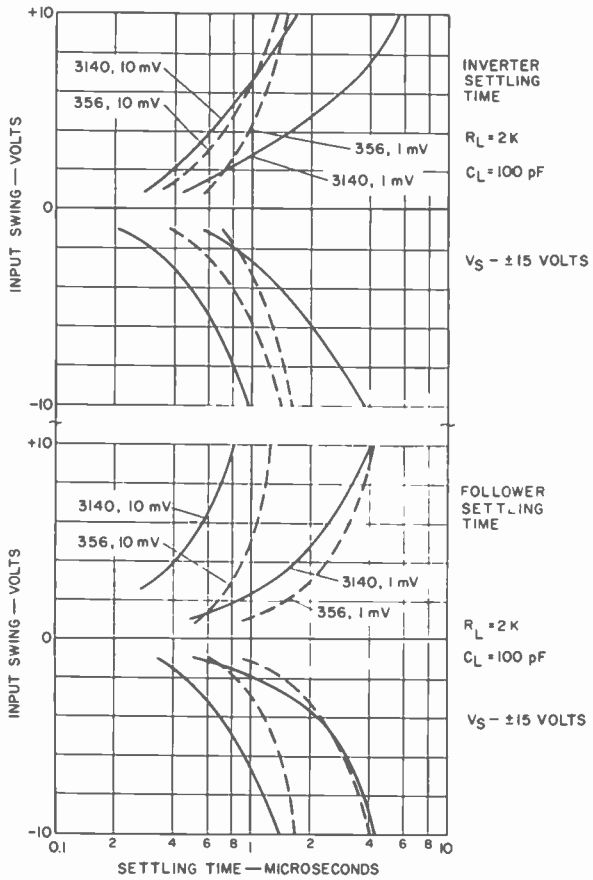


Fig. 20—LF356 and CA3140 inverter and follower settling times.

6. Future MOS/Bipolar Amplifiers

The gestation period of an integrated circuit may be a year or more, particularly when it employs a new process. As an example, the 3140 circuit design had already been established when the 3130 became commercially available. Further development of MOS circuits, many already thoroughly explored, are therefore presently under consideration

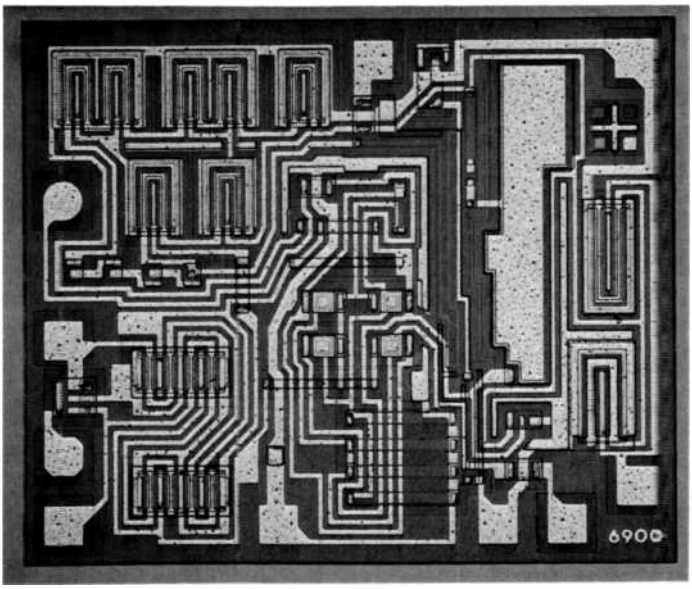


Fig. 21—CA3160 chip (input pair at lower left).

to improve existing amplifiers or for new applications. For example, the potential for an MOS/bipolar amplifier to achieve 100V/μsec unity gain slew rate is readily demonstrated. In a more compact layout (Fig. 21), a 50 pF compensating capacitor has been added to the 3130. The revised input-pair design provides a modest improvement in phase margin and transient response, exhibiting a lesser and more uniform input offset drift (Fig. 22). Another promising feature of this amplifier is a "guard band"

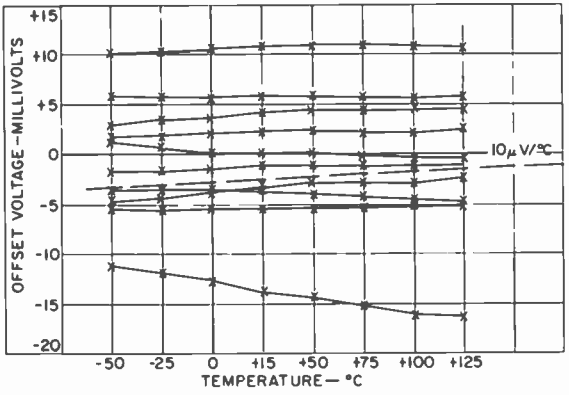


Fig. 22—CA3160 input offset drift (laboratory data).

wire-bond option, whereby the protective diode at either input terminal can be biased to near-zero potential, bringing input-bias current into the 0.2 pA range.

A recent op-amp layout (Fig. 23) employs an input pair made with advanced process techniques. Early samples demonstrate a radical improvement in the pair match, reducing input offset by a factor approaching 3:1 and increasing common-mode and power-supply rejections by 20 dB. Such performance suggests MOSFET offset distributions comparable to those of "general-purpose" bipolar pairs, without area or cost increase. Yield improvements alone provide sufficient impetus to employ the new approach, and all MOS/bipolar amplifiers should eventually benefit from it.

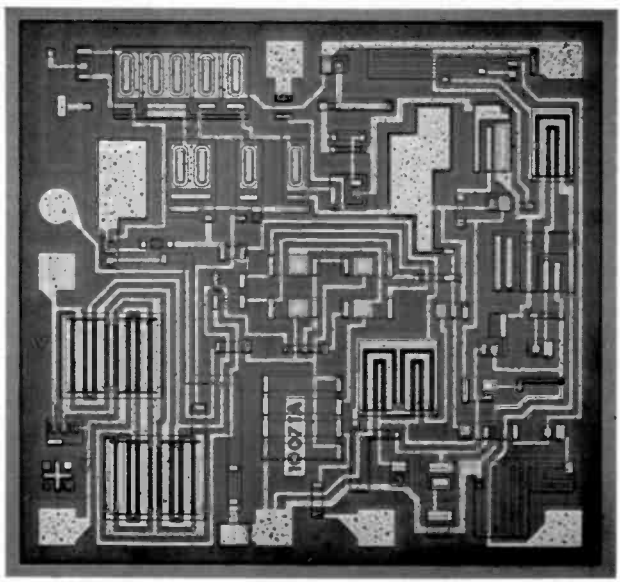


Fig. 23—Developmental op amp using advanced matching techniques.

7. Summary

The performance characteristics of several new MOS/bipolar amplifiers have been described and compared with those of monolithic JFET and bipolar devices. It is concluded that

(1) A monolithic MOS/bipolar process is suitable for extended performance economical linear IC's.

(2) The MOSFET op-amp excells in the compromise between input bias and offset and is presently superior in common-mode and output excursions.
(3) The reproducibility of a MOSFET transfer characteristic (threshold) appears more readily controlled in manufacture than that of a JFET (pinch-off).
(4) The MOS/bipolar process provides rugged input protection.
(5) MOSFET and JFET amplifiers are roughly comparable in speed/bandwidth capability, extending performance beyond bipolar IC levels. Offset temperature drifts are similar.
(6) MOSFET amplifier offset life-drift can approach that of a JFET amplifier, over moderate temperature ranges. Both are inferior to bipolar input drifts.
(7) The JFET is superior to the MOSFET in noise performance.

The acceptance and use of FET/bipolar IC's as fully accredited partners to standard devices is well under way. Both JFET and MOSFET approaches offer enhanced performance, notably superior input characteristics and improved speed/gain/bandwidth tradeoffs, assuring their evaluation in promising new applications. In addition, MOS/bipolar devices introduce MOSFET performance to low-cost, general purpose applications.

Acknowledgments

Recognition is due S. Ahrens, L. Baar, R. Baird, G. Checkowski, M. Hoover, J. P. Keller, R. Krachun, K. Orlowsky, M. Polinsky, T. Robe and H. Wittlinger for process, layouts, characterization, testing, applications, and sound advice. Without their skills and effort and the support of R. L. Sanquini, the BiMOS product line would not have materialized.

References:

[1] J. E. Solomon, "The Monolithic Op Amp: A Tutorial Study," *IEEE J. Solid-State Circuits*, **SC-9**, No. 6, p. 314, Dec. 1974.
[2] J. R. Butler and R. Q. Lane, "An Improved Performance MOS/Bipolar Op Amp," *Proc. 1974 IEEE International Solid-State Circuits Conf.*, p. 138.
[3] R. W. Russell and D. D. Culmer, "Ion Implanted JFET-Bipolar Monolithic Analog Circuits," *Proc. 1974 IEEE International Solid-State Circuits Conf.*, p. 140.
[4] O. H. Schade, Jr., "CMOS/Bipolar Linear Integrated Circuits," *Proc. 1974 IEEE International Solid-State Circuits Conf.*, p. 136.
[5] M. A. Polinsky, O. H. Schade, Jr., and J. P. Keller, "CMOS-Bipolar Monolithic Integrated-Circuit Technology," *Proc. 1973 International Electron Dev. Mtg.*, p. 229.

[6] R. L. Sanquini, "Building CMOS Bipolar Circuits on Monolithic Chip Enhances IC Specs," *Electronics*, Oct. 3, 1974.

[7] H. A. Wittlinger, "A New Breed of Op Amp Amplifies User's Options," *Electronic Products Mag.*, p. 51, April 28, 1975.

[8] W. J. Kirk, Jr., L. S. Carter, and M. L. Waddell, "Eliminate Static Damage to Circuits," *Electronic Design* 7, p. 80, March 29, 1976.

[9] J. Fishman, "Beware Those FET Op-Amp Specs," *Electronic Design 1*, p. 105, Jan. 4, 1975.

[10] O. H. Schade, Jr., "Digital-to-Analog Conversion Using the RCA-CD4007A COS/MOS IC," RCA Application Note ICAN 6080, Aug. 1972.

Recent Papers by RCA Authors

Listing is alphabetical by name of primary author. For copies of reprints, the reader should contact the publication directly.

M. S. Abrahams, "Epitaxy, Heteroepitaxy and Misfit Dislocations," **Crystal Growth and Characterization Proc. ISSCG2 Springschool,** Japan, 1974, p. 187, 1975.

Gerald A. Alphonse, "Optical Storage in Lithium Niobate," **Third Conf. on the Laser,** Vol. 267, p. 373.

P. K. Baltzer, J. A. Weisbecker, and R. O. Winder, "Interpretive Programming of Small Microprocessor-Based Systems," **Electro 76 Program,** Session 23, High Level Languages for Microprocessors, p. 1, May 11–14, 1976.

V. S. Ban and E. A. D. White, "Mass Spectrometric Study of Processes in the Closed-Tube Vapor Growth of CdS and ZnS," **J. Crystal Growth,** Vol. 33, p. 365, 1976.

R. A. Bartolini, A. Bloom, and J. S. Escher, "Multiple Storage of Holograms in an Organic Medium," **J. Appl. Phys.,** Vol. 28, No. 9, p. 506, May 1976.

R. A. Bartolini, A. Bloom, and D. A. Weakliem, "Volume Holographic Recording Characteristics of an Organic Medium," **Appl. Optics,** Vol. 15, No. 5, p. 1261, May 1976.

E. F. Belhoubek, A. Presser, and H. S. Veloric, "Improved Circuit-Device Interface for Microwave Bipolar Power Transistors," **J. Solid-State Circuits,** Vol. 11, No. 2, p. 256, April 1976.

A. E. Bell, "Dilute Transition Metal Alloys and the Kondo Problem," **Contemp. Phys.,** Vol. 16, No. 4, p. 375, 1975.

D. E. Carlson and C. R. Wronski, "Amorphous Silicon Solar Cell," **Appl. Phys. Lett.,** Vol. 28, No. 1, p. 671, June 1, 1976.

D. E. Carlson, "Injection of Ions into Thin Insulating Films from a Glow Discharge," **J. Appl. Phys.,** Vol. 47, No. 6, p. 2754, June 1976.

Y. S. Chiang, E. J. Denlinger, and C. P. Wen, "Contact Resistance of Metal-Silicon Systems at Microwave Frequencies," **RCA Review,** Vol. 37, No. 1, p. 107, March 1976.

R. B. Comizzoli, "Bulk and Surface Conduction in CVD SiO_2 and PSG Passivation," **J. Electrochemical Soc.,** Vol. 123, No. 3, p. 386, March 1976.

D. J. Channin and D. E. Carlson, "Rapid Turnoff in Triode Optical Gate Liquid-Crystal Devices," **Appl. Phys. Lett.,** Vol. 28, No. 6, p. 301, March 15, 1976.

R. S. Crandall, "Measurement of the Chemical Potential of Hydrogen in H_xWO_3," **Solid State Communications,** Vol. 18, p. 1409, 1976.

B. J. Curtis, "Temperature Asymmetrics and Fluctuations," **J. Electrochem. Soc.,** Vol. 123, No. 3, p. 438, March 1976.

W. R. Curtice and J. J. Risko, "Varactor-Tunable Pulsed Avalanche-Diode-Oscillator for X-Band Operation," **Electronics Lett.,** Vol. 12, No. 6, March 18, 1976.

W. Czaja, "BOSE-EINSTEIN Condensation and High Density Exciton Behaviour in AgBr," Molecular Spectroscopy of Dense Phases, **Proc. 12th European Cong. on Molecular Spectroscopy,** p 155, July 1–4, 1975.

D. A. De Wolf, "Comments on Uniqueness of Statistics Derived from Moments of Random Variables," **Proc. IEEE,** p. 371, March 1976.

B. Dorner, R. E. Ghosh, and G. Harbeke, "Phonon Dispersion in the Layered Compound PbI_2," **Phys. Stat. Sol. (B),** Vol. 73, p. 655, 1976.

J. Dresner and B. Goldstein, "Dissociation of MgO Films Under Heat and Electron Bombardment and Its Effects on Secondary Emission," **J. Appl. Phys.,** Vol. 47, No. 3, p. 1038, March 1976.

M. Ettenberg and H. Kressel, "Interfacial Recombination at (AlGa)As/GaAs Heterojunction Structures," **J. Appl. Phys.,** Vol. 47, p. 1542, April 1976.

Michael Ettenberg, Henry Kressel, and James P. Wittke, "Very High Radiance Edge-Emitting LED," **IEEE Jour. Quantum Electronics,** Vol. QE-12, No. 6, p. 360, June 1976.

A. W. Fisher and G. L. Schnable, "Minimizing Process-Induced Slip in Silicon Wafers by Slow Heating and Cooling," **J. Electrochem. Soc.,** Vol. 123, No. 3, p. 434, March 1976.

J. I. Gittleman, "Application of Granular Semiconductors to Photothermal Conversion of Solar Energy,"

Appl. Phys. Lett., Vol. 28, No. 7, p. 370, April 1, 1975.
W. E. Ham, "Test Data Reduction," **IEEE Trans. Manufacturing Tech.** Vol. MFT-5, No. 1, p. 24, March 1976.
J. J. Hanak and J. P. Pellicane, "Effect of Secondary Electrons and Negative Ions on Sputtering of Films," **J. Vac. Sci. Technol.,** Vol. 13, No. 1, p. 406, Jan/Feb. 1976.
T. T. Hitch, H. H. Whitaker, E. M. Botnick, and B. L. Goydish, "Chemical Analyses of Thick-Film Gold Conductor Inks," **IEEE Trans. Parts, Hybrids, and Packaging,** Vol. PHP-11, No. 4, p. 248, Dec. 1975.
P. T. Ho, A. Rosen, and J. Klatskin, "Power Combination of Broadband Trapatt Amplifiers," **Electronics Lett.,** Vol. 12, No. 1, Jan. 8, 1976.
D. M. Hoffman and F. J. Tams III, "Inexpensive Radiant Substrate Heater for Vacuum Use," **J. Vac. Sci. Tech.,** Vol. 13, No. 2, p. 647, Mar./Apr. 1976.
D. M. Hoffman and M. D. Coutts, "Preparation and Properties of Adherent, Transparent, Conducting Rh Films," **J. Vac. Sci. Technology,** Vol. 13, No. 1, p. 122, Jan./Feb. 1976.
S. T. Hsu, "A COS/MOS Linear Amplifier Stage," **RCA Review,** Vol. 37, No. 1, p. 136, March 1976.
A. C. Ipri and J. C. Sarace, "CMOS/SOS Semi-Static Registers," **J. Solid-State Circuits,** p. 337, April 1976.
A. C. Ipri and J. C. Sarace, "Low-Threshold Low-Power CMOS/SOS for High-Frequency Counter Applications," **J. Solid-State Circuits,** Vol. SC-11, No. 2, p. 327, April 1976.
T. Iwasa, "Chemical Etching and Fabrication of Ridge Waveguides on $Bi_{12}GeO_{20}$ Single Crystals," **J. Appl. Phys.,** Vol. 47, No. 6, p. 2746, June 1976.
W. Kern, "Analysis of Glass Passivation Layers on Integrated-Circuit Pellets by Precision Etching," **RCA Review,** Vol. 37, No. 1, p. 79, March 1976.
W. Kern, G. L. Schnable, and A. W. Fisher, "CVD Glass Films for Passivation of Silicon Devices: Preparation, Composition, and Stress Properties," **RCA Review,** Vol. 37, No. 1, p. 3, March 1976.
W. Kern, "Densification of Vapor-Deposited Phosphosilicate Glass Films," **RCA Review,** Vol. 37, No. 1, p. 55, March 1976.
E. S. Kohn, "A Charge-Coupled Infrared Imaging Array with Schottky-Barrier Detectors," **IEEE J. Solid-State Circuits,** Vol. SC-11, No. 1, p. 139, Feb. 1976.
W. F. Kosonocky and D. J. Sauer, "The ABCs of CCDs—Part I; Consider CCDs for a Wide Range of Uses—Part II," **Electronic Design,** April 12, 1975/March 15, 1976.
H. Kressel and F. Z. Hawrylo, "Red-Light-Emitting Laser Diodes Operating CW at Room Temperature," **Appl. Phys. Lett.,** Vol. 28, No. 10, p. 598, May 15, 1976.
H. W. Lehmann and R. Widmer, "Radio Frequency Sputtering of CdSe Films," **Thin Solid Films,** Vol. 33, p. 301, 1976.
M. Lurie, "Evaluation of Expanded Laser Beams Using an Optical Flat," **Optical Engineering,** Vol. 15, No. 1., p. 68, Jan.–Feb. 1976.
E. W. Maby, "Bombardment-Enhanced Diffusion of Arsenic in Silicon," **J. Appl. Phys.,** Vol. 47, No. 3, p. 830, March 1976.
R. U. Martinelli and G. H. Olsen, "Improved Transmission Secondary Emission from $In_xGa_{1-x}P$/GaAs Self-Supporting Films Activated to Negative Electron Afinity," **J. Apply. Phys.,** Vol. 47, No. 4, p. 1332, April 1976.
M. D. Miller, H. Schade, and C. J. Nuese, "Lifetime Controlling Recombination Centers in Platinum-Diffused Silicon," **J. Appl. Phys.,** Vol. 47, No. 6, p. 2569, June 1976.
R. M. Moore, C. J. Busanovich, F. Kozielec, Jr., and J. T. Fischer, "An Experimental Crystalline Selenium Vidicon," **IEEE Trans. Electron Devices,** Vol. 23, No. 6, p. 534, June 1976.
G. H. Olsen and V. S. Ban, "The Use of Thin Carbon Films for Selective Chemical Etching and Epitaxial Deposition of III–V Semiconductors," **Appl. Phys. Lett.,** Vol. 28, No. 12, p. 734, June 1976.
G. H. Olsen and R. T. Smith, "Misorientation and Tetragonal Distortion in Heteroepitaxial Vapor-Grown III–V Structures," **Phys. Stat. Sol. (A),** Vol. 31, p. 739, 1975.
S. M. Perlow, "Noise Performance Factors in Television Tuners," **RCA Review,** Vol. 37, No. 1, p. 119, March 1976.
R. J. Powell, "Vacuum Ultraviolet Induced Space Charge in Al_2O_3 Films," **Appl. Phys. Lett.,** Vol. 28, No. 11, p. 643, June 1, 1976.
W. Rehwald, K. Frick, G. K. Lang, and E. Meier, "Doping Effects Upon the Ultrasonic Attenuation of $Bi_{12}SiO_{20}$," **J. Appl. Phys.,** Vol. 47, No. 4, p. 1292, April 1976.
F. N. Sechi, "Linearized Class-B Transistor Amplifiers," **J. Solid-State Circuits,** Vol. 11, p. 264, April 1976.
H. Schade, "Evidence for Electron Emission Stimulated Desorption from Negative Electron Affinity GaAs Surfaces," **Surface Sciences,** Vol. 55, No. 1, p. 20, 1976.
P. Sheng, "Effect of Director Fluctuations on the Nematic Distribution Factor," **Solid State Communications,** Vol. 18, p. 1165, 1976.
I. Shidlovsky, "The Mechanism of Color Center Production in Iron Doped Photochromic Sodalities," **Solid State Communications,** Vol. 18, p. 155, 1976.

E. K. Sichel, R. E. Miller, M. S. Abrahams, and C. J. Buiocchi, "Heat Capacity and Thermal Conductivity of Hexagonal Pyrolytic Boron Nitride," **Phys. Review B,** Vol. 13, No. 10, p. 460, May 15, 1976.
H. S. Sommers and D. O. North, "Rotation of Polarization Axis of Injection Lasers with Viewing Direction," **J. Quantum Electronics,** Vol. 12, No. 4, p. 225, April 1976.
D. L. Staebler, "Binocularly Induced Motion of Flicker Patterns," **J. Opt. Soc. Amer.,** Vol. 66, No. 2, p. 156, Feb. 1976.
E. F. Steigmeier, R. Loudon, G. Harbeke, H. Auderset, and G. Scheiber, "Raman Scattering in $K_2Pt(CN)_4Br_{0.3}.3H_2O$," **Solid State Communications,** Vol. 17, p. 1447, 1975.
A. W. Stephens and J. L. Vossen, "Abstract: Measurement of Interfacial Bond Strength by Laser Spallation," **J. Vac. Sci. Technol.,** Vol. 13, No. 1, p. 38, Jan./Feb. 1976.
T. Takahashi and O. Yamada, "Growth of Cadmium Boracite Single Crystals by Chemical Vapor Transport," **J. Crystal Growth,** Vol. 33, p. 361, 1976.
H. S. Veloric and J. Mitchell, Jr., "Capacitors for Microwave Applications," **IEEE Trans. Parts, Hybrids, and Packaging,** Vol. 12, No. 2, p. 83, June 1976.
J. L. Vossen, "Transparent Conducting Films," **J. Vac. Sci. Technol.,** Vol. 13, No. 1, p. 116, Jan./Feb. 1976.
H. A. Weakliem and B. F. Williams, "Review and Analysis of Optical Recording Media," **Optical Eng.,** Vol. 15, No. 2, p. 99, March/April 1976.
R. K. Wehner and D. Baeriswyl, "Equation of State of a Classical Anharmonic Oscillator," **Physica A,** Vol. 81, p. 129, 1975.
R. Williams, "Electrochemical Reactions of Semiconductors," **J. Vac. Sci. Technol.,** Vol. 13, No. 1, p. 12, Jan./Feb. 1976.
J. P. Wittke, I. Gorog, and A. H. Firester, "CW Semiconductor Injection Lasers in Information Handling Systems," **Optical Eng.,** Vol. 15, No. 2, p. 128, March/April 1976.
J. P. Wittke, "Optical Fiber Communications Link Design," **Proc. SPIE,** 19th Annual Meeting, Vol. 63, p. 58, 1975.
M. H. Wood and R. Williams, "Hole Traps in Silicon Dioxide," **J. Appl. Phys.,** Vol. 47, No. 3, p. 1083, March 1976.

Patents Issued to RCA Inventors Second Quarter 1976

April

A. A. Ahmed Ground Fault Detection Apparatus (3,953,767)
J. A. Allen Pickup Arm Cartridge Apparatus (3,952,145)
A. Anchutin Replaceable Solar Array Panels (3,948,468)
A. F. Arnold Method of Electrolessly Depositing Nickel Phosphorus-Alloys (3,953,624)
T. L. Chase, D. Duranti, and R. Sassoli Method for Preparing Supplemental Filter for Lighthouse (3,953,209)
R. H. Dean Method of Manufacturing a Semiconductor Device (3,951,708)
J. Evans, Jr. In-Line Electron Guns Having Consecutive Grids with Aligned Vertical, Substantially Elliptical Apertures (3,952,224)
N. Feldstein Temperature-Stable Non-Magnetic Alloy (3,953,654)
D. M. Gavrilovic Novel Liquid Crystal Electro-Optic Devices (3,951,846)
A. M. Goldschmidt Reference Signal Generator for Tape Tension Servomechanism (3,949,244)
R. J. Gries Gating Circuit for Television SCR Deflection System (3,950,673)
J. R. Hall Gain-Controlled Amplifier (3,950,708)
R. E. Hanson and T. E. Nolan, Jr. Determining Engine Compression from Starter Motor Current (3,952,586)
R. J. Hollingsworth Logic Circuit (3,953,743)
J. Kane and H. Schweizer Process for Depositing Transparent Electrically Conductive Tin Oxide Coatings on a Substrate (3,949,146)
H. Kawamoto and D. J. Miller Injection Laser Modulation (3,953,809)
M. A. Leedom Pickup Arm Cartridge (3,952,147)
P. A. Levine Method of Operating Imagers (3,953,733)
D. D. Mawhinney and T. E. Walsh Constant Temperature Control for Transferred Electron Devices (3,953,878)
L. A. Olson Method of Tuning a Tunable Microelectronic LC Circuit (3,947,934)
M. Rayl, H. D. Hanson, and B. W. Beyers, Jr. High Voltage Aerosol Detector (3,949,390)
A. Rose and A. D. Cope Pickup Tube Target (3,952,222)
D. L. Ross Organic Volume Phase Holographic Recording Media Comprising an Alpha Di-Ketone (3,951,663)
D. L. Ross and L. A. Barton Electron Beam Recording Article with Q-Quinone Diazide Compound (3,950,173)
O. H. Schade, Jr. Current Proportioning Circuits (3,952,257)
O. H. Schade, Jr. Current Amplifier (3,953,807)
A. C. Sheng and M. V. Hoover Timer Circuits (3,950,657)
T. M. Schrader Adjustable Spring Mount for a Cathode Ray Tube Yoke (3,950,720)
E. M. Smith Erasing Method for Storage Tube Employing Raster Scan (3,950,669)
J. C. Turnbull CRT Comprising Strontium Metal Getter Films and Method of Preparation (3,952,226)
D. H. Vilkomerson Wave-Energy Imaging Technique (3,953,822)

May

T. G. Athanas Deep Depletion Insulated Gate Field Effect Transistors (3,958,266)
J. B. Beck Transistor Amplifier Stage with Device in its Temperature Compensated Bias Network used as Preliminary Amplifier (3,955,108)
W. L. Behrend Amplitude Modulation System (3,956,715) (3,955,155)
B. F. Bogner and D. F. Bowman RF Power Coupling Network Employing a Parallel Plate Transmission Line (3,958,247)
B. M. Childers and C. D. Fesperman, Jr. Diphenyl Continuous Foam Dyeing with Fabric Running over Rolls in Foam Bath (3,954,404)
E. L. Crosby, Jr. Process for Non-Destructive Inspection (3,956,631)
M. T. Duffy Surface Acoustic Wave Device (3,955,160)
G. S. Gadbois and F. W. Ragland, Jr. Method and Apparatus for Determining the Average Size of Apertures in an Apertured Member (3,955,095)

J. B. George Varactor Tuner Frequency Controller (3,958,180)
H. E. Haslau and W. E. Rigsbee Coil Winding Machine (3,957,216)
K. Knop Diffractive Subtractive Color Filtering Technique (3,957.354)
M. A. Leedom Signal Pickup Return Mechanism (3,954,272)
P. A. Levine Charge Coupled Device Systems (3,958,210)
I. Ladany and D. P. Marinelli Method for Forming an Ohmic Contact (3,959,522)
G. S. Lozier Metal Plated or Platable Article (3,956,535)
M. A. Polinsky Method of Making an Insulated Gate Field Effect Transistor (3,959,025)
F. R. Ragland, Jr. Method and Apparatus for Determining the Average Size of Apertures in an Apertured Member (3,954,337)
B. D. Rosenthal Current Mirror Amplifiers (3,958,135)
R. L. Shanley, II Set-up Arrangement for a Color Television Receiver (3,959,811)
G. A. Swartz and R. E. Chamberlain Method of Manufacturing a Semiconductor Device having a Lead Bonded to a Surface Thereof (3,956,820)
B. K. Taylor Pickup Arm Apparatus (3,956,581)
Z. Turski and J. L. Vossen, Jr. Metallized Lithium Niobate and Method of Making (3,959,747)
L. C. Upadhyayula and M. Nowogrodzki Microwave Delay Line (3,955,158)
L. C. Upadhyayula and S. Y. Narayan Planar Transferred Electron Device with Integral Nonlinear Load Resistor (3,959,807)
D. H. Willis D. C. Reinsertion in Video Amplifier (3,955,047)

June

J. Avins and B. J. Yorkanis Gain Control Arrangement Useful in a Television Signal Processing System (3,961,361)
L. A. Barton Process for Improved Development of Electron-Beam-Sensitive Resist Films (3,961,101)
W. P. Bennett Apparatus for Non-Destructively Testing the Voltage Characteristics of a Transistor (3,965,420)
P. Brown, Jr. Memory Dial for Teletypewriter Subscribers (3,961,130)
T. W. Burrus Velocity Correction Circuit for Video Discs (3,965,482)
W. F. Dietz Side Pincushion Correction System (3,962,602)
S. S. Eaton, Jr. CMOS Oscillator (3,965,442)
R. S. Engelbrecht Handwriting Identification Technique (3,962,679)
M. Ettenberg Method of Making a Transmission Mode Semiconductor Photocathode (3,960,620)
N. Feldstein Method of Making Duplicates of Optical or Sound Recordings (3,962,495)
J. S. Fuhrer Velocity Correction for Video Discs (3,967,311)
J. B. George Bandstart Detector System for a Television Tuning System (3,961,263)
R. E. Hanson and T. E. Nolan, Jr. Engine Brake Horsepower Test without External Load (3,964,301)
W. A. Harmening Steerable Mount (3,964,336)
S. A. Harper Method for Improving Adherence of Phosphor-Photobinder Layer During Luminescent-Screen Making (3,966,474)
D. I. Harris and L. B. Johnston Method for Developing Electron Beam Sensitive Resist Films (3,961,100)
R. J. Himics, S. O. Graham, and D. L. Ross Method of Preparing a Pattern on a Silicon Wafer (3,964,909)
T. T. Hitch and T. E. McCurdy Reactively-Bonded Thick-Film Ink (3,962,143)
S. K. Khanna Disc Record and Method of Compounding Disc Record Composition (3,960,790)
K. Knop Focused-Image Hologram System Providing Increased Optical Readout Efficiency (3,961,836)
W. F. Kosonocky and D. J. Sauer Charge Transfer Memory (3,967,254)
R. D. Larrabee Fluorescent Liquid Crystals (3,960,753)
P. R. Liller Low-Voltage Aging of Cathode-Ray Tubes (3,966,287)
G. Mark Method of Forming an Overlayer Including a Blocking Contact for Cadmium Selenide Photoconductive Imaging Bodies (3,964,986)
W. L. Oates Method of Assembling a Liquid Crystal Cell (3,960,534)
M. A. Polinsky Leakage Current Prevention in Semiconductor Integrated Circuit Devices (3,961,358)

H. M. Scott Regulated Switched Mode Multiple Output Power Supply (3,967,182)
K. J. Sonneborn Pattern Definition in an Organic Layer (3,962,004)
R. G. Stewart Input Transient Protection for Integrated Circuit Element (3,967,295)
G. F. Stockdale Method for Embossing a Pattern in Glass (3,961,929)
B. K. Taylor Disc Record Locked Groove Escape Apparatus (3,961,131)
J. C. Turnbull Method for Producing a Strontium Metal Film on Internal Surfaces of a CRT (3,964,182)
C. F. Wheatley, Jr. Amplifier with Over-Current Protection (3,967,207)
K. R. Woolling, Jr. Digital Control System (3,961,281)
C. T. Wu Self-Clocking, Error Correcting Low Bandwidth Digital Recording System (3,961,367)
B. Zuk Threshold Detector Circuitry, as for PCM Repeaters (3,962,549)

AUTHORS

Carl W. Benyon received his A.S. degree (cum laude) in Basic Engineering from Trenton Junior College in 1962. In 1975 he completed the electronics and transistor courses of RCA Institutes. He has done subsequent work in semiconductor device technology and physics with Rutgers University. Mr. Benyon joined RCA Labs in 1958. He has been involved in the study of semiconductor device problems including bonding of leads, damage by bonding techniques, high-temperature contacts, liquid epitaxial growth of semiconductors, radiation hardening of Al_2O_3 MIS devices, and advanced techniques in integrated circuit manufacture.

Joseph H. McCusker received a B.S. in Electrical Engineering from Northeastern University, Boston, Mass., in 1947, and a M.S. in Electrical Engineering from Massachusetts Institute of Technology, Cambridge, Mass., in 1949, and took further graduate work at Pennsylvania State University. From 1943 to 1946 he served in the U.S. Army and worked on the Atomic Energy Project from 1944 to 1946. At M.I.T., 1947–1949, he worked as a research assistant in the Digital Computer Laboratory on the development of an electrostatic storage tube for use as a high-speed random-access memory. From 1949–1951, he was an instructor in the Electrical Engineering Department at Pennsylvania State University, State College, Pennsylvania. From 1952–1956, he worked at Lincoln Laboratories, Bedford, Mass. He was the leader of the Magnetic Section which, in cooperation with other sections, developed the first working random-access high-speed ferrite memory. From 1956–1960, he helped set up the RCA Needham Materials Laboratory, where he worked in both initial development and later in production. He was the Manager of Advanced Development which synthesized ferrites with better temperature characteristics in both memory and high-frequency ferrite cores. Since then he has been a member of the technical staff at RCA Laboratories, where he has worked on fabrication of diode matrices by electron-beam bombardment, packaging of MOS transistors, the development of thin-film transistors, and on solid-state adaptive devices. Recently, he has been working on microsonic surface wave filters for signal processing.

Harry J. Moody graduated in 1948 from the University of Saskatchewan with the B.Eng. in Engineering Physics. He obtained a National Research Council Bursary, and Studentship to continue work with the University of Saskatchewan Betatron. He obtained the MSc in Physics in 1950. He spent a year at the University of Illinois working with 300 MeV Betatron, and then returned to Canada to take a position at the National Research Council. He obtained a PhD from McGill in 1955. He then joined the Research Laboratory of the Canadian Marconi Company. In 1961 he joined the Research Laboratories of RCA Limited as a Member of Scientific Staff to work in the field of millimeter waves. In 1966 he extended his interest in the electromagnetic spectrum to the infrared doing work on various military systems, including infrared detection

systems, infrared countermeasures systems, infrared simulation systems, and laser ranging systems. For some time he worked in the field of electronically-scanned antenna arrays, during which he determined the systematic design techniques for the Butter Matrix and made detailed studies of waveguide slot parameters. In late 1968 he provided the antenna design for the RCA study for the Canadian domestic Communication Satellite. He has been involved in a number of satellite system studies. These include, besides the Telesat Project, several uhf communications satellites studies, the communication technology satellite, the U.S. Domestic Communication System, and a millimeter-wave radiometer study for measuring atmospheric water vapor remotely. He is member of the Canadian Association of Physicists.

Stuart S. Perlman received the B.S. degree in Electrical Engineering from Rensselaer Polytechnic Institute, Troy, New York, in 1959, and the M.S. and Ph.D. degrees in Electrical Engineering from Carnegie Institute of Technology, Pittsburgh, Pa., in 1960 and 1963, respectively. During his undergraduate years, he was a cooperative student with the IBM Corporation, where his work included studies in semiconductor photoconductivity, thin-film cryogenic elements, ferrite core materials, and semiconductor junction devices using electron-beam bombardment. His doctorial thesis was an investigation of semiconductor heterojunction structures. Since 1963, he has been a member of the technical staff of RCA Laboratories, Princeton, N.J., where he has been engaged in studies of electrical and optical properties of high resistivity gallium phosphide crystals, of adaptive piezo-ferroelectric solid-state devices, and, more recently, of acoustic surface-wave filters for signal processing. In 1969, he was a recipient of an RCA Laboratories Outstanding Achievement Award.

Dr. Perlman is a member of Eta Kappa Nu, Tau Beta Pi, Sigma Xi, the Institute of Electrical and Electronic Engineers, and the American Physical Society.

Otto H. Schade, Jr., received the BEE degree from Rensselaer Polytechnic Institute in June, 1953. Following graduation, he joined the Electron Tube Division of RCA at Harrison, N.J., where he designed damper and deflection tubes for TV applications, Nuvistors, and mechanical filters; he also conducted studies of heat flow and electron optics in vacuum tubes. During 1961, Mr. Schade was responsible for the electrical performance and testing of silicon-germanium thermoelectric converters for Atomics International's SNAP 10A space program. He then built high-efficiency gaseous and liquid-fuel radiant burners for 1000 to 60,000 BTU/hr applications, and performed the system design of a 100-watt portable thermoelectric generator.

In 1966, Mr. Schade transferred to the Solid State Division, Somerville, N.J., where he developed silane-deposition systems, studied means for automatic photomask alignment and projection printing, and analyzed steady-state and transient thermal behavior in rf power transistors. During 1968, he assisted O. H. Schade Sr. in the development of the high-definition TV art by designing solid-state horizontal and vertical deflection amplifiers for a 100-MHz, 4000-line system. In 1969, Mr. Schade joined the Linear IC group at Somerville, where he has been active in the design, development, and application of industrial and custom circuits for uses such as CMOS D/A conversion, ground-fault interrupt service, and conventional and radiation-hard operational amplifiers. He is the architect of RCA's first MOS/bipolar monolithic amplifiers.

Mr. Schade is a member of Eta Kappa Nu.

Kenneth M. Schlesier received his BS in Electrical Engineering from Newark College of Engineering in 1965. From 1965 to 1970 he attended Princeton University to study device physics. In 1970 he received the PhD degree in Electrical Engineering from Princeton University. His dissertation dealt with "Photoconductivity and Contact Effects in Amorphous Selenium." During his graduate work he was a Summer Fellow at the General Electric Research and Development Center where he worked on infrared television cameras. Dr. Schlesier joined RCA Laboratories in 1970, where he has been concerned with the development of advanced integrated circuits. He has worked on process development of Al_2O_3 for use in radiation hardened CMOS circuits and has investigated the electrical and radiation behavior of Al_2O_3/MOS devices. He has also worked on ion implantation doping of silicon-on-sapphire films for CMOS circuits.

Dr. Schlesier is a member of Tau Beta Pi, Eta Kappa Nu, American Physical Society and IEEE.

Joseph M. Shaw attended Seton Hall University where he received his B.S. in Physics in 1963. He has done graduate work in Metallurgy at the University of Pennsylvania. After his discharge from the U. S. Navy in 1955, he was employed by the E. I. Dupont Company, in Newark, New Jersey, as a member of the Physical Chemistry Laboratory of the Pigments Department, where he was concerned with the physical characterization of organic pigments and later in the development of flake pigments produced by chemical vapor deposition of metals and oxide films. In 1964, he joined the Process Research and Development Laboratory at RCA Laboratories, working on the application of vapor deposition techniques in the metallization of semiconductor materials. In 1970 he joined the Process and Applied Materials Research Laboratory where he was engaged in research and techniques relating to the vapor growth of insulators and oxides on MOS devices. He is presently a member of the Integrated Circuit Process Research Group. In 1975, Mr. Shaw was the recipient of an RCA Laboratories Achievement Award for team performance in developing improved dielectrics for MOS devices.

Issie P. Shkarofsky graduated in 1952 from McGill University, Montreal with a B.Sc. degree and first class honours in physics and mathematics. In the following year, 1953, he obtained his M.Sc. degree, conducting his research at the Eaton Electronics Research Laboratory, McGill University, in the fields of microwave optics and antennas. He then joined the microwave tube and noise group at the Eaton Electronics Research Laboratory, McGill University, and received his Ph.D. degree in 1957 with a thesis on modulated electron beams in space-charge-wave tubes and klystrons. After graduation, he joined the Research Laboratory of RCA Limited, Montreal, where he has participated in research on microwave diffraction, millimeter waves, obstacle gain, laser applications, electromagnetic wave interaction with plasmas, plasma kinetics, and plasmas in space. In 1973 he became a R&D fellow at the RCA Research and Development Laboratories, presented in the Physical Electronics Laboratory. In the field of plasma and space studies his particular interest has been in the following topics: plasma transport coefficients, collisional effects in plasmas (slightly, partially and strongly ionized), Boltzmann and Fokker-Planck theory and appropriate expansions, bremsstrahlung, magnetohydrodynamics, re-entry plasma physics, generalized Appleton-Hartree equation for the ionosphere, laboratory simulation of geophysical phenomena such as the interaction between the solar wind and the magnetosphere and the sheaths around satellites, cyclotron harmonic reso-

nances and related dispersion effects, diagnostics of plasmas by laser scattering, nonlinear mixing of plasma modes, the accuracy of Langmuir probes on satellites, VLF sheath admittance of antennas in the ionisphere, high power laser interaction with matter, turbulence and fusion.

Dr. Shkarofsky is a member of the Canadian Association of Physicists, of the American Physical Society, the American Geophysical Union, and of the National Research Council Advisory Committee on Physics. He is a professeur invité du Centre de Recherche INRS-Energie, Un. du Québec, and he is also the co-leader of the Task Force on Toroidal Machines for Project Fusion Canada.

Harold S. Veloric received the B.A. in chemistry from the University of Pennsylvania in 1951 and the M.S. and Ph.D. degrees in physical chemistry from the University of Delaware in 1952 and 1954, respectively. Dr. Veloric joined the Bell Telephone Laboratories in 1954 where he worked on the development of various types of silicon diodes including power rectifiers, voltage-regulatory diodes, computer diodes, and solar cells. He joined RCA Solid-State Division in 1958, where he worked on the design and development of high frequency transistors for consumer and computer applications. In 1960 he was responsible as Engineering Manager for the design and development of signal and switching transistors, and subsequently for MOS Engineering, Device Process Development, and RF transistor Engineering.

He is currently a Member of the Technical Staff of the Microwave Technology Center at RCA Laboratories, working on technology developments for the microwave group.

Richard A. Ulene is a graduate student at the Massachusetts Institute of Technology where he is completing work toward the Bachelor's degrees in Electrical Engineering and the Life Sciences, as well as a Master's degree in Biomedical Engineering. Under the direction of the M.I.T. and RCA Cooperative programs, he has worked on Fourier optics for television applications, acoustic surface-wave filters, separate-sound i-f filtering for television, computer-aided filter design, and PCM color television.

Mr. Ulene is a member of Sigma Xi and the Audio Engineering Society.