Digital Amplitude-Phase Keying with M-ary Alphabets

C. MELVIL THOMAS, MEMBER, IEEE, MICHAEL Y. WEIDNER, MEMBER, IEEE, AND S. H. DURRANI, SENIOR MEMBER, IEEE

Abstract—**Signal sets employing amplitude and phase keying (APK) with large alphabets conserve bandwidth and do not require as high a signal-to-noise ratio (SNR) as phase-shift keying (PSK). Twenty-nine empirically generated APK sets with M-ary alphabet sizes from 4 to 128 are investigated to determine optimum designs. Selected sets are compared on the basis of a symbol-error-probability bound for both average and peak SNR. Results are presented in the form of symbol error curves.**

The degradation caused by nonlinearities in a typical TWT amplifier is examined. The effects of AM-AM and AM-PM distortion on APK are presented as a function of power backoff for single-carrier operation (when APK is at the greatest disadvantage compared with PSK). Modem implementation methods, including carrier phase reconstruction techniques, are discussed for basic APK designs.

I. Introduction

THE INCREASING demands of data transmission via satellites with limited bandwidth allocations point to the need for modulation techniques that are highly efficient in terms of bandwidth. Phase-shift keying (PSK) with an M-ary symbol alphabet is one such technique requiring less bandwidth than FM or biphase PSK. It has been shown [1]-[3] that a hybrid modulation combining both amplitude and phase keying (APK) requires less power than PSK for the same error probability and alphabet size. APK is therefore a potentially attractive modultation technique for satellite communication and an investigation [4] was undertaken to examine its performance for a typical transponder channel. The study included signal set design, intersymbol interference, TWT distortion, and modem implementation. M-ary PSK, because of its widespread use, is employed as a standard of performance comparison.

To assist in signal-set design and performance evaluation, an asymptotic expression is obtained for the probability of symbol error at high SNR's for signal sets with unequal energies and any dimensionality. While this expression for error probability has not yet led to a solution of the theoretically optimum design for APK signal sets, it does permit an accurate comparison of candidate designs. This paper reports the results of such a comparison for 29 M-ary ($2 \leqslant \log_2 M \leqslant 7$) APK designs, offering counter examples to previously reported optimum 8-ary and 16-ary sets. These signal sets are described in Section II, and their symbol error performance is compared in Section III for peak and average SNR's.

Paper approved by the Associate Editor for Data Communications of the IEEE Communications Society for publication after presentation at the 1972 International Telemetry Conference, Los Angeles, Calif. Manuscript received May 3, 1973. This paper is based on work performed in COMSAT Laboratories and TRW Systems Group under the sponsorship of the International Telecommunications Satellite Organization (INTELSAT). Any views expressed in this paper are not necessarily those of INTELSAT or COMSAT.

C. M. Thomas and M. Y. Weidner are with TRW Systems, Inc., Redondo Beach, Calif. 90278.

S. H. Durrani was with COMSAT Laboratories. He is now with Operations Research, Inc., Silver Spring, Md. 20910.

Since actual communication links are seldom distortion-free, APK performance evaluation is most meaningful when these distortions are taken into account. In a typical link the major distortion elements are the channel filter and amplifier nonlinearities. The channel filter causes a loss in signal power and introduces intersymbol interference. The symbol-error performance of various APK signal sets has been analyzed as a function of symbol rate-to-bandwidth ratio [4]. The degradation in error performance of the APK sets due to the filtering is very nearly the same as for PSK of equal alphabet size [5] and will not be discussed.

The nonlinear amplifier is a traveling-wave tube (TWT) producing both AM-AM and AM-PM distortion. Because the APK signals lack a constant envelope, they are highly sensitive to these nonlinearities. The TWT must therefore be operated below maximum power unless some form of signal predistortion is employed. Section IV analyzes the tradeoff between distortion degradation and power backoff for single-carrier operation.

The two-dimensional nature of APK leads to modem implementations that are significantly different from PSK or ASK and realizations in terms of Cartesian and polar signal representation are presented. The benefit that may be gained from examining both the matched filter and the decision region viewpoints of maximum likelihood demodulation is discussed, as well as a method for carrier-phase reconstruction.

II. Signal Designs

A. Vector Representation

APK signal sets of alphabet size M can be viewed mathematically as M two-dimensional signal vectors. To define these vectors, let the signal waveform be represented over the symbol period T by

$$s_i(t) = \sqrt{\frac{2}{T}}\, a_i \cos \omega_0 t - \sqrt{\frac{2}{T}}\, b_i \sin \omega_0 t, \qquad -T/2 \leqslant t < T/2, \quad i = 1, \cdots, M.$$

If the orthonormal set

$$\psi_1(t) = \sqrt{\frac{2}{T}} \cos \omega_0 t$$

$$\psi_2(t) = -\sqrt{\frac{2}{T}} \sin \omega_0 t$$

is selected as the basis waveforms, $s_i(t)$ becomes

$$s_i(t) = a_i\psi_1(t) + b_i\psi_2(t) \tag{1}$$

and can be represented by the complex quantity

$$S_i = a_i + jb_i, \quad i = 1, \cdots, M. \tag{2}$$

At the detector input the signal is accompanied by noise which has passed through a filter matched to the signal. If this noise is defined as

$$n(t) = \sqrt{\frac{2}{T}}\, n_1(t) \cos \omega_0 t - \sqrt{\frac{2}{T}}\, n_2(t) \sin \omega_0 t, \qquad -T/2 < t < T/2, \tag{3}$$

it can also be represented in complex form by

$$N = n_1 + jn_2. \tag{4}$$

The noise is assumed to be Gaussian and white with single-sided power spectral density N_0 W/Hz. The two-sided spectral densities $n_1(t)$ and $n_2(t)$ are therefore

$$\Phi_{n_1}(\omega) = \Phi_{n_2}(\omega) = TN_0/2$$

so that the noise component variances are

$$E\{n_1^2\} = E\{n_2^2\} = N_0/2 \triangleq \sigma^2/2$$
$$E\{n_1 n_2\} = 0 \tag{5}$$

where $E\{\cdot\}$ denotes the expected value.

The average-signal-power to average-noise-power ratio is

$$\mathrm{SNR}_{\mathrm{av}} = \frac{\sum_{i=1}^{M} (a_i^2 + b_i^2)}{M\sigma^2}. \tag{6}$$

Similarly, the peak-signal-power to average-noise-power ratio is

$$\mathrm{SNR}_{\mathrm{pk}} = \max_i \frac{a_i^2 + b_i^2}{\sigma^2}, \quad i = 1, \cdots, M. \tag{7}$$

B. Previous APK Signal Designs

Theoretical work in signal design related to APK sets was performed by Dunbridge [6] as he extended the work of Weber [7] and Balakrishnan [8] to asymmetrical signal sets. He established that the optimum signal arrangement for low SNR was to locate three of the $M(\geqslant 4)$ signals on a circle of radius $\sqrt{M/3}$ and map the remaining $M-3$ signals into the origin. Unfortunately, no correspondingly elegant design is available for the more useful case of high SNR.

Most investigators of APK signal design have concerned themselves with optimizing a particular type of configuration—either in the form of a grid [2] or concentric circles, with [7] or without [1], [9] constraining the signals on each circle to lie on a ray from the origin.[1] A triangular configuration has also been reported since this study was completed [10].

Some theoretical guidance for the design of very large signal sets is available from Shannon [11] and Blachman [12] in their consideration of the signal statistics of optimum analog AM and PM systems at high SNR's. Shannon has shown that under an average power constraint the channel capacity is maximized with a signal that is uniformly distributed in phase and Rayleigh-distributed in amplitude, i.e., two-dimensional Gaussian. Under a peak power constraint the capacity is maximized by a signal uniformly distributed in phase with an amplitude probability density that increases linearly with amplitude up to the peak amplitude constraint.

[1] Lucky has indicated by private communication that his designs were constrained to lie on concentric circles with nonzero radius [1].

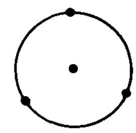

Fig. 1. 4-ary signal set design; (1, 3).

The approach employed in this investigation has been to use the above results as guidelines in an empirical search for good designs. Reasonable designs are selected as candidates for error probability evaluation. While the optimum design may not be discovered by this method, the optimum performance should not be greatly superior to that of the best candidates considered, if a large class of designs is tested.

C. Candidate Designs

The signal designs selected for error probability evaluation will now be described. Most fall into four basic categories: those with signals arranged on concentric circles, and those with triangular, rectangular, or hexagonal grid patterns. All alphabet sizes that are powers of two from 4-ary to 128-ary are considered.

In the following discussion and figures, the signal designs have been scaled such that, for all but the circular sets, the nearest neighbor distance is unity. For the circular sets the signals on each ring are separated by unit distance. For notational convenience the circular sets are designated as, for example, (a, b, c) meaning there are three circles with "a" signals on the inner circle, "b" signals on the next larger circle, etc.

4-Ary Signal Set: Only one 4-ary signal set is considered for comparison with PSK. It consists of three signals equally spaced about a circle with the fourth at the origin, and is designated as a (1, 3) circular set. This set is illustrated in Fig. 1.

8-Ary Signal Sets: For 8-ary signal sets there are four candidate configurations as shown in Fig. 2. The first design is a (1, 7) circular set consisting of a seven-signal circle surrounding a signal at the origin. The second is a double circle with four signals per circle, denoted as a (4, 4) design, which Lucky and Hancock [1] calculate to be the optimum 8-ary design under an average power constraint. The remaining two candidates are a 3 × 3 grid minus the center signal and a triangular design.

Two additional designs were investigated but considered to be unsuitable as candidates. A hexagonal design, due to its 6-ary nature, cannot be formed into a reasonable 8-ary set. The (3, 5) circular design is clearly inferior to the (4, 4) design in terms of intercircle versus intracircle distances (poor peak power).

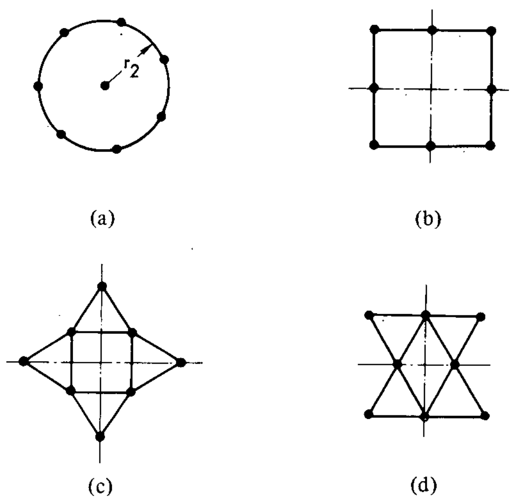

Fig. 2. 8-ary signal set design. (a) (1, 7) circular, $r_1 = 0$, $r_2 = 1.153$. (b) Rectangular. (c) (4, 4) circular, $r_1 = 0.707$, $r_2 = 1.366$. (d) Triangular.

16-Ary Signal Sets: The seven 16-ary candidate sets are shown in Fig. 3 and each of the four design categories are represented. The rectangular set is 4 X 4 square, while the triangular set has alternating rows of three and four signals each, minus the center point. Four circular designs are compared, including an (8, 8) arrangement previously thought to be optimum for both peak and average power constraints [1], [10].

The (5, 11), (4, 12), and (1, 5, 10) circular sets were obtained by applying the $M \Rightarrow \infty$ criterion of Shannon for peak power limited sets; i.e., that the signal distribution be uniform within the peak circle.[2] This can be satisfied approximately with circular sets by making the number of signals per circle proportional to the circle radius. For a two-level set one then has

$$n + 2n = 16$$

$$n = 5.33$$

or, since n must be an integer,

$$n = 5.$$

With 16 signals this leads at once to the (5, 11) design. The (4, 12) design is a perturbation of the (5, 11) which appears attractive because the inner circle can be rotated so that all inner circle signal angles fall half-way between those of the outer circle. The (1, 5, 10) set with one signal point at the origin can be treated as a two-level set with 15 points so that $n + 2n = 15$ or $n = 5$.

The circular sets can be further optimized by evaluating the effect on the error probability of variations in circle radii and rotations of the various circles with respect to each other. Because the additional improvement is small, these optimum sets were not considered in this study.

32-Ary Signal Sets: There are seven candidate configurations for 32-ary signal sets as shown in Fig. 4. These include two circular designs, a rectangular grid, and triangular and hexagonal patterns. Again the circular designs were chosen to approximate the uniform signal distribution. Fig. 4(e) and (f) are unique in that they do not fall into one of the four basic pattern categories.

[2]The (1, 5, 10) and (1, 7) designs were suggested by J. E. Taber of TRW Systems.

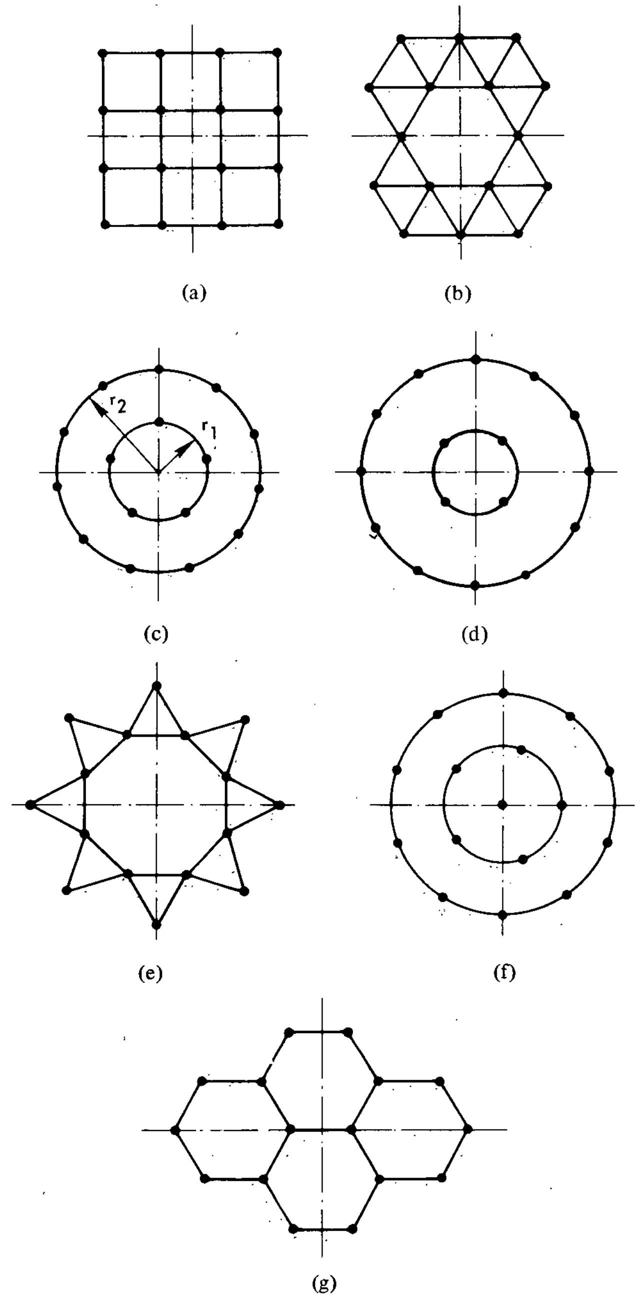

Fig. 3. 16-ary signal set designs. (a) Rectangular. (b) Triangular. (c) (5, 11) circular, $r_1 = 0.853$, $r_2 = 1.72$. (d) (4, 12) circular, $r_1 = 0.707$, $r_2 = 1.93$. (e) (8, 8) circular, $r_1 = 1.305$, $r_2 = 2.17$. (f) (1, 5, 10) circular, $r_1 = 0$, $r_2 = 0.9$, $r_3 = 1.8$. (g) Hexagonal.

The (5, 10, 17) and (4, 10, 18) circular designs are rejected; the former because its minimum distance is much less than the (5, 11, 16) and the latter because its outer radius is too large (poor peak power). The four-circle (3, 6, 10, 13) design was examined and it is apparent that four circles are too many for a 32-ary alphabet because the intercircle distance is too small.

64-Ary Signal Sets: The five candidate 64-ary designs are illustrated in Fig. 5 where only the basic symmetry region is shown. Uniform distribution with circles leads to the (6, 12,

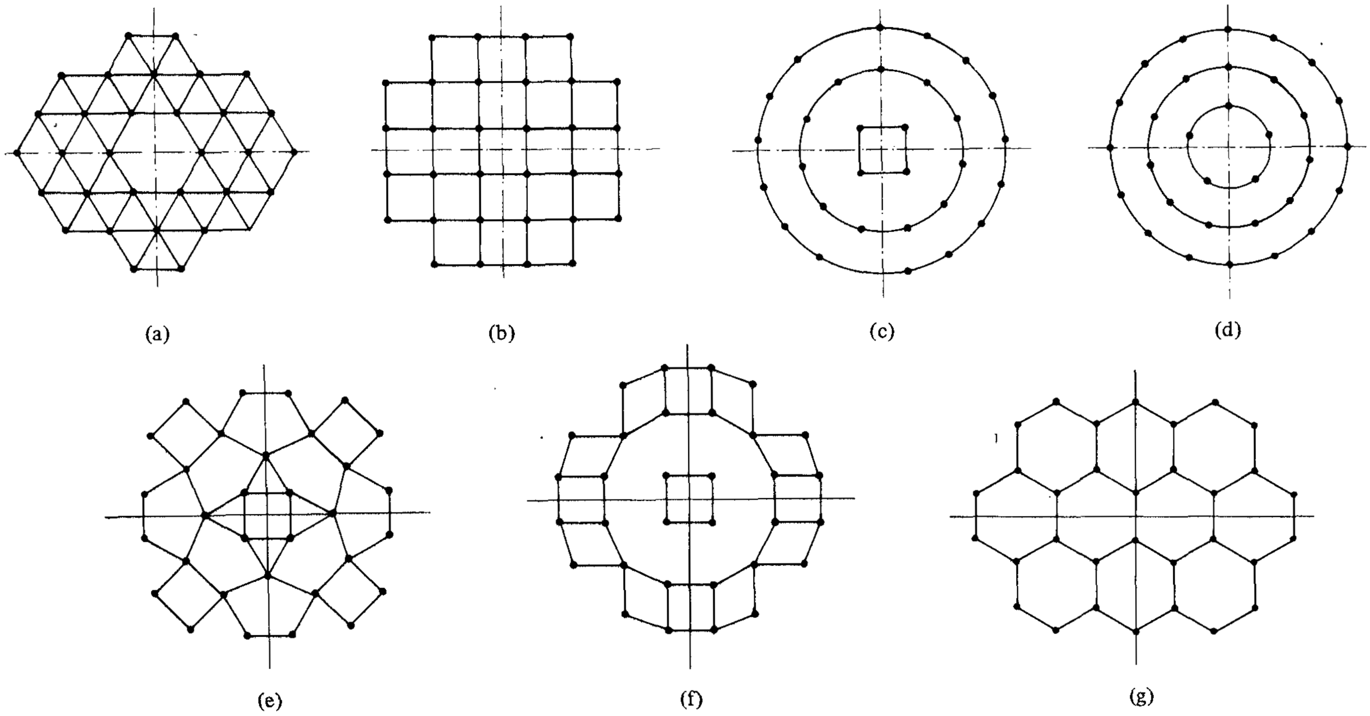

Fig. 4. 32-ary signal set designs. (a) Triangular. (b) Rectangular. (c) (4, 11, 17) circular, $r_1 = 0.707$, $r_2 = 1.72$, $r_3 = 2.68$. (d) (5, 11, 16) circular, $r_1 = 0.853$, $r_2 = 1.72$, $r_3 = 2.56$. (e). (f). (g) Hexagonal.

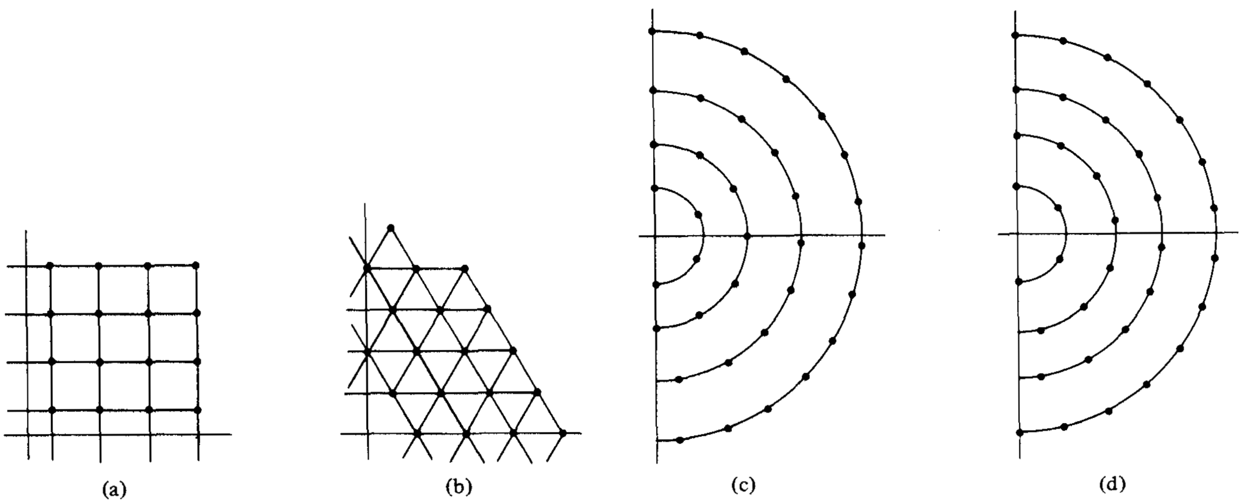

Fig. 5. 64-ary signal set designs. (a) Rectangular. (b) Triangular. (c) (6, 12, 19, 27) circular, $r_1 = 1$, $r_2 = 1.93$, $r_3 = 3.04$, $r_4 = 4.29$. (d) (6, 13, 19, 26) circular, $r_1 = 1$, $r_2 = 2.08$, $r_3 = 3.04$, $r_4 = 4.15$.

19, 27) and (6, 13, 19, 26) designs [Fig. 5(c) and (d)] for four circles. Five circles were found to be too many since the intercircle distance is too small.

128-Ary Signal Sets: The five candidate designs selected for 128-ary signal sets, as illustrated in Fig. 6, are the rectangular, or stepped square, triangular, five-circle (8, 17, 25, 34, 44), and six-circle (6, 12, 18, 24, 30, 38) configurations. The circles approximate a uniform signal distribution. Note that the triangular design has one signal point at the top which does not fit into the triangular pattern.

III. Performance in the Presence of Noise

The symbol error probabilities for the foregoing signal sets will now be compared as functions of peak and average SNR.

A general expression for the probability of symbol error for M-ary signal sets with unequal energies and maximum likelihood detection was apparently first considered by Dunbridge [6]. Using the tetrachoric series expansion approach of Balakrishnan and Weber [13], he obtained an asymptotic expression for the probability of signal detection for high SNR. Unfortunately an integral approximation [13, eq. 14.101 and ff] employed by Weber and Dunbridge caused the expression to diverge for two-dimensional signal sets. R. M. Fielding [14] has employed a somewhat different approach which avoids this approximation and obtains an asymptotic expression for high SNR that is valid for signal sets of any dimensionality. For APK signal sets, Fielding shows that the probability of signal-detection error at a large SNR is asymptotically

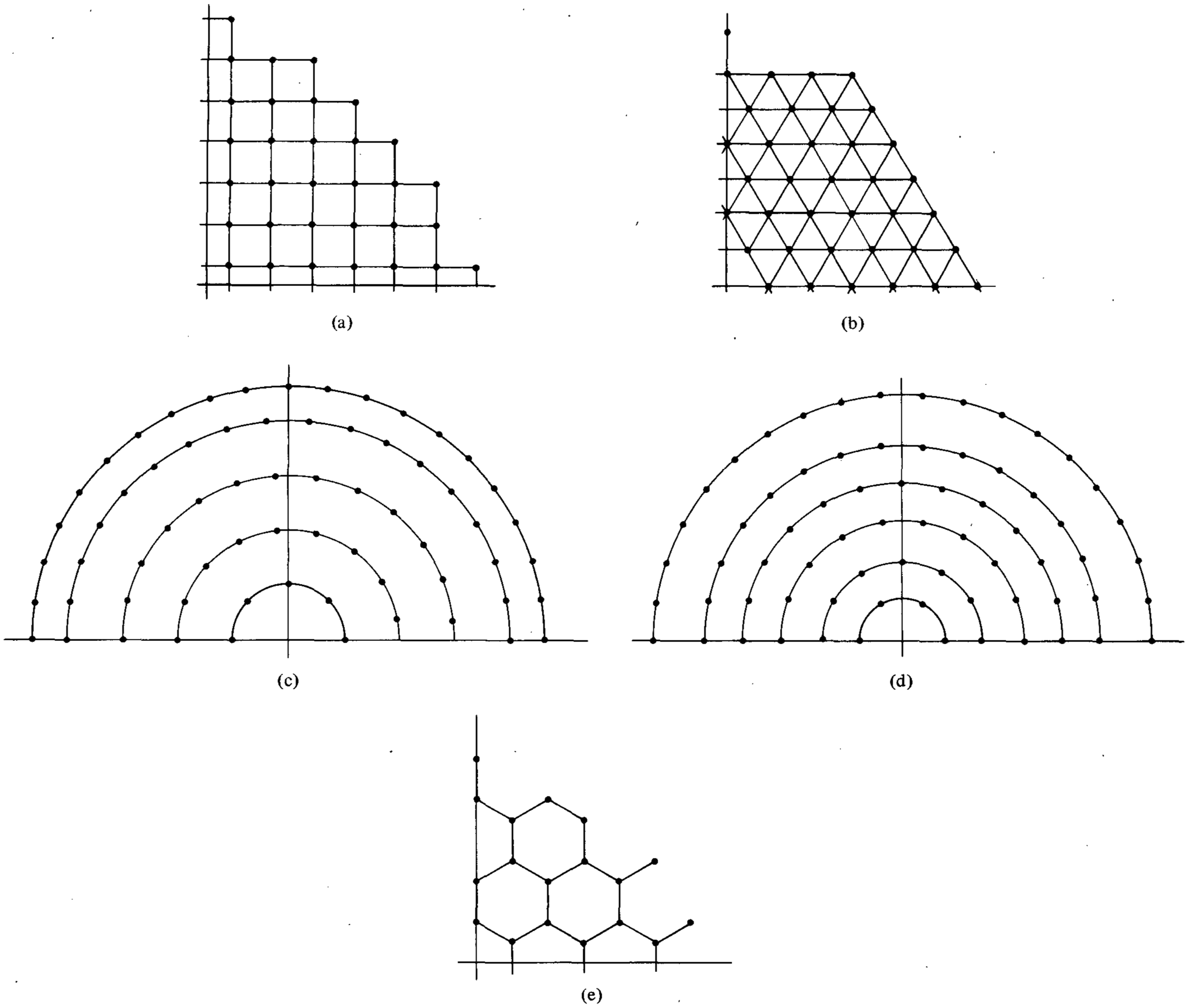

Fig. 6. 128-ary signal set designs. (a) Rectangular. (b) Triangular. (c) (8, 17, 34, 44) circular, $r_1 = 1.31, r_2 = 2.68, r_3 = 3.99, r_4 = 5.41, r_5 = 6.25$. (d) (6, 12, 18, 24, 30, 38) circular, $r_1 = 1, r_2 = 1.93, r_3 = 2.88, r_4 = 3.83, r_5 = 4.78, r_6 = 6.05$. (e) Hexagonal.

$$P_E = \frac{1}{M}\sqrt{\frac{1}{2\pi}} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \frac{\exp(-|S_i - S_j|^2/4\sigma^2)}{|S_i - S_j|/\sqrt{2}\sigma}. \qquad (8)$$

Note that, since P_E can be considered to be the error probability resulting from $M(M-1)$ binary sets made up from every possible distinct signal pair, it is also a union-bound for the error probability. It follows that the union bound is asymptotically tight. For M-ary PSK and rectangular sets, where the exact answers are known, the bound is found to be very tight for $P_E \leqslant 0.01$.

A. Signal Set Comparison

The performance of each signal set in Figs. 1–6 has been evaluated by means of (8). The results are shown in Figs. 7 and 8 as a function of peak and average SNR. The best designs for each M are summarized in Tables I and II, along with their relative advantages over other major configurations.

It should be borne in mind that these performance curves are actually upper bounds. The relative performance of the signal sets based on true error probability could differ from their relative performance bounds.

The performance of the (1, 3) quaternary APK set is shown in Figs. 7(a) and 8(a) along with quadriphase PSK. With only four signals the PSK set is superior to the APK set on the basis of both average and peak SNR. At 10^{-5} symbol error probability PSK offers an advantage of 2.9 dB for peak SNR and 1.6 dB for average SNR.

For $M = 8$, shown in Figs. 7(b) and 8(b), the circular (1, 7) set is the best of the four APK designs for both peak and average SNR. For peak SNR it offers an advantage of 1.2 dB at $P_E = 10^{-5}$ over PSK, which was previously thought to be

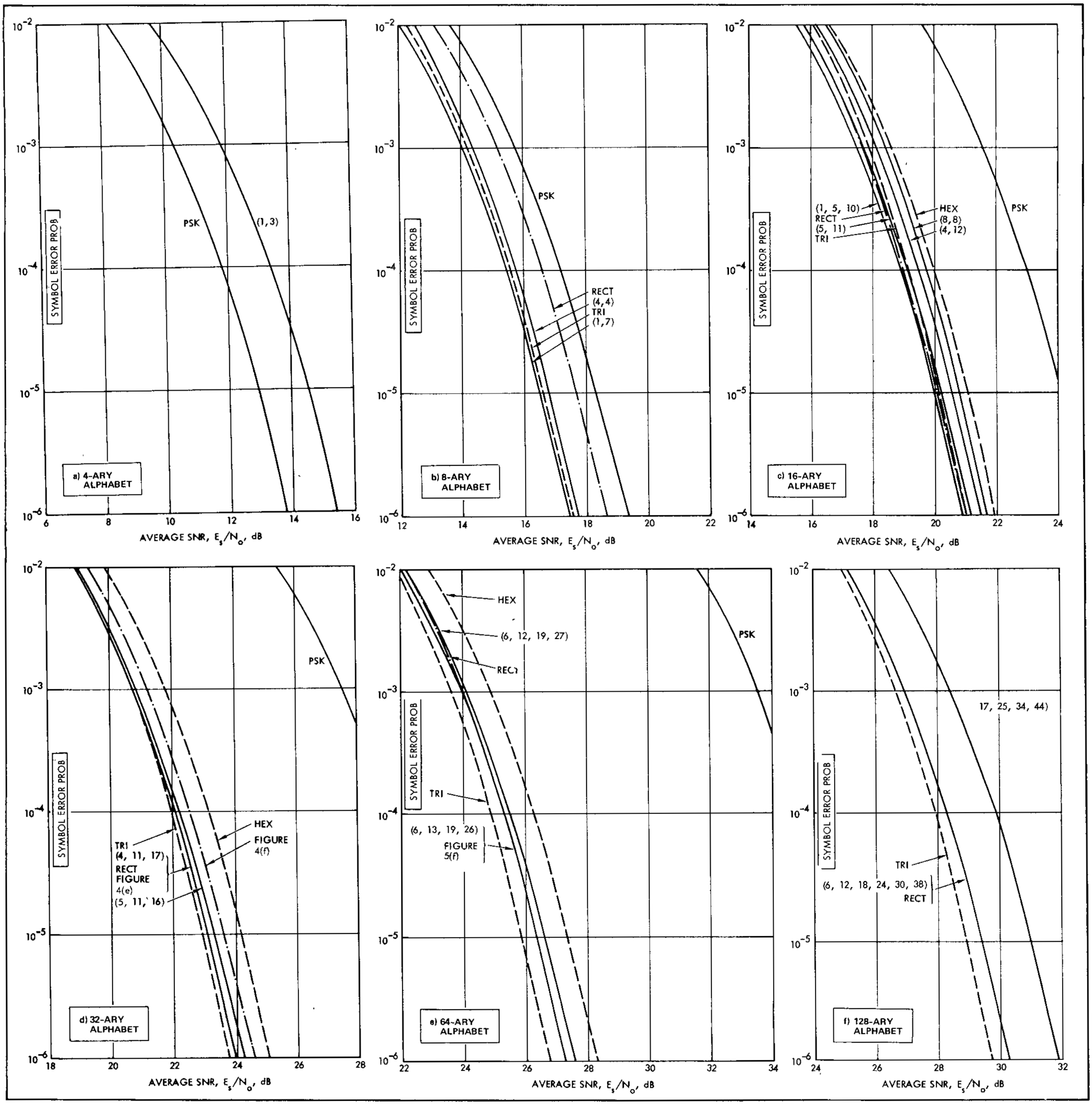

Fig. 7. Symbol error probability versus average symbol SNR.

optimum. It also yields a 0.4 dB advantage in average SNR over the (4, 4) set which had also been presented as an optimum design. The easily generated rectangular design suffers a 1.8 dB disadvantage in both average and peak SNR.

Among the seven 16-ary sets, the circular designs are shown in Figs. 7(c) and 8(c) to be clearly superior for both average and peak SNR performance. The advantage of APK over PSK is 2.7 dB peak SNR and 4.1 dB average SNR, as indicated in Tables I and II.

The performance of the seven 32-ary designs is shown in Figs. 7(d) and 8(d). Of the seven designs, the triangular set attains the best average SNR performance but the rectangular and (4, 11, 17) circular designs are only 0.2 dB inferior. The advantage offered by APK over PSK is approximately 7.1 and 5.2 dB for average and peak SNR. The special sets, which are difficult to generate, show no performance advantage.

For 64-ary and 128-ary signal sets, as for 32-ary sets, the triangular set yields the optimum performance for average SNR [Fig. 7(e) and (f)], with the rectangular and circular designs requiring about 0.5 dB greater SNR. Fig. 8(e) and (f) show again that the circular design offers the best peak SNR performance. The optimum 64-ary APK sets provide a 10.1

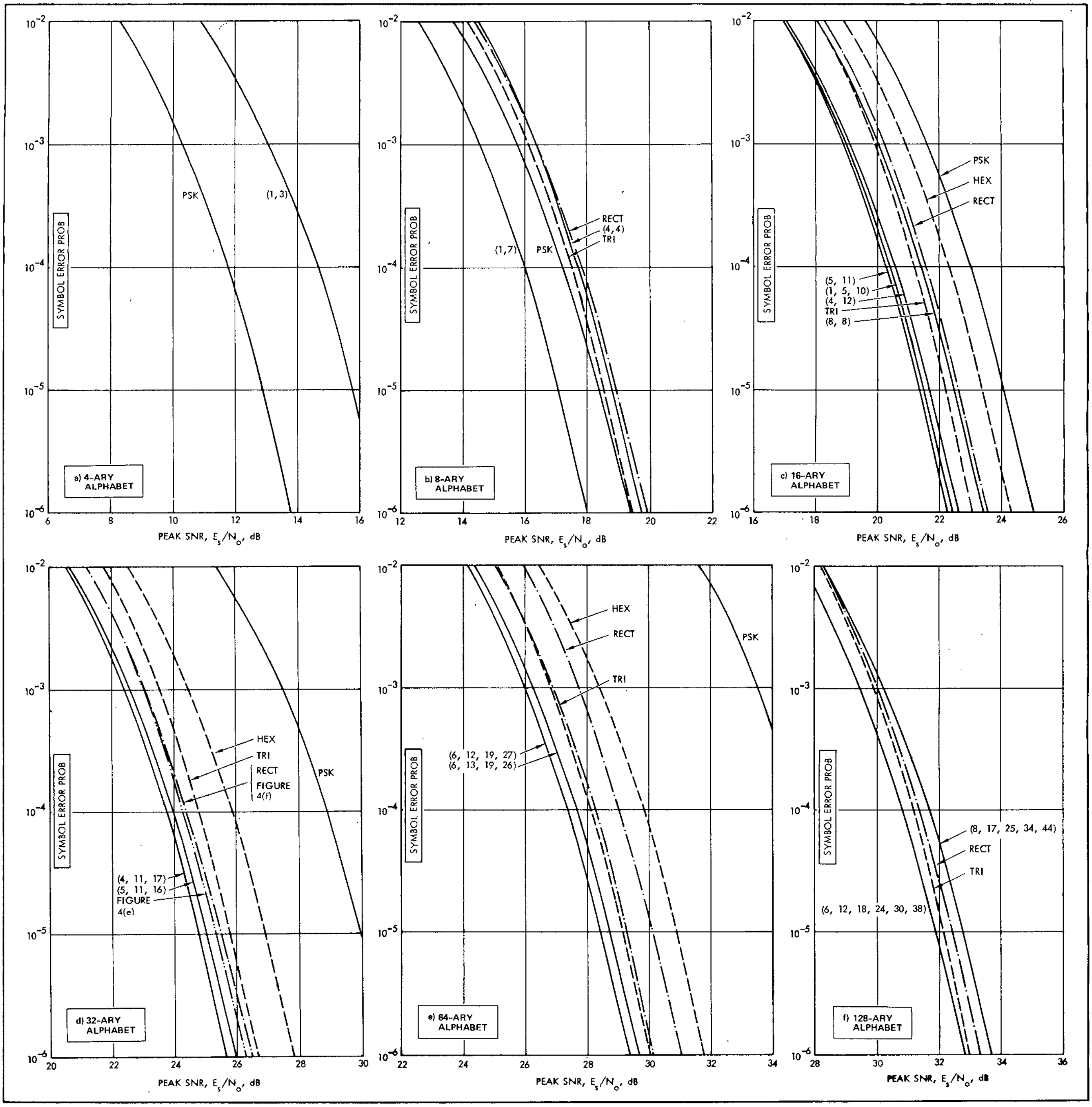

Fig. 8. Symbol error probability versus peak symbol SNR.

dB and 7.6 dB advantage over PSK for average and peak SNR. With an 128-ary alphabet, APK offers an advantage of about 13 dB for an average SNR and 10 dB for peak SNR.

As summarized in Table I it is evident that for $M \geqslant 32$ the triangular design yields the best performance on an average SNR basis. This is to be expected because of its optimum sphere-packing properties. However the more easily generated rectangular design is within 0.5 dB of the best set for $M \geqslant 16$. Table II shows that for peak SNR the circular designs are superior for all M's, with the rectangular sets suffering a penalty of 0.5 dB to 1.7 dB.

IV. TWT Amplifier Effects

When an APK signal is amplified by a TWT in a single carrier mode, the AM-AM and AM-PM nonlinearities of the tube distort the signal set and introduce performance degradation. Since the distortion can be reduced by backing-off from the saturated drive level of the tube, there is a tradeoff between signal quality and operating power. That tradeoff has been investigated to determine the best TWT performance when both distortion and backoff are considered.

To evaluate the degradation caused by the TWT, other link components must also be specified and modeled analytically.

TABLE I
COMPARISON OF PRINCIPAL APK SETS FOR AVERAGE SNR

Alphabet Size	Best Set	Other Sets	Penalty With Respect to Best Set ($P_E = 10^{-5}$) (dB)
4	PSK		
		(1,3)	1.6
8	(1,7)		
		Triangular	0.1
		Rectangular	1.2
		PSK	1.8
16	(1,5,10)		
		Rectangular	0.1
		Triangular	0.2
		PSK	4.1
32	Triangular		
		Rectangular	0.2
		(4,11,17)	0.2
		PSK	7.1
64	Triangular		
		Rectangular	0.4
		(6,13,19,26)	0.4
		PSK	10.1
128	Triangular		
		Rectangular	0.5
		(6,12,18,24, 30,38)	0.5
		PSK	13.1

TABLE II
COMPARISON OF PRINCIPAL APK SETS FOR PEAK SNR

Alphabet Size	Best Set	Other Sets	Penalty With Respect to Best Set ($P_E = 10^{-5}$) (dB)
4	PSK		
		(1,3)	2.9
8	(1,7)		
		PSK	1.2
		Rectangular	1.7
16	(5,11)		
		Triangular	0.8
		Rectangular	1.2
		PSK	2.7
32	(4,11,17)		
		Rectangular	0.5
		Triangular	1
		PSK	5.2
64	(6,13,19,26)		
		Triangular	0.7
		Rectangular	1.7
		PSK	7.6
128	(6,12,18,24, 30,38)		
		Triangular	0.2
		Rectangular	0.5
		PSK	10.1

For example the demodulator, which is a maximum likelihood detector for the undistorted signal, is no longer optimum after the TWT distortion. This difficulty may be overcome by altering the demodulator structure so that it becomes a maximum likelihood detector for the distorted signal set; as an alternative the transmitted signal set can be predistorted so that after passing through the TWT it is undistorted and matched to the demodulator. In addition the demodulator must reconstruct references for carrier amplitude and phase. Due to the TWT distortion the form of phase and amplitude reconstruction can have a significant effect on the error performance. The following paragraphs discuss the models employed for the TWT and demodulator reference reconstruction. Performance curves for 16-ary and 64-ary APK signal sets are then presented as a function of TWT backoff. Their performance is compared to PSK signals of the same alphabet size operating at saturation.

A. Link Model

The total communication link pictured in Fig. 9 contains a TWT, modem, and a predistortion circuit. The predistortion circuit may or may not be included in the link. It compensates for the TWT's amplitude compression by expanding the outer (large amplitude) set members with respect to the inner (small amplitude) members. Similarly it compensates for the phase rotation due to AM–PM distortion by an opposite rotation of the outer signals with respect to the inner ones. The predistortion is assumed to be ideal in that it completely compensates for the TWT distortion.

TWT Model: The nonlinear effects in the TWT are modeled as being independent cascaded elements, as shown in Fig. 9. Clearly, the AM–PM effect must precede the AM–AM in a series model because AM–PM conversion is present at saturation.

Since the input and output of the TWT for a steady-state measurement can be represented by complex numbers, it becomes convenient to think of the TWT transfer function as a complex attenuation defined by the ratio of the input and output that is solely a function of the magnitude of the excitation; that is

$$\mathrm{TWT}\,(r_i) = \frac{r_0}{r_i}\, e^{j(\phi_0 - \phi_i)}. \tag{9}$$

The contribution to this transfer function of both nonlinearities must now be determined.

The amplitude nonlinearity can be analytically modeled by fitting the single-carrier input/output power data with a portion of a cosine centered about saturation. This leads to the simple functional representation

$$r_0 = \begin{cases} 10^{\alpha[\cos(\log_{10}(r_i/\hat{r}_i)/\beta - 1]}, & r_i > \tilde{r}_i \\ r_i, & r_i \leqslant \tilde{r}_i \end{cases} \tag{10}$$

where $\hat{r}_i$ is the input voltage level corresponding to saturation and α and β are constants chosen to fit the data. The resulting power curve is shown in Fig. 10 for coefficients matched to measured data for the TRW DSCS-II satellite TWT. The maximum error of the analytical model is less than 0.3 dB.

For the AM–PM transfer function, the Berman and Mahle [15] model is used. It has the form

$$\Delta\phi = k_1(1 - \exp(-k_2 r_i^2)) + k_3 r_i^2 \tag{11}$$

where the k_2 and k_3 coefficients have been modified from Berman and Mahle's closest computed set ($K_2 = 3.54, K_4 =$

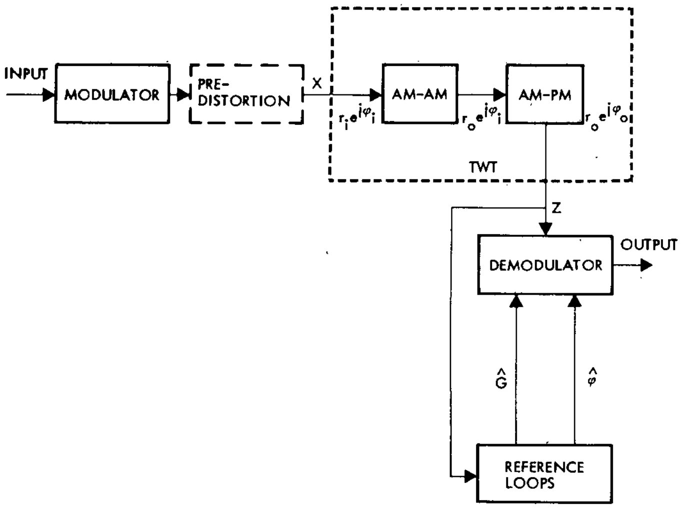

Fig. 9. TWT link model.

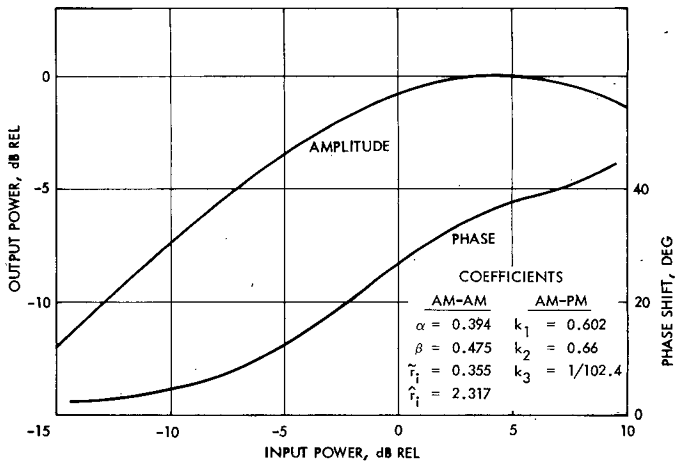

Fig. 10. Analytic model for TWT.

0.0524) to account for operations directly upon the magnitude of the complex voltage envelope and the normalization of the TWT gain transfer function to unity small signal gain with 0-dB saturated output. The AM–PM transfer function is also plotted in Fig. 10.

Reference Reconstruction: In order to properly demodulate the received signal, the demodulator must reconstruct carrier phase and amplitude references and symbol sync timing. This reconstruction must be performed on the distorted signal and in the presence of receiver noise. Fortunately, the latter can be neglected because the reference tracking loops will be narrow compared to the symbol rate so that their SNR is even higher than the 15–30 dB symbol SNR's of interest.

A simple and effective method of carrier phase reconstruction is to estimate either the long term average of the incoming waveform phase or the average phase of the distorted signal set at the TWT output. This corresponds to a narrowband phase-lock loop that settles during the preamble, uses a special averaging preamble, or tracks a carrier that has burst-to-burst coherence. Letting $Z_i(t)$ be the ith TWT output signal, the phase reference $\hat{\phi}$ is

$$\hat{\phi} = \frac{1}{M} \sum_{i=1}^{M} \operatorname{Im}\left\{\log_e\left(\int_0^T Z_i(t)dt\right)\right\}. \tag{12}$$

In the demodulator, then, each signal $Z_i(t)$ is phase compensated by multiplying it by $e^{-j\hat{\phi}}$.

The amplitude of the carrier is compensated for the TWT effect by adjusting the amplitude of the peak input signal to match the amplitude of the peak reference signal at the demodulator. The receiver gain is therefore

$$\hat{G} = \frac{|S_{pk}|}{|Z_{pk}|} \tag{13}$$

where Z_{pk} and S_{pk} are the peak received and transmitted signals, respectively.

B. Performance

The performance of APK transmission through a TWT is illustrated with 16-ary and 64-ary circular signal sets in Figs. 11 and 12. These curves assess the tradeoff between degradation due to TWT distortion near saturation and the power loss (penalty) incurred if distortion is reduced through power backoff. Since the TWT beam current is determined by the saturated power level, the fact that the APK signal may not utilize all of this available power is of no advantage in the power consumption requirements of the TWT. Thus, the tradeoff between distortion and backoff penalty can be displayed by evaluating the symbol error performance in terms of the saturated, or available, downlink SNR, which is obtained by multiplying the peak symbol SNR by the backoff.[3] By plotting such curves for various backoff levels, the optimum backoff can be found and the resulting performance conveniently compared with PSK operating at saturation.

For the 16-ary circular set the error probability was calculated for backoff levels of 0 dB, 3 dB, and 6 dB. Fig. 11 shows that 3-dB backoff is near optimum. (The 0-dB backoff curve was not plotted because of its extreme degradation.) The most significant feature is the poor performance of APK relative to PSK, which *can* operate at 0-dB backoff. The 2.7-dB advantage of APK in a linear channel has become a 3.7-dB disadvantage.

When a 64-ary circular set is employed, PSK is still superior by 0.6 dB, as shown in Fig. 12. Note that here the required backoff is somewhat above 6 dB, so that the TWT distortion for a high alphabet set must be kept small. In fact, at 6-dB backoff the average phase shift for this signal set due to AM–PM is only 3.75°, while the peak signals are compressed only 0.6 dB.

The "ideal APK" curve in both figures is the performance of APK in a linear channel. By employing a predistortion compensation of the transmitted signal set that is perfectly matched to the TWT characteristic at 0-dB backoff, this ideal performance could be obtained.

It is apparent that APK[4] suffers a severe performance degradation in single-carrier operation through a typical uncompensated TWT. If APK is to be employed in the single-carrier

[3] Backoff is defined as a number greater than unity.
[4] The (1, 7) design is an obvious exception as it suffers no TWT degradation relative to PSK since it requires no backoff.

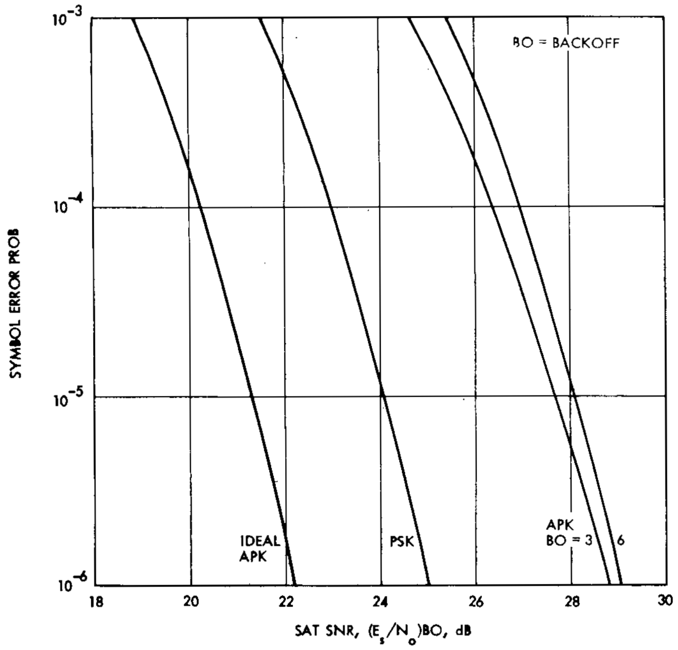

Fig. 11. Performance of (5, 11) circular set transmitted through a TWT.

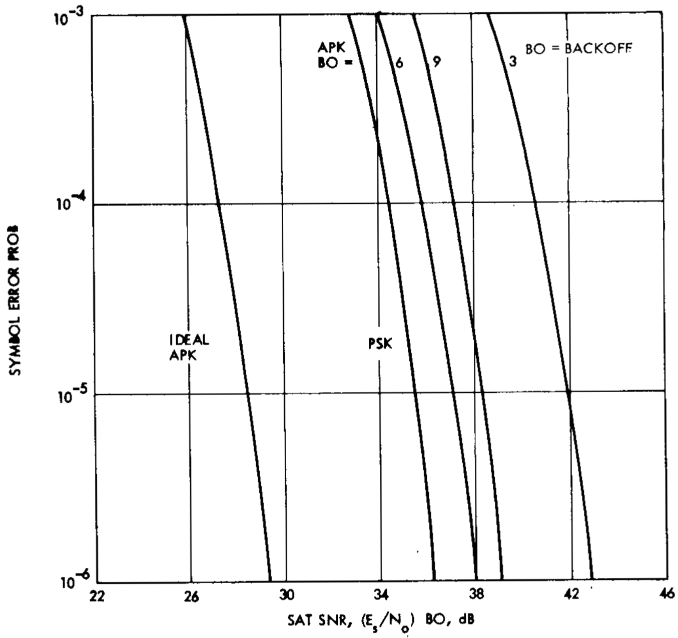

Fig. 12. Performance of (6, 13, 19, 26) circular set transmitted through a TWT.

mode, the TWT must be linearized or the signal set must be predistorted.

The performance of APK in multi-carrier operation was not investigated but the relative performance of APK with respect to PSK will be much improved in this mode. The improvement results because the PSK carriers require a backoff of 4 to 6 dB to achieve a low level of intermodulation products. APK, on the other hand, already operates at large backoffs for low distortion so that the intermodulation products are small, and little if any additional backoff is required to make them negligible. For example, Fig. 12 shows that PSK with 5-dB backoff is 3 dB worse than APK with 6-dB backoff.

V. Modem Implementation

Because APK sets are two-dimensional, APK modems are significantly different from the more familiar one-dimensional PSK or ASK modem. Basic considerations in the construction of APK modulators, demodulators, and phase reference reconstruction loops will now be discussed. More specific details are found in [4].

Since APK signals can be represented in either Cartesian or polar coordinates, modems can be constructed based on either representation. In general, Cartesian coordinates result in the simplest modem design for rectangular sets, as do polar coordinates for circular sets. For other types of signal sets, however, both coordinate configurations should be examined.

A. Modulator

In a Cartesian coordinate modulator, the amplitudes of the two orthogonal (X, Y) coordinates are varied in accordance with the data using attenuators. Campopiano and Glazer [2] illustrate such a modulator for the rectangular sets. Its simplicity is a major advantage for rectangular sets.

A polar modulator varies the amplitude and phase of the transmitted carrier in accordance with the data. Such a modulator for the (1, 5, 10) set would consist of five phase shifters, an attenuator, and appropriate switches. (A PSK modulator would have $M/4$ phase shifters with two quadrature channels.)

Clearly any arbitrary signal set can be generated with either type of modulator but one may be less complex than the other. For example, a (1, 5, 10) modulator in Cartesian coordinates requires 9 attenuator settings in one direction and 11 in the other; these 20 attenuators must be compared in cost to 5 phase shifters and 1 attenuator.

B. Demodulator

The optimum demodulator for APK utilizes maximum likelihood detection (MLD) in which the Euclidian distances between the incoming signal and all members of the (stored) signal set are calculated. The closest member is determined to be the transmitted symbol. Thus if Z represents the input signal-plus-noise, the MLD selects $\hat{S} = S_k$, where

$$S_k = \min_i |Z - S_i|^2, \quad i = 1, \cdots, M \tag{14a}$$

$$= \max_i \, [2\mathrm{Re}\,\{ZS_i^*\} - |S_i|^2]. \tag{14b}$$

This operation can be interpreted in two different but equivalent ways. The first emphasizes (14b) and forms the dot product by cross correlation or matched filtering of the incoming signal. The second interpretation emphasizes the distance aspect of (14a) and forms decision regions about each signal set member; the identification of the region containing the incoming signal is the detection process. Demodulator implementations can be based on either interpretation and the most suitable depends upon the particular signal set to be used. The first is termed the "matched filter demodulator" and the second the "decision boundary demodulator."

Consider first the matched filter demodulator with a Cartesian coordinate implementation. If the received vector Z is represented as $Z = I + jQ$, the decision variable is

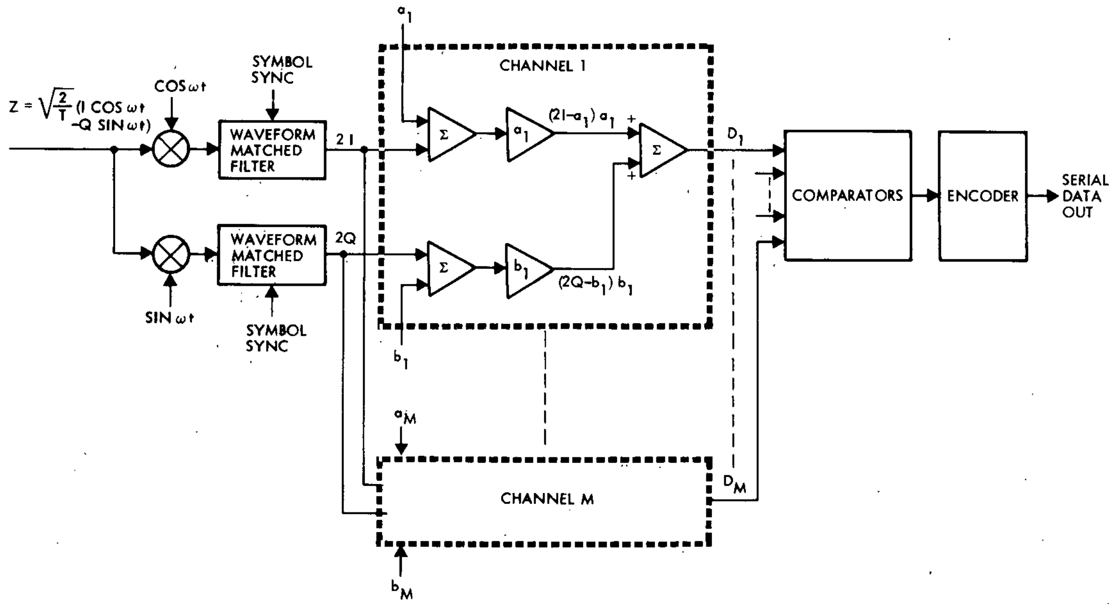

Fig. 13. M-ary APK matched filter demodulator.

$$D_i = 2\,\mathrm{Re}\,\{ZS_i^*\} - |S_i|^2$$
$$= 2(a_iI + b_iQ) - (a_i^2 + b_i^2), \qquad i = 1, \cdots, M.$$

The resulting demodulator is shown in Fig. 13. If the demodulator is to be implemented digitally, this formulation is particularly attractive.

In polar form, let

$$Z(t) = \sqrt{\frac{2}{T}}\,R\cos(\omega t + \theta) \tag{15a}$$

and

$$S_i(t) = \sqrt{\frac{2}{T}}\,A_i\cos(\omega t + \theta_i). \tag{15b}$$

The decision variable becomes

$$D_i = 2A_iR\cos(\theta - \theta_i) - A_i^2 \tag{16}$$

and leads to a demodulator structure similar to Fig. 13 with the addition of phase detectors.

The decision boundary demodulator compares the received signal-plus-noise with a pre-established set of boundaries to identify the decision region to which the signal belongs. The optimum (MLD) decision regions are polygons formed by the perpendicular bisection of lines joining nearest neighbors of the signal set. For some types of signal sets the decision boundary demodulator with optimum decision regions leads to a simpler implementation than does the matched filter demodulator. For others, the optimum regions can be approximated by ray and arc boundaries which have simple realizations.

The simplest decision boundary demodulator in Cartesian coordinates results with the 16-ary and 64-ary rectangular signal sets. The decision regions are squares centered about the signal members. They are obtained by simply quantizing the inphase and quadrature components of the received signal Z. For $M = 32$ and 128, more logic is required to handle the signal members along the set edge unless simplifying boundary approximations are employed. Campopiano and Glazer [2] discuss this type of modulator.

In general, the decision boundaries for circle sets are quite complex. However, by approximating the decision regions in terms of rays and arcs, the incoming signal is simply quantized into amplitude and phase to determine the region to which it belongs. This type of approximate demodulator was considered by Hancock and Lucky [16].

M-ary PSK demodulators are discussed in Ma, Stone, and Sullivan [17]. They are typically of the decision boundary type and require $M/2$ boundaries. Thus for 16-ary PSK, eight boundaries are needed. In comparison only six boundaries are required for the rectangular set and seven for the (1, 5, 10) circular set.

C. Carrier Phase Reconstruction

As in coherent PSK, a phase reference is required for coherent APK demodulation. However the frequency multiply-and-divide technique [17] often employed for data removal in PSK demodulators is not appropriate for APK and a decision-directed remodulation technique is suggested instead.

Frequency multiplication and division is suited to PSK because the phase angles are simple fractions of 360°, i.e., $360/M$. For an alphabet of size M, a multiplication of M is required for PSK. By way of contrast, consider the APK 16-ary rectangular set with phase angles of ±16.7°, ±45°, ±73.3°, 180 ± 16.7°, 180 ± 45°, and 180 ± 73.3°. The multiplication factor F is found from the relation

$$16.7\,F = 360\,n$$

where n is an integer; thus F must be 3600. Even approximating the 16.7° by 17.1° leads to a multiplication factor of 21—clearly excessive.

In place of the multiply-and-divide technique, the Costas loop could be extended to estimate inphase and quadrature component levels and so calculate the phase [4]. However remodulation is a more promising technique. This is a decision-directed phase reconstruction technique that remodulates a local carrier with the demodulator's estimate of the received signal so as to reconstruct the transmitted signal. The reconstructed signal is phase-compared with the received signal and the VCO phase adjusted to minimize the resulting error.

As shown in Fig. 14 the incoming signal is the jth member of the set, $S_j = A_j \cos(\omega t + \theta_j)$. The APK demodulator makes a decision that it has received the signal $\hat{S}_i$, where $i \neq j$ means a data error has occurred. This estimate of the received set member is fed to an APK modulator that is identical to the transmitter's and forms S_m, an RF version of $\hat{S}_i$, which is identical to the transmitted version (for $i = j$), except for amplitude and phase errors. That is

$$S_m = A_j (1 + g) \cos(\omega t + \theta_j + \Phi) \qquad (17)$$

where g and Φ are the errors when $i = j$.

The incoming signal S_j is delayed long enough to compensate for the processing time of the demodulator and is then phase-compared with S_m. The error produced by the phase detector is

$$\epsilon = A_j A_i (1 + g) \sin(\theta_j - \theta_i - \Phi)$$

yielding

$$\epsilon = A_j^2 (1 + g) \sin \Phi \qquad (18)$$

when no data errors are made by the demodulator, i.e., when $\theta_j = \theta_i$.

The carrier tracking loop drives the VCO phase to minimize Φ and thereby minimize ϵ. Note that the phase error ϵ is amplitude modulated at the symbol rate by A_j^2. It is not necessary to remove this if the loop bandwidth is small because the loop gain can be based on the long term average of A_j^2.

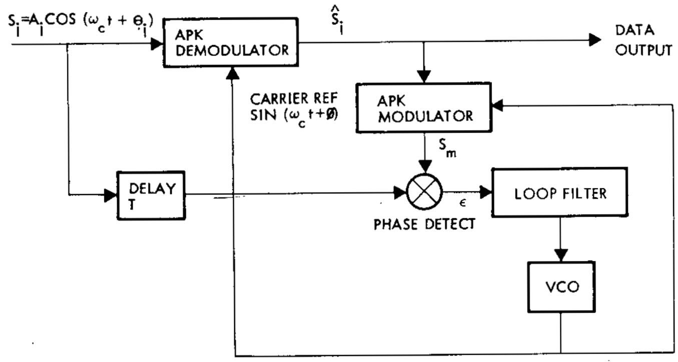

Fig. 14. APK carrier reconstruction remodulation technique.

VI. Conclusion

By means of an error probability bound, APK signal sets have been compared in an empirical search for the optimum design as a function of alphabet size. For all (binary) alphabet sizes greater than four, new designs are presented that outperform previously proposed sets on the basis of both peak SNR and average SNR. For $M \geqslant 8$, APK offers an advantage in average and peak SNR relative to PSK that increases with alphabet size.

When employed in a single-carrier mode with a typical nonlinear TWT amplifier, most APK designs lose their SNR advantage over PSK unless the signal set is predistorted at the transmitter. However in the multi-carrier mode where PSK requires a 4- to 6-dB backoff, APK may retain its superiority.

Modems for certain APK sets are not significantly more complex than those for PSK with the same alphabet size, especially when approximations are incorporated into the decision boundary demodulator.

Acknowledgment

The authors would like to thank G. R. Welti of COMSAT Laboratories for his guidance in formulating the problem and suggesting many of the signal designs. Dr. C. L. May of TRW Systems helped develop and run the computer program used in this analysis, and L. N. Ma of TRW Systems was responsible for the detailed modem design performed during the study.

References

[1] R. W. Lucky and J. C. Hancock, "On the optimum performance of N-ary systems having two degrees of freedom," *IRE Trans. Commun. Syst.*, vol. CS-10, pp. 185–192, June 1962.
[2] C. N. Campopiano and B. G. Glazer, "A coherent digital amplitude and phase modulation scheme," *IRE Trans. Commun. Syst.*, vol. CS-10, pp. 90–95, Mar. 1962.
[3] C. R. Cahn, "Combined digital phase and amplitude modulation communication systems," *IRE Trans. Commun. Syst.*, vol. CS-8, pp. 150–155, Sept. 1960.
[4] C. M. Thomas, "Final report on amplitude-phase keying techniques," prepared by TRW Systems for COMSAT Laboratories as Manager of INTELSAT under Contract CSC-SS-292, Aug. 1971.
[5] M. Y. Weidner, "Degraded error probability for multiphase digital signalling with linear distortion and Gaussian noise," *Proc. INTELSAT/Inst. Elec. Eng. Int. Conf. Digital Satellite Communication*, London, pp. 202–212, Nov. 1969.
[6] B. Dunbridge, "Asymmetric signal design for the coherent Gaussian channel," *IEEE Trans. Inform. Theory*, vol. IT-13, pp. 422–431, July 1967.
[7] C. Weber, "New solutions to the signal design problem for coherent channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 161–167, Apr. 1966.
[8] A. V. Balakrishnan, "A contribution to the sphere packing problem of communication theory," *J. Math. Anal. Appl.*, vol. 3, pp. 485–506, Dec. 1961.
[9] J. Salz *et al.*, "Data transmission by combined AM and PM," *Bell Syst. Tech. J.*, vol. 50, pp. 2399–2419, Sept. 1971.
[10] K. Kawai *et al.*, "Optimum combinations of amplitude and phase modulation scheme and its application to data transmission modem," *Proc. Int. Communication Conf.*, pp. 29-6–29-11, 1972.
[11] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July and Oct. 1948.
[12] N. M. Blachman, "A comparison of the informational capacities of amplitude- and phase-modulation communication systems," *Proc. IRE*, vol. 41, pp. 748–759, June 1953.
[13] C. L. Weber, *Elements of Detection and Signal Design*. New York: McGraw-Hill, 1968, pp. 182–183.
[14] R. M. Fielding, [4, appendix C].
[15] A. L. Berman and C. E. Mahle, "Nonlinear phase shift in traveling-wave tubes as applied to multiple access communications satellites," *IEEE Trans. Commun. Technol.*, vol. COM-18, pp. 37–48, Feb. 1970.
[16] J. C. Hancock and R. W. Lucky, "Performance of combined amplitude and phase-modulated communication systems," *IRE Trans. Commun. Syst.*, vol. CS-8, pp. 232–237, Dec. 1960.
[17] L. N. Ma, M. S. Stone, and D. P. Sullivan, "Synthesis of high data rate coherent telemetry systems," in *Int. Telemetering Conf.*, Los Angeles, Calif., vol. IV, pp. 440–454, 1968.

C. Melvil Thomas (M'61) was born in Glendale, Calif. on March 7, 1935. He received the B.S.E.E. degree from Texas University, Austin, in 1957, the M.S.E.E. degree from Stanford University, Stanford, Calif.,

in 1958, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1967.

From 1958 to 1961 he was employed by General Dynamics, Pomona, Calif., in the design of missile guidance receivers. In 1961 he joined Lear Siegler Electronics, Anaheim, Calif., to design phase-lock receivers. Since 1963 he has been with TRW Systems, Redondo Beach, Calif., where he is now a Member of the Senior Research Staff. At TRW he has been engaged in the analysis and development of advanced communication systems for military and commercial applications throughout the frequency region from ELF to EHF.

Michael Y. Weidner (S'66–M'67) was born in St. Louis, Mo., on May 14, 1944. He received the B.S.E.E., and M.S.E.E. degrees from the Massachusetts Institute of Technology, Cambridge, in 1966 and 1967, respectively.

He has worked for TRW Systems, Redondo Beach, Calif., since 1967. As Head of the Satellite Communications Section he was responsible for the performance analysis of digital communication links for satellite applications. This group concentrated on computer evaluation of degradation due to nonideal distortion mechanisms. Currently, he is a Senior Staff Engineer for communication systems, responsible for the conceptual design and development of digital satellite communication systems.

S. H. Durrani (M'57–SM'62) was born in Pakistan in 1928. He received the B.A. degree in mathematics and the B.Sc. (Hons.) degree in electrical engineering from the University of the Punjab, Lahore College of Engineering and Technology, Lahore, Pakistan, in 1946 and 1949, respectively, the Master's degree from the University of Manchester, Manchester, England, in 1953, and the Ph.D. degree from the University of New Mexico, Albuquerque, in 1962.

He taught at the Lahore College of Engineering and Technology from 1949 to 1951. From 1962 to 1964 he was employed by General Electric Company. He served as Professor and Chairman of the Department of Electrical Engineering, Lahore College of Engineering and Technology, from 1964 to 1965. For a year after that he was an Associate Professor at Kansas State University, Manhattan. He worked at the Astro Electronics Division of RCA, Princeton, N.J., from 1966 to 1968. For the next four years he was with COMSAT, where he was a Branch Manager in the Systems Analysis Lab and a Staff Scientist in the Advanced Studies Lab. He served as a visiting Professor at the University of Maryland for Fall 1972. Currently he is a Senior Scientist/Project Manager at Operations Research, Inc. He has published numerous technical articles.

Dr. Durrani has been active in the Princeton and Washington Sections of the IEEE. He is a member of the Educational Activities Board and Chairman of the Region 2 Professional Activities Committee. He is an Associate Editor of The *IEEE Transactions on Aerospace and Electronics Systems.* He is also a member of the International Scientific Radio Union Commission 2 and an Associate Fellow of the American Institute of Aeronautics and Astronautics. He has organized and chaired technical sessions of topics related to space communications at several conferences sponsered by IEEE, URSI, and AIAA.

Negatively Correlated Branches in Frequency Diversity Systems to Overcome Multipath Fading

FRED HABER, SENIOR MEMBER, IEEE, AND MAHMOOD R. NOORCHASHM, STUDENT MEMBER, IEEE

Abstract—In multipath environments in which there is a compact concentration of the values of time-delay-difference (TDD) between signals arriving over each of the indirect paths and the direct path, it is possible, by suitably choosing frequency differences, to design a frequency diversity system in which the diversity branches are negatively correlated. This results in a reduced probability of simultaneous deep fades on all branches over that obtained with branches designed to be independent.

In the first part of the paper the magnitude of the negative correlation is shown to depend on the number of indirect reflectors contributing to the total indirect signal and that it is maximum when the indirect signal arises from a single reflector. Also shown is how the magnitude of the negative correlation can be maximized using the probability density function of the TDD to establish the frequency assignment. This is a precursor to the second part of the paper in which an investigation of the improvement of the error probability by a judicious choice of frequency differences is reported.

The error probability as a function of frequency differences is found for the specific case of noncoherent frequency-shift keying (FSK) in a specular multipath channel using square-law combining. A general result for Mth order diversity is obtained but a detailed study is made for two and three order diversity only. Results are presented for uniformly distributed TDD's with various ranges and mean values, the parameters being suggested by the physical conditions encountered when low flying aircraft are the multipath source or are themselves trying to communicate. Results for other densities of TDD, the normal and the gamma, are cited. For the second order case, it is found that an advantageous frequency spacing can be determined in terms of the average TDD. For the third order case, two frequency differences are available

Paper approved by the Associate Editor for Communications Theory of the IEEE Communications Society for publication without oral presentation. Manuscript received February 8, 1973; revised July 9, 1973. The work reported here was done as part of the Pennsylvania–Princeton Avionics Research Program for the United States Army Electronics Command of Fort Monmouth, N.J. under Contract DA 28-043-AMC-02411 (E).

The authors are with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa.