

The signal duration is $T = m\tau$, and by (2) m is proportional to R . Therefore, we define a *reduced signal-to-noise ratio* d_0 by

$$d_0^2 = \frac{1}{RN} \sum_i p_i \sum_j r_i^{(j)} d_i^2. \quad (44)$$

Systems will now be compared by plotting the quantity $2^{-U/RN}$ vs d_0 .

Using the convexity of the function $\log x$, we can derive the inequality

$$U \geq -\log \bar{Q} \quad (\text{or}) \quad \bar{Q} \geq 2^{-U}, \quad (45)$$

where

$$\bar{Q} = \sum_i p_i Q_i$$

is the average probability of correct reception of the letters of the alphabet. Equality holds when all signals have equal probabilities of correct reception ($Q_i = Q$). We then have $U = -\log Q$ and $2^{-U/RN} = Q_e$, and we are back to the equal-bandwidth comparison derived earlier. The source possesses maximum entropy¹ when all signals are sent equally often: $p_i = 1/L$, and we then have $H = \log L$ and $\log L_e = (\log L)/RN$ as before. Our reduced signal-to-noise ratio becomes that defined in (8) when all transmitted signals have the same energy.

If one follows a similar line of reasoning, it is not hard to show that for the equal-rate comparison of a number

of systems one calculates for each the quantity $2^{-U/H}$ for the same set of values of an equivalent signal-to-noise ratio d_e given now by

$$d_e^2 = \frac{1}{H} \sum_i p_i \sum_j r_i^{(j)} d_i^2. \quad (46)$$

The bandwidth W required when the signal durations are adjusted for this comparison is proportional to $RN/H = 1/\log L_e$. Again the equal-bandwidth and the equal-rate comparisons rank strictly comparable systems in the same order of preference.

In this formulation the performance of a system depends upon the quantity U of (42); the smaller U the better. One would like to minimize U for each given system by varying the probabilities p_i and, possibly, the distribution of power among the several signals, as well perhaps as even the decision rules at the receiver. These variations will be subject to such constraints as a fixed average transmitter power and a fixed entropy H . Consideration of the problems that arise here lies outside the scope of this paper.

VI. ACKNOWLEDGMENT

We wish to thank E. Scherer, Dr. R. Filipowsky and others of the Communications Group, Advanced Development, Westinghouse Electronics Division, Baltimore, Md., whose discussions of communication problems stimulated the work described here.

Combined Digital Phase and Amplitude Modulation Communication Systems*

CHARLES R. CAHN†, MEMBER IRE

Summary—The performance of phase modulation communication systems is investigated for the asymptotic case of a high signal-to-noise ratio. A simple and excellent approximation to the probability of error is derived for both coherent and phase-comparison demodulation. In addition, a simple expression is derived for the asymptotic degradation of phase-comparison demodulation.

When the number of bits per sample of the signal is large, a combination of digital amplitude and phase modulation is found to make more efficient use of transmitter power than phase modulation alone. The number of amplitude levels may be optimized for either minimum peak power or minimum average power, using the approximate expressions for probability of error. While an analytical solution can be found for the peak power case, a numerical solution with restriction to powers of two appears preferable, and is necessary for the average power case. Accordingly, tables showing the optimum number of levels with this restriction are presented.

* Received by the IRE PGSC, March 14, 1960.

† Space Tech. Labs., Inc., Los Angeles, Calif.

INTRODUCTION

THE use of multiple amplitude levels to transmit digital information in a restricted bandwidth is well known under the name quantized pulse-amplitude modulation (PAM). In a recent paper,¹ the possibility of employing multiple phase positions for the same purpose was explored. Performance calculations for the case of nonfading signal in Gaussian noise indicate that digital multiphase modulation requires a lower signal-to-noise ratio than PAM to maintain a specified error rate. However, since multiphase modulation produces an essentially constant amplitude signal, its use is not consistent with

¹ C. R. Cahn, "Performance of digital phase-modulation communication systems," IRE TRANS. ON COMMUNICATION SYSTEMS vol. CS-7, pp. 3-6; May, 1959.

Shannon's result that a Gaussian amplitude distribution provides maximum transmission capacity when the signal is average-power limited. A combination of digital-amplitude and digital-phase modulation is thereby suggested as a desirable form of digital modulation.

It is the purpose of this paper to analyze the performance of combined phase and amplitude modulation systems, with emphasis on obtaining asymptotic, or high signal-to-noise ratio comparisons of various modulation combinations. The asymptotic comparison is justified because it applies, with reasonable accuracy, over the practical range of error rate; in addition, simple results may be obtained and related to significant geometrical parameters. It will be possible, thereby, to select combinations which are asymptotically optimum from the point of view of minimum required signal-to-noise ratio, or other suitable criteria.

The amplitude and phase modulations are considered as separately applied to each independent sample of the signal. An envelope detector is assumed for amplitude demodulation in the receiver. (This is equivalent to coherent detection for high signal-to-noise ratios.) However, two methods of practical importance for phase demodulation are considered: 1) coherent detection with a synchronized local reference and 2) phase comparison of two successive received samples. As discussed in previous papers,^{1,2} the former is theoretically optimum but may not be applicable because of fluctuations of the propagation medium or oscillator instability. The latter requires only short-term phase stability in the transmission system, but engenders an inherent degradation in performance.

Only systems in which the same number of phase positions is used with each amplitude level are considered. As pointed out by J. Taber of Space Technology Laboratories, a further saving in transmitter power will result if this restriction is removed. However, in view of the problem of practical equipment implementation, this paper has been limited to a consideration of systems with independent phase and amplitude modulations.

ASYMPTOTIC PERFORMANCE OF DIGITAL MULTIPHASE SYSTEM

Coherent Phase Demodulation

A digital multiphase system utilizes a set of n equally spaced phase positions of the carrier to transmit data expressed in terms of an n -level alphabet. Coherent detection of such a signal is relatively simple to analyze because of the geometrical symmetry. Thus, if there are n possible phases, an error in transmission will occur if noise causes the received signal to be displaced in phase by more than π/n radians. In a previous paper,¹ curves showing the variation of the required signal-to-noise ratio with number of phase positions were presented for several specified probabilities of error. An excellent analytic

approximation to these curves is derived in the following, and will facilitate analytical comparison of different modulation systems. It has been shown¹ that when the signal-to-noise ratio, S/N , is high, the probability density of phase for a sine wave in Gaussian noise may be approximated by

$$p(\theta) \cong \sqrt{\frac{S/N}{\pi}} \cos \theta \exp\left(-\frac{S}{N} \sin^2 \theta\right), \quad |\theta| < \pi/2, \quad (1)$$

where θ denotes the phase displacement from the noise-free carrier. By an appropriate change of variable, (1) may be directly integrated to yield the probability of error, the result being

$$P_e \cong 2 \left[1 - \Phi\left(\sqrt{2 \frac{S}{N}} \sin \pi/n\right) \right] \cong \frac{\exp\left(-\frac{S}{N} \sin^2 \pi/n\right)}{\sqrt{\pi S/N} \sin \pi/n}, \quad (2)$$

where $\Phi(u)$ denotes the probability integral (*i.e.*, the Gaussian distribution) and the last term in (2) is an asymptotic evaluation for large values of $\sqrt{2S/N} \sin \pi/n$.

Eq. (2) is a "normalized" or "universal" curve providing an approximation to the probability of error for any number of phase positions. The equation is plotted in Fig. 1 as the curve marked $n = \infty$, the notation indicating

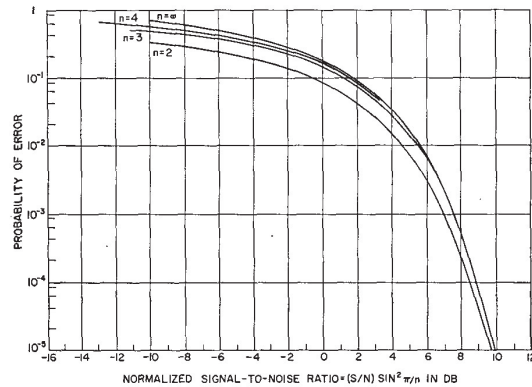


Fig. 1—Probability of error for phase-modulation systems with coherent phase demodulation.

that the approximation is exact in the limit as n becomes large. The other curves in Fig. 1 are exact curves obtained by numerical integration for small values of n , and are plotted by using the indicated normalization on the abscissa. It is seen that the approximation ($n = \infty$) is extremely good for $n \geq 3$ and is fairly close even for the binary case, $n = 2$. The closed form evaluation for $n = 2$ shows that (2) is in error only by a factor of two for this case.

The exponential form of (2) may be seen to be related to the amplitude distribution of random noise. The dependence of the exponent upon the number of n phases

² J. G. Lawton, "Theoretical error rates of 'differentially-coherent' binary and 'kineplex' data transmission systems," *Proc., IRE* vol. 47, pp. 333-334; February, 1959.

can be obtained by a simple geometric argument, as follows. Because of the exponential nature of the Gaussian distribution, the dominant behavior of the probability of error is determined by the smallest noise amplitude which can cause the resultant of signal and noise to cross a decision threshold. This shortest distance is proportional to $\sin \pi/n$, as shown in Fig. 2. By a quantitative extension of this argument, in fact, (2) may be derived without initial recourse to an approximation of the phase distribution.

Phase-Comparison Demodulation

In a previous paper,¹ demodulation by phase-comparison of successive independent signal samples was shown to introduce a 3-db degradation, for large n . On the other hand, the limiting degradation is zero for a binary system.² The exact asymptotic degradation for any intermediate value of n can be derived by an extension of the geometric argument used above. It may be seen that an error in transmission is caused by any combination of phase displacements in the two samples, the difference of which exceeds π/n in magnitude. Again, because of the characteristics of Gaussian noise, the dominant contribution to the error probability occurs for phase displacements each of magnitude $\pi/2n$, but in opposite directions. The exponential behavior of this dominant contribution is, accordingly,

$$\left[\exp \left(-\frac{S}{N} \sin^2 \pi/2n \right) \right]^2 = \exp \left(-2 \frac{S}{N} \sin^2 \pi/2n \right). \quad (3)$$

Comparison of (2) and (3) shows that the signal-to-noise ratios required for equal probabilities of error with the two-phase demodulation procedures are asymptotically related by

$$\frac{(S/N)_{\text{phase comparison}}}{(S/N)_{\text{coherent}}} = \frac{\sin^2 \pi/n}{2 \sin^2 \pi/2n}. \quad (4)$$

Eq. (4) is plotted in Fig. 3. It is seen that the asymptotic degradation is 1.8 db for a tri-phase system, 2.3 db for a quadri-phase system, and very close to 3 db for larger numbers of phases.

Eq. (4) may be combined with (2) to obtain a universal curve for probability of error with phase-comparison demodulation. Thus,

$$P_e \cong 2 \left[1 - \Phi \left(2 \sqrt{\frac{S}{N}} \sin \pi/2n \right) \right] \cong \frac{\exp \left(-2 \frac{S}{N} \sin^2 \pi/2n \right)}{\sqrt{\pi 2S/N} \sin \pi/2n}. \quad (5)$$

Eq. (5) is plotted in Fig. 4 as the curve labeled $n = \infty$, the notation having the same significance described for (2). For comparison, exact curves, obtained by numerical integration for small values of n , are also plotted in Fig. 4.

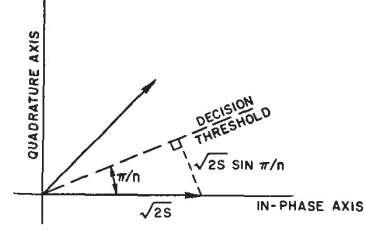


Fig. 2—Illustration of minimum distance to decision threshold.

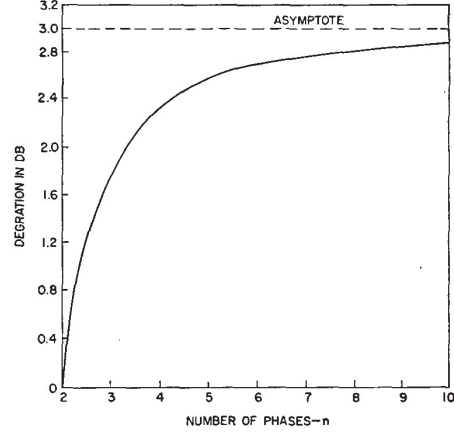


Fig. 3—Limiting degradation of phase-comparison demodulation over coherent phase demodulation.

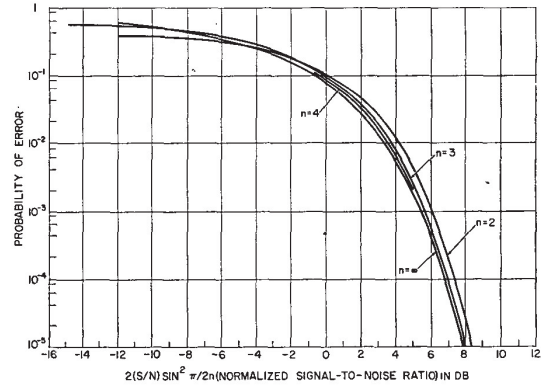


Fig. 4—Probability of error for phase-modulation systems with phase-comparison demodulation.

It appears that (5) is an excellent approximation for all values of n .

COMBINED MULTIAMPLITUDE AND MULTIPHASE MODULATION

A modulation combination of m amplitude levels and n phase positions will allow transmission of digital data having an alphabet size $r = nm$. All modulation combinations with equal alphabet sizes will yield the same channel capacity when the number of signal samples per second

is fixed. A qualitative explanation showing why a power saving can result when several amplitude levels are used is easily obtained, as follows. Suppose that it is desired to double the alphabet size in a phase-modulation system to provide an additional bit per sample. This can, of course, be accomplished by doubling the number n of phases. However, if n is sufficiently large that $\sin \pi/n$ may be replaced by its argument, (2) shows a 6-db additional power requirement to maintain a specified error rate with the doubled number of phases. Alternatively, the additional bit can be provided by use of two amplitude levels rather than by an increase in the number of phases. To a good approximation, the two amplitudes need be spaced only by the distance between the tips of signal phasors in adjacent phase positions (see Fig. 2) and, therefore, both peak and average power requirements are increased by amounts less than 6 db when n is large.

A more accurate specification of the spacing between amplitude levels may be obtained by consideration of the probability distribution of the envelope. Since the quadrature noise component can be neglected when the signal-to-noise ratio is high, this distribution is essentially Gaussian. Consequently, for uniform probabilities of error, the m amplitudes should be equally spaced, and the amplitude-decision thresholds should be symmetrically located between adjacent levels. Denoting the ratio of amplitude spacing to *rms* noise by γ , the average probability of error in demodulating the amplitude level is easily shown to be,

$$(P_e)_{\text{amplitude}} = \frac{2(m-1)}{m} [1 - \Phi(\gamma/2)] \cong \frac{2(m-1)e^{-\gamma^2/8}}{m\sqrt{2\pi}\gamma/2} \quad (6)$$

The probability of error in demodulating the phase position is given in (2); however, it depends upon the amplitude level used for each sample. The dominant contribution is from the lowest level, which is used $1/m$ of the time. The total probability of error is then essentially the sum of the two individual probabilities of error, when these are small. Since the amplitude level and the phase position are often selected from separate binary data, it is desirable that the two probabilities of error be approximately equal. Thus, equating the exponential parts of (6) and either (2) or (5) will determine the proper amplitude spacing. The resulting levels are given by the expressions

$$S_k/S_1 = [1 + 2(k-1) \sin \pi/n]^2 \quad (7a)$$

for coherent phase demodulation, and

$$S_k/S_1 = [1 + \sqrt{8}(k-1) \sin \pi/2n]^2 \quad (7b)$$

or phase-comparison demodulation. In (7), S_k denotes the power of the k th level. The power S_1 of the lowest amplitude level is then the parameter establishing the error rate (when the number of phase positions is specified).

The optimum choice for number of amplitude levels m and number of phase positions n , for a given required

number of states $r = nm$, of course, depends upon the optimization criterion employed. Two criteria will be considered here: 1) minimum peak power, and 2) minimum average power. More complex criteria incorporating the relative costs of modulation and demodulation equipment can be formulated; however, the above are of direct interest in many practical situations. The peak and average powers can be expressed using (7). Thus, for coherent phase demodulation

$$\frac{S_{\text{peak}}}{S_1} = [1 + 2(m-1) \sin \pi/n]^2 \quad (8)$$

$$\frac{S_{\text{average}}}{S_1} = \frac{1}{m} \sum_{k=0}^{m-1} [1 + 2k \sin \pi/n]^2, \quad (9)$$

and for phase-comparison demodulation

$$\frac{S_{\text{peak}}}{S_1} = [1 + \sqrt{8}(m-1) \sin \pi/2n]^2 \quad (8a)$$

$$\frac{S_{\text{average}}}{S_1} = \frac{1}{m} \sum_{k=0}^{m-1} [1 + \sqrt{8}k \sin \pi/2n]^2. \quad (9a)$$

A meaningful optimizing procedure requires that the probability of error be introduced as a fixed constraint. Using the asymptotic approximation or (2) or (5), this constraint becomes

$$S_1 \sin^2 \pi/n = \text{constant} \quad (10a)$$

for coherent phase demodulation, and

$$S_1 \sin^2 \pi/2n = \text{constant} \quad (10b)$$

for phase-comparison demodulation.

The minimization can be very easily carried through for the peak power case, and for reasonably large $r = nm$ the approximate solutions are

$$m = \sqrt{\frac{r}{2\pi}}; \quad \frac{n}{m} = 2\pi \quad (11a)$$

for coherent phase demodulation, and

$$m = \sqrt{\frac{r}{\sqrt{2\pi}}}; \quad \frac{n}{m} = \sqrt{2\pi} \quad (11b)$$

for phase-comparison demodulation. However, it is unlikely that the true minimum would be used in practice, since restriction to a number of levels which is a power of two appears desirable in view of the coding complexity otherwise generated. Table I presents the optimum choice of levels when this restriction is made. Both coherent phase demodulation and phase-comparison demodulation lead to the same result.

The minimization procedure for the average power case leads to considerable algebraic complexity, and a simple analytic solution does not appear to be feasible. Since the primary interest is for solutions which are powers of two, a numerical minimization can be performed with this restriction, yielding the results of Table II. The

TABLE I
OPTIMUM NUMBER OF AMPLITUDE LEVELS AND PHASE
POSITIONS FOR MINIMUM PEAK POWER
(NUMBERS RESTRICTED TO POWERS OF 2)

Bits per sample	Total number of states	Number of amplitude levels	Number of phase positions
1	2	1	2
2	4	1	4
3	8	1	8
4	16	2	8
5	32	2	16
6	64	4	16
7	128	4	32
8	256	8	32

TABLE II
OPTIMUM NUMBER OF AMPLITUDE LEVELS AND PHASE
POSITIONS FOR MINIMUM AVERAGE POWER
(NUMBER RESTRICTED TO POWERS OF 2)

Bits per sample	Total number of states	Number of amplitude levels	Number of phase positions
1	2	1	2
2	4	1	4
3	8	2	4
4	16	2	8
5	32	4	8
6	64	4	16
7	128	8	16
8	256	8	32

asymptotic saving of transmitter power with use of the optimum (restricted to powers of 2) modulation combination is indicated in Figs. 5 and 6. These figures compare the increase in peak and in average power to maintain a specified probability of error as the total number of states r is increased from two for: 1) the optimum combination of phase and amplitude modulation, and 2) a single amplitude level (phase modulation only).

CONCLUSIONS

The results presented in the above discussion are of particular interest, since they indicate the existence of excellent and relatively simple approximations of the probability of error in digital multiphase systems. These approximations should facilitate the analysis of such systems, such as PAM systems have been analyzed in the past, and may point the way toward useful concepts not easily discovered by graphical techniques.

The solution for the asymptotic degradation introduced by phase-comparison demodulation is believed to be a new result and replaces previous approximations which do not fully take into account the nonlinear relation between phase displacement and noise amplitude.

Optimum combinations of multi-amplitude and multi-phase modulations have been obtained which take maximum advantage of available transmitter power. Of course, the solutions for a large alphabet size may have only limited applications in practice, because of the difficulties of instrumenting multistate systems. When the number

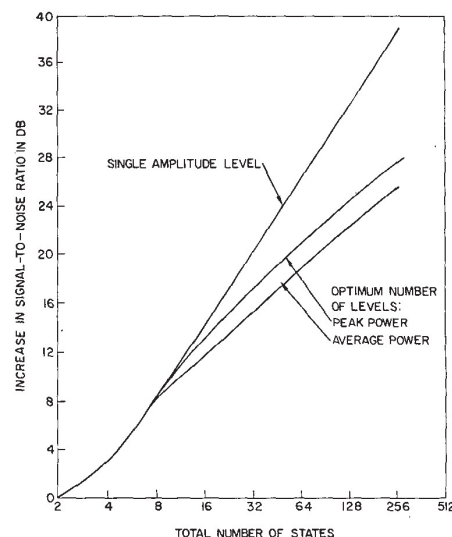


Fig. 5—Increase in signal-to-noise ratio to maintain fixed error rate with coherent phase demodulation.

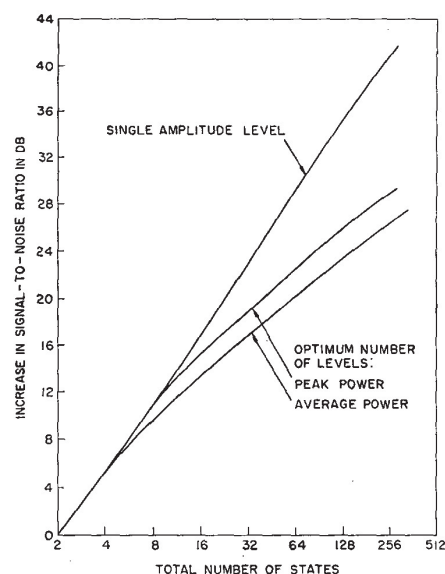


Fig. 6—Increase in signal-to-noise ratio to maintain fixed error rate with phase-comparison demodulation.

of states is relatively small, as is the usual case, phase modulation alone is found to be optimum. On the other hand, the theory indicates that significant power saving can be realized by use of combined amplitude and phase modulation when the information per sample of the signal is required to be high. A transition case is 4 bits per sample (16 states), in which a choice of 2 amplitude levels and 8 phase positions, is preferred.

From the point of view of equipment complexity, a system with combined amplitude and phase modulation is probably more simply instrumented than one using phase modulation alone, particularly when binary data are to be encoded for transmission. For example, when

bits per sample are transmitted by a choice of 8 phase positions and 2 amplitude levels, code conversion from binary to octal form is required. With the use of 16 phase positions, code conversion from binary to the 16-state form is necessary. Thus, practical benefits in addition to a saving in power may be obtained by employing combined amplitude and phase modulation in a digital communication system.

ACKNOWLEDGMENT

The author wishes to thank Dr. J. C. Hancock of Purdue University, Lafayette, Ind., for his contributions during the formulation of the pertinent problem areas. Much of the work was performed while the author was employed at Ramo-Wooldridge, a division of Thompson Ramo Wooldridge, Inc., Los Angeles, Calif.

A New Transmission Method for Pulse-Code Modulation Communication Systems*

A. P. BROGLE†

*cited in Brodie & Green, June 1972 TR 146-102 of U. Alabama
& "Com 6/73 p. 781"*

Summary—This paper presents the results of a study of the basic factors underlying the selection of a preferable method of transmission for pulse-code modulation communication systems. A new method of transmission, called biternary, is described and is compared with quaternary transmission, as a means of doubling the channel capacity of PCM systems without requiring increased bandwidth or transmission facilities. It is found that biternary transmission is more tolerant of the effects of all types of interference than quaternary transmission, and significantly reduces a major problem of quaternary transmission excessive errors, caused by low frequency cut-off effects. The use of biternary transmission is planned in present and future military pulse-code modulation communication systems.

INTRODUCTION

PULSE-CODE modulation systems are being developed to serve as the major means of multichannel communications for the Army. In addition to the need for these systems to satisfy the requirements for combat communications, future plans include widespread use of these systems for global communications.

A fundamental problem of these developments is the selection of the method of transmission which yields best over-all performance for least total cost in required equipment. For an acceptable solution, this choice must be made within the permissible limits of these factors and under the constraints of restricted bandwidth, specified transmission media and modulations, given power outputs and transmission losses, predicted noise levels, and so forth. The demands for high accuracy, together with the wide range of transmission conditions to be overcome in these systems, further restrict this choice to a very

few types of known transmission techniques which could conceivably satisfy the requirements.

In PCM, the signals to be transmitted are sampled at regular intervals, and the amplitude of each sample is sent by a coded group of pulses. The usual method of coding is to represent each sample by a group of on-off or binary pulses. However, the definition of PCM does not preclude coding of these samples by higher order numerical forms, such as ternary, quaternary, quinary etc., codes.

At the receiving end, the radio frequency carrier, if any, is amplified and demodulated. The resulting signal is then resampled and applied to a slicing circuit, which produces regenerated PCM pulses whenever the voltage applied to it falls within the prescribed range. Because of noise on the transmission path, bandwidth limitations and system imperfections, the received pulses before sampling and regeneration will differ in shape from the transmitted pulses. These distortions and uncertainties will affect the output, however, only if they are of sufficient amplitude at the sampling times to cause the slicing circuit to make an improper decision. The principal problem of pulse transmission is to find out how often this will happen for given system characteristics and noise.

TRANSMISSION CAPACITY

In PCM, the bandwidth establishes the maximum rate at which pulses can be transmitted through a perfect low-pass channel, without mutual interference between pulses. The presence of Gaussian noise in the channel then limits the number of independent pulse amplitudes that can be reliably detected in any pulse interval. Thus, the use of PCM, in theory, allows bandwidth and signal-to-noise ratio to be exchanged according to the same law as an ideal channel.

* Received by the PGCS, June 24, 1960. Presented at the Fifth Natl. Communications Symp., Utica, N. Y., October 5, 1959. Reprinted from the 1960 IRE INTERNATIONAL CONVENTION RECORD.

† RCA Corp., New York, N. Y. Formerly with U. S. Army Signal Corp., Ft. Monmouth, N. J.