THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXVIII | MAY 1959 | NUMBER 3

Probability of Error for Optimal Codes in a Gaussian Channel

By CLAUDE E. SHANNON

(Manuscript received October 17, 1958)

A study is made of coding and decoding systems for a continuous channel with an additive gaussian noise and subject to an average power limitation at the transmitter. Upper and lower bounds are found for the error probability in decoding with optimal codes and decoding systems. These bounds are close together for signaling rates near channel capacity and also for signaling rates near zero, but diverge between. Curves exhibiting these bounds are given.

I. INTRODUCTION

Consider a communication channel of the following type: Once each second a real number may be chosen at the transmitting point. This number is transmitted to the receiving point but is perturbed by an additive gaussian noise, so that the ith real number, s_i, is received as $s_i + x_i$. The x_i are assumed independent gaussian random variables all with the same variance N.

A *code word* of length n for such a channel is a sequence of n real numbers $(s_1, s_2, \cdots, s_n)$. This may be thought of geometrically as a point in n-dimensional Euclidean space. The effect of noise is then to move this point to a nearby point according to a spherical gaussian distribution.

A *block code* of length n with M words is a mapping of the integers 1, 2, $\cdots$, M into a set of M code words $w_1, w_2, \cdots, w_M$ (not necessarily all distinct). Thus, geometrically, a block code consists of a collection of M (or less) points with associated integers. It may be thought of as a way of transmitting an integer from 1 to M to the receiving point (by sending the corresponding code word). A *decoding system* for such a code is a partitioning of the n-dimensional space into M subsets corresponding to the integers from 1 to M. This is a way of deciding, at the receiving point, on the transmitted integer. If the received signal is in subset S_i, the transmitted message is taken to be integer i.

We shall assume throughout that all integers from 1 to M occur as messages with equal probability $1/M$. There is, then, for a given code and decoding system, a definite probability of error for transmitting a message. This is given by

$$P_e = \frac{1}{M} \sum_{i=i}^{M} P_{ei} ,$$

where P_{ei} is the probability, if code word w_i is sent, that it will be decoded as an integer other than i. P_{ei} is, of course, the total probability under the gaussian distribution, centered on w_i in the region complementary to S_i.

An *optimal decoding system* for a code is one which minimizes the probability of error for the code. Since the gaussian density is monotone decreasing with distance, an optimal decoding system for a given code is one which decodes any received signal as the integer corresponding to the geometrically nearest code word. If there are several code words at the same minimal distance, any of these may be used without affecting the probability of error. A decoding system of this sort is called *minimum distance* decoding or *maximum likelihood* decoding. It results in a partitioning of the n-dimensional space into n-dimensional polyhedra, or polytopes, around the different signal points, each polyhedron bounded by a finite number (not more than $M - 1$) of $(n - 1)$-dimensional hyperplanes.

We are interested in the problem of finding good codes, that is, placing M points in such a way as to minimize the probability of error P_e. If there were no conditions on the code words, it is evident that the probability of error could be made as small as desired for any M, n and N by placing the code words at sufficiently widely separated points in the n space. In normal applications, however, there will be limitations on the choice of code words that prevent this type of solution. An interesting case that has been considered in the past is that of placing some kind of *average power limitation* on the code words; the distance of the points from the origin should not be too great. We may define three different possible limitations of this sort:

i. All code words are required to have *exactly the same power* P or the same distance from the origin. Thus, we are required to choose for code words points lying on the surface of a sphere of radius $\sqrt{nP}$.

ii. All code words have power P *or less*. Here all code words are required to lie interior to or on the surface of a sphere of radius $\sqrt{nP}$.

iii. The *average power* of all code words is P or less. Here, individual code words may have a greater squared distance than nP but the average of the set of squared distances cannot exceed nP.

These three cases lead to quite similar results, as we shall see. The first condition is simpler and leads to somewhat sharper conclusions — we shall first analyze this case and use these results for the other two conditions. *Therefore, until the contrary is stated, we assume all code words to lie on the sphere of radius* $\sqrt{nP}$.

Our first problem is to estimate, as well as possible, the probability of error $P_e(M, n, \sqrt{P/N})$ for the best code of length n containing M words each of power P and perturbed by noise of variance N. This minimal or *optimal probability of error* we denote by $P_{e\,\text{opt}}(M, n, \sqrt{P/N})$. It is clear that, for fixed M, n, $P_{e\,\text{opt}}$ will be a function only of the quotient $A = \sqrt{P/N}$ by change of scale in the geometrical picture. We shall obtain upper and lower bounds on $P_{e\,\text{opt}}$ of several different types. Over an important range of values these bounds are reasonably close together, giving good estimates of $P_{e\,\text{opt}}$. Some calculated values and curves are given and the bounds are used to develop other bounds for the second and third type conditions on the code words.

The geometrical approach we use is akin to that previously used by the author[1] but carried here to a numerical conclusion. The problem is also close to that studied by Rice,[2] who obtained an estimate similar to but not as sharp as one of our upper bounds. The work here is also analogous to bounds given by Elias[3] for the binary symmetric and binary erasure channels, and related to bounds for the general discrete memoryless channel given by the author.[4]

In a general way, our bounds, both upper and lower, vary exponentially with n for a fixed signaling rate, R, and fixed P/N. In fact, they all can be put [letting $R = (1/n) \log M$, so that R is the transmitting rate for the code] in the form

$$e^{-E(R)\,n+o(n)}, \tag{1}$$

where $E(R)$ is a suitable function of R (and of P/N, which we think of as a fixed parameter). [In (1), $o(n)$ is a term of order less than n; as $n \to \infty$ it becomes small relative to $E(R)n$.]

Thus, for large n, the logarithm of the bound increases linearly with n or, more precisely, the ratio of this logarithm to n approaches a constant $E(R)$. This quantity $E(R)$ gives a crude measure of how rapidly the probability of error approaches zero. We will call this type of quantity a *reliability*. More precisely, we may define the reliability for a channel as follows:

$$E(R) = \lim_{n\to\infty} \sup -\frac{1}{n} \log P_{e\,\mathrm{opt}}(R,n), \tag{2}$$

where $P_{e\,\mathrm{opt}}(R,n)$ is the optimal probability of error for codes of rate R and length n. We will find that our bounds determine $E(R)$ *exactly* over an important range of rates, from a certain critical rate R_c up to channel capacity. Between zero and R_c, E is not exactly determined by our bounds, but lies within a not too wide range.

In connection with the reliability E, it may be noted that, in (1) above, knowledge of $E(R)$ and n does not closely determine the probability of error, even when n is large; the term $o(n)$ can cause a large and, in fact, increasing multiplier. On the other hand, given a desired probability of error and $E(R)$, the necessary value of the code length n *will* be sharply determined when n is large; in fact, n will be asymptotic to $-(1/E) \log P_e$. This inverse problem is perhaps the more natural one in applications: given a required level of probability of error, how long must the code be?

The type of channel we are studying here is, of course, closely related to a band-limited channel (W cycles per second wide) perturbed by white gaussian noise. In a sense, such a band-limited channel can be thought of as having $2W$ coordinates per second, each independently perturbed by a gaussian variable. However, such an identification must be treated with care, since to control these degrees of freedom physically and stay strictly within the bandwidth would require an infinite delay.

It is possible to stay very closely within a bandwidth W with a large but finite delay T, for example, by using $(\sin x)/x$ pulses with one tail deleted T from the maximum point. This deletion causes a spill-over outside the band of not more than the energy of the deleted part, an amount less than $1/T$ for the unit $(\sin x)/x$ case. By making T large, we can approach the situation of staying within the allotted bandwidth and also, for example, approach zero probability of error at signaling rates close to channel capacity.

However, for the problems we are studying here, delay as related to probability of error is of fundamental importance and, in applications of our results to such band-limited channels, the additional delay involved in staying closely within the allotted channel must be remembered. This is the reason for defining the channel as we have above.

II. SUMMARY

In this section we summarize briefly the main results obtained in the paper, both for easy reference and for readers who may be interested in the results without wishing to work through the detailed analysis. It might be said that the algebra involved is in several places unusually tedious.

We use the following notations:

P = signal power (each code word is on the surface of a sphere of radius $\sqrt{nP}$);

N = noise power (variance N in each dimension);

$A = \sqrt{P/N}$ = signal-to-noise "amplitude" ratio;

n = number of dimensions or block length of code;

M = number of code words;

$R = (1/n) \log M$ = signaling rate for a code (natural units);

$C = \frac{1}{2} \log (P + N)/N = \frac{1}{2} \log (A^2 + 1)$ = channel capacity (per degree of freedom);

θ = variable for half-angle of cones appearing in the geometrical problem which follows;

$\Omega(\theta)$ = solid angle in n space of a cone of half-angle θ, or area of unit n sphere cut out by the cone;

$\theta_0 = \cot^{-1} A$ = cone angle relating to channel capacity;

θ_1 = cone angle such that the solid angle $\Omega(\theta_1)$ of this cone is $(1/M)\Omega(\pi)$, [the solid angle of a sphere is $\Omega(\pi)$]; thus, θ_1 is a cone angle related to the rate R;

$G = G(\theta) = \frac{1}{2}(A \cos \theta + \sqrt{A^2 \cos^2 \theta + 4})$, a quantity which appears often in the formulas;

θ_c = the solution of $2 \cos \theta_c - AG(\theta_c) \sin^2 \theta_c = 0$ (this critical angle is important in that the nature of the bounds change according as $\theta_1 > \theta_c$ or $\theta_1 < \theta_c$);

$Q(\theta) = Q(\theta, A, n)$ = probability of a point X in n space, at distance $A\sqrt{n}$ from the origin, being moved outside a circular cone of half-angle θ with vertex at the origin O and axis OX (the perturbation is assumed spherical gaussian with unit variance in all dimensions);

$E_L(\theta) = A^2/2 - \frac{1}{2}AG \cos \theta - \log (G \sin \theta)$, an exponent appearing in our bounds;

$P_{e\,\text{opt}}(n, R, A)$ = Probability of error for the best code of length n, signal-to-noise ratio A and rate R;

$\Phi(X)$ = normal distribution with zero mean and unit variance.

The results of the paper will now be summarized. $P_{e\,\mathrm{opt}}$ can be bounded as follows:

$$Q(\theta_1) \leqq P_{e\,\mathrm{opt}} \leqq Q(\theta_1) - \int_0^{\theta_1} \frac{\Omega(\theta)}{\Omega(\theta_1)}\, dQ(\theta). \tag{3}$$

[Here $dQ(\theta)$ is negative, so the right additional term is positive.] These bounds can be written in terms of rather complex integrals. To obtain more insight into their behavior, we obtain, in the first place, asymptotic expressions for these bounds when n is large and, in the second place, cruder bounds which, however, are expressed in terms of elementary functions without integrals.

The asymptotic lower bound is (asymptotically correct as $n \to \infty$)

$$\begin{aligned} Q(\theta_1) &\sim \frac{1}{\sqrt{n\pi}\, G\sqrt{1+G^2}\sin\theta_1\,(\cos\theta_1 - AG\sin^2\theta_1)}\, e^{-E_L(\theta_1)n} \\ &= \frac{\alpha(\theta_1)}{\sqrt{n}}\, e^{-E_L(\theta_1)n} \qquad (\theta_1 > \theta_0). \end{aligned} \tag{4}$$

The asymptotic upper bound is

$$Q(\theta_1) - \int_0^{\theta_1} \frac{\Omega(\theta)}{\Omega(\theta_1)}\, dQ(\theta) \sim \frac{\alpha(\theta_1)}{\sqrt{n}}\, e^{-E_L(\theta_1)n}\left(1 - \frac{\cos\theta_1 - AG\sin^2\theta_1}{2\cos\theta_1 - AG\sin^2\theta_1}\right). \tag{5}$$

This formula is valid for $\theta_0 < \theta_1 < \theta_c$. In this range the upper and lower asymptotic bounds differ only by the factor in parentheses independent of n. *Thus, asymptotically, the probability of error is determined by these relations to within a multiplying factor depending on the rate.* For rates near channel capacity (θ_1 near θ_0) the factor is just a little over unity; the bounds are close together. For lower rates near R_c (corresponding to θ_c), the factor becomes large. For $\theta_1 > \theta_c$ the upper bound asymptote is

$$\frac{1}{\cos\theta_c \sin^3\theta_c G(\theta_c)\sqrt{\pi E''(\theta_c)[1 + G(\theta_c)]^2}}\, e^{-n[E_L(\theta_c) - R]}. \tag{6}$$

In addition to the asymptotic bound, we also obtain firm bounds, valid for all n, but poorer than the asymptotic bounds when n is large. The firm lower bound is

$$P_e \geqq \frac{1}{6}\, \frac{\sqrt{n-1}\, e^{3/2}}{n(A+1)^3 e^{(A+1)^2/2}}\, e^{-E_L(\theta_1)n}. \tag{7}$$

It may be seen that this is equal to the asymptotic bound multiplied by a factor essentially independent of n. The firm upper bound {valid if the maximum of $G^n (\sin \theta)^{2n-3} \exp [-(n/2)(A^2 - AG \cos \theta)]$ in the range 0 to θ_1 occurs at θ_1} is

$$P_{e\text{ opt}} \leqq \theta_1 \sqrt{2n}\, e^{3/2} G^n(\theta_1) \sin \theta_1^{\,n-2} \exp\left[\frac{n}{2}(-A^2 + AG \cos \theta_1)\right] \cdot \left\{1 + \frac{1}{n\theta_1 \min [A, AG(\theta_1) \sin \theta_1 - \cot \theta_1]}\right\}. \tag{8}$$

For rates near channel capacity, the upper and lower asymptotic bounds are both approximately the same, giving, where n is large and $C - R$ small (but positive):

$$P_{e\text{ opt}} \doteq \Phi\left[\sqrt{n} \sqrt{\frac{2P(P+N)}{N(P+2N)}}\,(R - C)\right], \tag{9}$$

where Φ is the normal distribution with unit variance.

To relate the angle θ_1 in the above formulas to the rate R, inequalities are found:

$$\frac{\Gamma\left(\frac{n}{2}+1\right)(\sin \theta_1)^{n-1}}{n\Gamma\left(\frac{n+1}{2}\right)\pi^{1/2} \cos \theta_1}\left(1 - \frac{1}{n}\tan^2 \theta_1\right) \leqq e^{-nR} \leqq \frac{\Gamma\left(\frac{n}{2}+1\right)(\sin \theta_1)^{n-1}}{n\Gamma\left(\frac{(n+1)}{2}\right)\pi^{1/2} \cos \theta_1}. \tag{10}$$

Asymptotically, it follows that:

$$e^{-nR} \sim \frac{\sin^n \theta_1}{\sqrt{2\pi n} \sin \theta_1 \cos \theta_1}. \tag{11}$$

For low rates (particularly $R < R_c$), the above bounds diverge and give less information. Two different arguments lead to other bounds useful at low rates. The *low rate upper bound* is:

$$P_{e\text{ opt}} \leqq \frac{1}{\lambda A \sqrt{\pi n}} e^{n[R - (\lambda^2 A^2)/4]}, \tag{12}$$

where λ satisfies $R = [1 - (1/n)] \log (\sin 2 \sin^{-1} \lambda/\sqrt{2})$. Note that as $R \to 0$, $\lambda \to 1$ and the upper bound is approximately

$$\frac{1}{A\sqrt{\pi n}}\, e^{-nA^2/4}.$$

The *low rate lower bound* may be written

$$P_{e\,\mathrm{opt}} \geqq \frac{1}{2}\,\Phi\left[-A\left(\frac{2M}{2M-1}\,\frac{n}{2}\right)^{1/2}\right]. \tag{13}$$

For M large, this bound is close to $\frac{1}{2}\Phi(-A\sqrt{n/2})$ and, if n is large, this is asymptotic to $1/(A\sqrt{\pi n})\,e^{-nA^2/4}$. Thus, for rates close to zero and large n we again have a situation where the bounds are close together and give a sharp evaluation of $P_{e\,\mathrm{opt}}$.

With codes of rate $R \geqq C + \epsilon$, where ϵ is fixed and positive, $P_{e\,\mathrm{opt}}$ approaches unity as the code length n increases.

III. THE LOWER BOUND BY THE "SPHERE-PACKING" ARGUMENT

Suppose we have a code with M points each at distance $\sqrt{nP}$ from the origin in n space. Since any two words are at equal distance from the origin, the $n - 1$ hyperplane which bisects the connecting line passes through the origin. Thus, all of the hyperplanes which determine the polyhedra surrounding these points (for the optimal decoding system) pass through the origin. These polyhedra, therefore, are pyramids with apexes at the origin. The probability of error for the code is

$$\frac{1}{M}\sum_{i=1}^{M} P_{ei},$$

where P_{ei} is the probability, if code word i is used, that it will be carried by the noise outside the pyramid around the ith word. The probability of being *correct* is

$$1 - \frac{1}{M}\sum_{i=1}^{M} P_{ei} = \frac{1}{M}\sum_{i=1}^{M}(1 - P_{ei});$$

that is, the average probability of a code word being moved to a point *within* its own pyramid.

Let the ith pyramid have a solid angle Ω_i (that is, Ω_i is the area cut out by the pyramid on the unit n-dimensional spherical surface). Consider, for comparison, a right circular n-dimensional cone with the same solid angle Ω_i and having a code word on its axis at distance $\sqrt{nP}$. We assert that *the probability of this comparison point being moved to within its cone is greater than that of w_i being moved to within its pyramid.* This is because of the monotone decreasing probability density with distance from the code word. The pyramid can be deformed into the cone by moving small conical elements from far distances to nearer distances, this movement continually increasing probability. This is suggested for a three-dimensional case in Fig. 1. Moving small conical elements from outside the cone to inside it increases probability, since the probability density is greater inside the cone than outside. Formally, this follows by integrating the probability density over the region R_1 in the cone but not in the pyramid, and in the region R_2 in the pyramid but not in the cone. The first is greater than the solid angle Ω of R_1 times the density at the edge of the cone. The value for the pyramid is less than the same quantity.

We have, then, a bound on the probability of error P_e for a given code:

$$P_e \geqq \frac{1}{M} \sum_{i=1}^{M} Q^*(\Omega_i), \tag{14}$$

where Ω_i is the solid angle for the ith pyramid, and $Q^*(\Omega)$ is the probability of a point being carried outside a surrounding cone of solid angle Ω. It is also true that

$$\sum_{i=1}^{M} \Omega_i = \Omega_0 ,$$

the solid angle of an n sphere, since the original pyramids corresponded to a partitioning of the sphere. Now, using again the property that the density decreases with distance, it follows that $Q^*(\Omega)$ is a convex function of Ω. Then we may further simplify this bound by replacing each Ω_i by

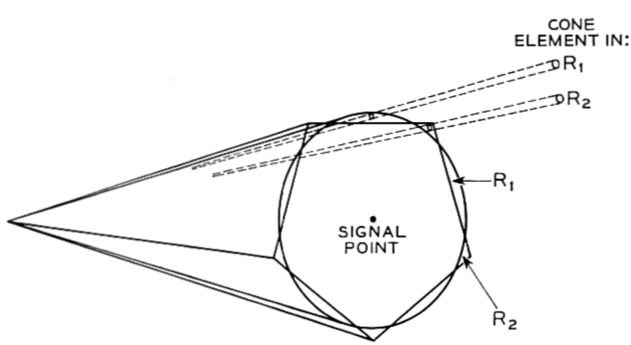

Fig. 1 — Pyramid deformed into cone by moving small conical elements from far to nearer distances.

the average Ω_0/M. In fact,

$$\frac{1}{M}\sum_{i=1}^{M} Q^*(\Omega_i) \geqq Q^*\left(\frac{\Omega_0}{M}\right),$$

and hence

$$P_e \geqq Q^*\left(\frac{\Omega_0}{M}\right).$$

It is more convenient to work in terms of the half-cone angle θ rather than solid angles Ω. We define $Q(\theta)$ to be the probability of being carried outside a cone of half-angle θ. Then, if θ_1 corresponds to the cone of solid angle Ω_0/M, the bound above may be written

$$P_e \geqq Q(\theta_1). \tag{15}$$

This is our fundamental lower bound for P_e. It still needs translation into terms of P, N, M and n, and estimation in terms of simple functions.

It may be noted that this bound is exactly the probability of error that would occur if it were possible to subdivide the space into M congruent cones, one for each code word, and place the code words on the axes of these cones. It is, of course, very plausible intuitively that any actual code would have a higher probability of error than would that with such a conical partitioning. Such a partitioning clearly is possible only for $n = 1$ or 2, if $M > 2$.

The lower bound $Q(\theta_1)$ can be evaluated in terms of a distribution familiar to statisticians as the noncentral t-distribution.[5] The noncentral t may be thought of as the probability that the ratio of a random variable $(z + \delta)$ to the root mean square of f other random variables

$$\sqrt{\frac{1}{f}\sum x_i^2}$$

does not exceed t, where all variates x_i and z are gaussian and independent with mean zero and unit variance and δ is a constant. Thus, denoting it by $P(f, \delta, t)$, we have

$$P(f,\delta,t) = \Pr\left\{\frac{z+\delta}{\sqrt{\frac{1}{f}\sum_{1}^{f} x_i^2}} \leqq t\right\}. \tag{16}$$

In terms of our geometrical picture, this amounts to a spherical gaussian distribution with unit variance about a point δ from the origin in $f + 1$ space. The probability $P(f, \delta, t)$ is the probability of being outside a cone from the origin having the line segment to the center of the distribution as axis. The cotangent of the half-cone angle θ is $t/\sqrt{f}$. Thus the probability $Q(\theta)$ is seen to be given by

$$Q(\theta) = P\left(n - 1, \sqrt{\frac{nP}{N}}, \sqrt{n-1}\cot\theta\right). \tag{17}$$

The noncentral t-distribution does not appear to have been very extensively tabled. Johnson and Welch[5] give some tables, but they are aimed at other types of application and are inconvenient for the purpose at hand. Further, they do not go to large values of n. We therefore will estimate this lower bound by developing an asymptotic formula for the cumulative distribution $Q(\theta)$ and also the density distribution $dQ/d\theta$. First, however, we will find an *upper* bound on $P_{e\,\mathrm{opt}}$ in terms of the same distribution $Q(\theta)$.

IV. UPPER BOUND BY A RANDOM CODE METHOD

The upper bound for $P_{e\,\mathrm{opt}}$ will be found by using an argument based on random codes. Consider the ensemble of codes obtained by placing M points randomly on the surface of a sphere of radius $\sqrt{nP}$. More precisely, each point is placed independently of all others with probability measure proportional to surface area or, equivalently, to solid angle. Each of the codes in the ensemble is to be decoded by the minimum distance process. We wish to compute the *average probability of error* for this *ensemble of codes.*

Because of the symmetry of the code points, the probability of error averaged over the ensemble will be equal to M times the average probability of error due to any particular code point, for example, code point 1. This may be computed as follows. The probability of message number 1 being transmitted is $1/M$. The differential probability that it will be displaced by the noise into the region between a cone of half-angle θ and one of half-angle $\theta + d\theta$ (these cones having vertex at the origin and axis out to code word 1) is $-dQ(\theta)$. [Recall that $Q(\theta)$ was defined as the probability that noise would carry a point outside the cone of angle θ with axis through the signal point.] Now consider the cone of half-angle θ surrounding such a *received point* (not the cone about the message point just described). If this cone is empty of signal points, the received word will be decoded correctly as message 1. If it is not empty, other points will be nearer and the received signal will be incorrectly decoded. (The probability of two or more points at exactly the same distance is readily seen to be zero and may be ignored.)

The probability in the ensemble of codes of the cone of half-angle θ being empty is easily calculated. The probability that any particular code word, say code word 2 or code word 3, etc. is in the cone is given by $\Omega(\theta)/\Omega(\pi)$, the ratio of the solid angle in the cone to the total solid angle. The probability a particular word is *not* in the cone is $1 - \Omega(\theta)/\Omega(\pi)$. The probability that all $M - 1$ other words are not in the cone is $[1 - \Omega(\theta)/\Omega(\pi)]^{M-1}$ since these are, in the ensemble of codes, placed independently. The probability of error, then, contributed by situations where the point 1 is displaced by an angle from θ to $\theta + d\theta$ is given by $-(1/M)\{1 - [1 - \Omega(\theta)/\Omega(\pi)]^{M-1}\}dQ(\theta)$. The total average probability of error for all code words and all noise displacements is then given by

$$P_{er} = -\int_{\theta=0}^{\pi} \left\{1 - \left[1 - \frac{\Omega(\theta)}{\Omega(\pi)}\right]^{M-1}\right\} dQ(\theta). \tag{18}$$

This is an exact formula for the average probability of error P_{er} for our random ensemble of codes. Since this is an average of P_e for particular codes, there must exist particular codes in the ensemble with at least this good a probability of error, and certainly then $P_{e\,\mathrm{opt}} \leqq P_{er}$.

We may weaken this bound slightly but obtain a simpler formula for calculation as follows. Note first that $\{1 - [\Omega(\theta)/\Omega(\pi)]^{M-1}\} \leqq 1$ and also, using the well-known inequality $(1 - x)^n \geqq 1 - nx$, we have $\{1 - [1 - \Omega(\theta)/\Omega(\pi)]^{M-1}\} \leqq (M - 1)[\Omega(\theta)/\Omega(\pi)] \leqq M[\Omega(\theta)/\Omega(\pi)]$. Now, break the integral into two parts, $0 \leqq \theta \leqq \theta_1$ and $\theta_1 \leqq \theta \leqq \pi$. In the first range, use the inequality just given and, in the second range, bound the expression in braces by 1. Thus,

$$\begin{aligned} P_{er} &\leqq -\int_0^{\theta_1} M\left[\frac{\Omega(\theta)}{\Omega(\pi)}\right] dQ(\theta) - \int_{\theta_1}^{\pi} dQ(\theta), \\ P_{er} &\leqq -\frac{M}{\Omega(\pi)}\int_0^{\theta_1} \Omega(\theta)dQ(\theta) + Q(\theta_1). \end{aligned} \tag{19}$$

It is convenient to choose for θ_1 the same value as appeared in the lower bound; that is, the θ_1 such that $\Omega(\theta_1)/\Omega(\pi) = 1/M$ — in other words, the θ_1 for which one expects one point within the θ_1 cone. The second term in (19) is then the same as the lower bound on $P_{e\,\mathrm{opt}}$ obtained previously. In fact, collecting these results, we have

$$Q(\theta_1) \leqq P_{e\,\mathrm{opt}} \leqq Q(\theta_1) - \frac{M}{\Omega(\pi)}\int_0^{\theta_1} \Omega(\theta)dQ(\theta), \tag{20}$$

where $M\Omega(\theta_1) = \Omega(\pi)$. *These are our fundamental lower and upper bounds on $P_{e\,\mathrm{opt}}$.*

We now wish to evaluate and estimate $\Omega(\theta)$ and $Q(\theta)$.

V. FORMULAS FOR RATE R AS A FUNCTION OF THE CONE ANGLE θ

Our bounds on probability of error involve the code angle θ_1 such that the solid angle of the cone is $1/M = e^{-nR}$ times the full solid angle of a sphere. To relate these quantities more explicitly we calculate the solid angle of a cone in n dimensions with half-angle θ. In Fig. 2 this means calculating the $(n - 1)$-dimensional area of the cap cut out by the cone on the unit sphere. This is obtained by summing the contributions due to ring-shaped elements of area (spherical surfaces in $n - 1$ dimensions

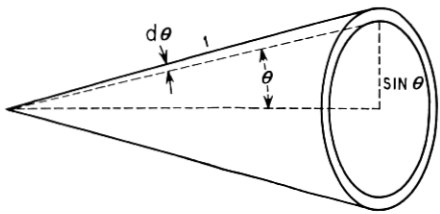

Fig. 2 — Cap cut out by the cone on the unit sphere.

of radius $\sin \theta$ and of incremental width $d\theta$). Thus, the total area of the cap is given by

$$\Omega(\theta_1) = \frac{(n-1)\pi^{(n-1)/2}}{\Gamma\left(\frac{n+1}{2}\right)} \int_0^{\theta_1} (\sin \theta)^{n-2} d\theta. \tag{21}$$

Here we used the formula for the surface $S_n(r)$ of a sphere of radius r in n dimensions, $S_n(r) = n\pi^{n/2}r^{n-1}/\Gamma(n/2 + 1)$.

To obtain simple inequalities and asymptotic expressions for $\Omega(\theta_1)$, make the change of variable in the integral $x = \sin \theta, d\theta = (1 - x^2)^{-1/2}dx$. Let $x_1 = \sin \theta_1$ and assume $\theta_1 < \pi/2$, so that $x_1 < 1$. Using the mean value theorem we obtain

$$(1 - x^2)^{-1/2} = (1 - x_1^2)^{-1/2} + \frac{\alpha}{(1 - \alpha^2)^{3/2}} (x - x_1), \tag{22}$$

where $0 \leqq \alpha \leqq x_1$. The term $\alpha(1 - \alpha^2)^{-3/2}$ must lie in the range from 0 to $x_1(1 - x_1^2)^{-3/2}$ since this is a monotone increasing function. Hence we have the inequalities

$$(1 - x_1^2)^{-1/2} + \frac{(x - x_1)x_1}{(1 - x_1^2)^{3/2}} \leqq (1 - x^2)^{-1/2} \leqq (1 - x_1^2)^{-1/2} \qquad 0 \leqq x \leqq x_1. \tag{23}$$

Note that $x - x_1$ is negative, so the correction term on the left is of the right sign. If we use these in the integral for $\Omega(\theta_1)$ we obtain

$$\frac{(n-1)\pi^{(n-1)/2}}{\Gamma\left(\frac{n+1}{2}\right)} \int_0^{x_1} x^{n-2}\left[(1-x_1^2)^{-1/2} + \frac{(x-x_1)x_1}{(1-x_1^2)^{3/2}}\right] dx$$
$$\leqq \Omega(\theta_1) \leqq \frac{(n-1)\pi^{(n-1)/2}}{\Gamma\left(\frac{n+1}{2}\right)} \int_0^{x_1} x^{n-2} \frac{dx}{\sqrt{1-x_1^2}}, \tag{24}$$

$$\frac{(n-1)\pi^{(n-1)/2}}{\Gamma\left(\frac{n+1}{2}\right)\sqrt{1-x_1^2}}\left[\frac{x_1^{n-1}}{n-1} + \frac{x_1^{n+1}}{n(1-x_1^2)} - \frac{x_1^{n+1}}{(n-1)(1-x_1^2)}\right]$$
$$\leqq \Omega(\theta_1) \leqq \frac{(n-1)\pi^{(n-1)/2}x_1^{n-1}}{\Gamma\left(\frac{n+1}{2}\right)(n-1)\sqrt{1-x_1^2}}, \tag{25}$$

$$\frac{\pi^{(n-1)/2}(\sin\theta_1)^{n-1}}{\Gamma\left(\frac{n+1}{2}\right)\cos\theta_1}\left(1 - \frac{1}{n}\tan^2\theta_1\right)$$
$$\leqq \Omega(\theta_1) \leqq \frac{\pi^{(n-1)/2}(\sin\theta_1)^{n-1}}{\Gamma\left(\frac{n+1}{2}\right)\cos\theta_1}. \tag{26}$$

Therefore, as $n \to \infty$, $\Omega(\theta_1)$ is asymptotic to the expression on the right.

The surface of the unit n sphere is $n\pi^{n/2}/\Gamma(n/2 + 1)$, hence,

$$\frac{\Gamma\left(\frac{n}{2}+1\right)(\sin\theta_1)^{n-1}}{n\Gamma\left(\frac{n+1}{2}\right)\pi^{1/2}\cos\theta_1}\left(1 - \frac{1}{n}\tan^2\theta_1\right) \leqq e^{-nR}$$
$$= \frac{\Omega(\theta_1)}{\Omega(\pi)} \leqq \frac{\Gamma\left(\frac{n}{2}+1\right)(\sin\theta_1)^{n-1}}{n\Gamma\left(\frac{n+1}{2}\right)\pi^{1/2}\cos\theta_1}. \tag{27}$$

Replacing the gamma functions by their asymptotic expressions, we obtain

$$e^{-nR} = \frac{\sin^n\theta_1}{\sqrt{2\pi n}\,\sin\theta_1\cos\theta_1}\left[1 + O\left(\frac{1}{n}\right)\right]. \tag{28}$$

Thus $e^{-nR} \sim \sin^n \theta_1/\sqrt{2\pi n} \sin \theta_1 \cos \theta_1$ and $e^{-R} \sim \sin \theta_1$. The somewhat sharper expression for e^{-nR} must be used when attempting asymptotic evaluations of P_e, since P_e is changed by a factor when θ_1 is changed by, for example, k/n. However, when only the reliability E is of interest, the simpler $R \sim -\log \sin \theta_1$ may be used.

VI. ASYMPTOTIC FORMULAS FOR $Q(\theta)$ AND $Q'(\theta)$

In Fig. 3, O is the origin, S is a signal point and the plane of the figure is a plane section in the n-dimensional space. The lines OA and OB represent a (circular) cone of angle θ about OS (that is, the intersection of this cone with the plane of the drawing.) The lines OA' and OB' correspond to a slightly larger cone of angle $\theta + d\theta$. We wish to estimate the probability $-dQ_n(\theta)$ of the signal point S being carried by noise into the region between these cones. From this, we will further calculate the probability $Q_n(\theta)$ of S being carried outside the θ cone. What is desired in both cases is an asymptotic estimate — a simple formula whose

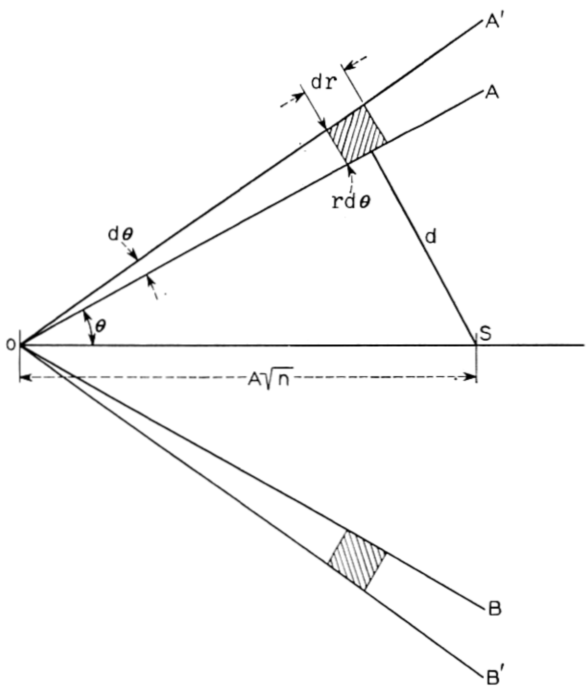

Fig. 3 — Plane of cone of half-angle θ.

ratio to the true value approaches 1 as n, the number of dimensions, increases.

The noise perturbs all coordinates normally and independently with variance 1. It produces a spherical gaussian distribution in the n-dimensional space. The probability density of its moving the signal point a distance d is given by

$$\frac{1}{(2\pi)^{n/2}} e^{-d^2/2} \, dV , \tag{29}$$

where dV is the element of volume. In Fig. 4 we wish to first calculate the probability density for the crosshatched ring-shaped region between

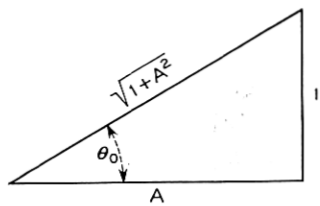

Fig. 4 — Special value θ_0 .

the two cones and between spheres about the origin of radius r and $r + dr$. The distance of this ring from the signal point is given by the cosine law as

$$d = (r^2 + A^2 n - 2rA\sqrt{n} \cos \theta)^{1/2}. \tag{30}$$

The differential volume of the ring-shaped region is $r \, dr \, d\theta$ times the surface of a sphere of radius $r \sin \theta$ in $(n - 1)$-dimensional space; that is

$$r \, dr \, d\theta \, \frac{(n-1)\pi^{(n-1)/2}(r \sin \theta)^{n-2}}{\Gamma\left(\dfrac{n+1}{2}\right)} . \tag{31}$$

Hence, the differential probability for the ring-shaped region is

$$\frac{1}{(\sqrt{2\pi})^n} \exp\left[\frac{-(r^2 + A^2 n - 2rA\sqrt{n}\cos\theta)}{2}\right] \cdot \left[\frac{(n-1)\pi^{(n-1)/2}(r \sin \theta)^{n-2}}{\Gamma\left(\dfrac{n+1}{2}\right)}\right] r \, dr \, d\theta \tag{32}$$

The differential probability $-dQ$ of being carried between the two cones is the integral of this expression from zero to infinity on dr:

$$-dQ = \frac{1}{2^{n/2}} \frac{(n-1)\, d\theta}{\sqrt{\pi}\, \Gamma\left(\frac{n+1}{2}\right)} \cdot \int_0^\infty \exp\left[\frac{-(r^2 + A^2 n - 2rA\sqrt{n}\cos\theta)}{2}\right] (r \sin\theta)^{n-2} r\, dr. \tag{33}$$

In the exponent we can think of A^2n as $A^2n(\sin^2\theta + \cos^2\theta)$. The $\cos^2$ part then combines with the other terms to give a perfect square

$$(r - A\sqrt{n}\cos\theta)^2$$

and the $\sin^2$ term can be taken outside the integral. Thus

$$-dQ = \frac{(n-1)\exp\left[-\frac{A^2 n \sin^2\theta}{2}\right](\sin\theta)^{n-2}\, d\theta}{2^{n/2}\sqrt{\pi}\,\Gamma\left(\frac{n+1}{2}\right)} \cdot \int_0^\infty \exp\left[\frac{-(r - A\sqrt{n}\cos\theta)^2}{2}\right] r^{n-1}\, dr. \tag{34}$$

We can now direct our attention to estimating the integral, which we call K. The integral can be expressed exactly as a finite, but complicated, sum involving normal distribution functions by a process of continued integration by parts. We are, however, interested in a simple formula giving the asymptotic behavior of the integral as n becomes infinite. This problem was essentially solved by David and Kruskal,[6] who prove the following asymptotic formula as a lemma:

$$\int_0^\infty z^\nu \exp\left(-\tfrac{1}{2}z^2 + z\sqrt{\nu+1}\, w\right) dz \sim \sqrt{2\pi}\left(\frac{\bar{z}}{e}\right)^\nu \exp\left(\tfrac{1}{2}\bar{z}^2\right) T, \tag{35}$$

as $\nu \to \infty$, w is fixed, $T = [1 + \frac{1}{4}(\sqrt{w^2+4} - w)^2]^{-1/2}$ and

$$\bar{z} = \tfrac{1}{2}\sqrt{\nu+1}\, w + \sqrt{\tfrac{1}{4}(\nu+1)w^2 + \nu}.$$

This is proved by showing that the main contribution to the integral is essentially in the neighborhood of the point $\bar{z}$ where the integral is a maximum. Near this point, when ν is large, the function behaves about as a normal distribution.

The integral K in (34) that we wish to evaluate is, except for a multiplying factor, of the form appearing in the lemma, with

$$z = r, \qquad w = A\cos\theta, \qquad \nu = n - 1.$$

The integral then becomes

$$\begin{aligned} K &= \exp\left(-\frac{A^2 n \cos^2\theta}{2}\right) \int_0^\infty z^{n-1} \exp -\left(\frac{z^2}{2} + zA\sqrt{n}\cos\theta\right) dz \\ &\sim \exp\left(-\frac{A^2 n \cos^2\theta}{2}\right) \sqrt{2\pi} \left(\frac{\bar{z}}{e}\right)^{n-1} T \exp\left(\frac{\bar{z}^2}{2}\right). \end{aligned} \tag{36}$$

We have

$$\begin{aligned} \bar{z} &= \tfrac{1}{2}\sqrt{n}\,A\cos\theta + \sqrt{\tfrac{1}{4}nA^2\cos^2\theta + n - 1} \\ &= \sqrt{n}\left[\tfrac{1}{2}A\cos\theta + \sqrt{\frac{A^2}{4}\cos^2\theta + 1 - \frac{1}{n}}\right] \\ &= \sqrt{n}\left[\tfrac{1}{2}A\cos\theta + \sqrt{\frac{A^2}{4}\cos^2\theta + 1} \right. \\ &\qquad\qquad \left. - \frac{1}{2n\sqrt{\frac{A^2}{4}\cos^2\theta + 1}} + O\left(\frac{1}{n^2}\right)\right]. \end{aligned} \tag{37}$$

Letting

$$G = \tfrac{1}{2}[A\cos\theta + \sqrt{A^2\cos^2\theta + 4}],$$

we have

$$\bar{z} = \sqrt{n}\,G\left[1 - \frac{1}{nG\sqrt{A^2\cos^2\theta + 4}} + O\left(\frac{1}{n^2}\right)\right],$$

so

$$\begin{aligned} \left(\frac{\bar{z}}{e}\right)^{n-1} &= \left(\frac{\sqrt{n}\,G}{e}\right)^{n-1}\left[1 - \frac{1}{nG\sqrt{A^2\cos^2\theta + 4}} + O\left(\frac{1}{n^2}\right)\right]^{n-1} \\ &\sim \left(\frac{\sqrt{n}\,G}{e}\right)^{n-1} \exp\left(-\frac{1}{G\sqrt{A^2\cos^2\theta + 4}}\right). \end{aligned} \tag{38}$$

Also,

$$\begin{aligned} \exp\frac{\bar{z}^2}{2} &= \exp\tfrac{1}{2}nG^2\left[1 - \frac{1}{nG\sqrt{A^2\cos^2\theta + 4}} + O\left(\frac{1}{n^2}\right)\right]^2 \\ &\sim \exp\left(\tfrac{1}{2}nG^2 - \frac{2G}{2\sqrt{A^2\cos^2\theta + 4}}\right) \\ &= \exp\left[\tfrac{1}{2}n(1 + AG\cos\theta) - \frac{G}{\sqrt{A^2\cos^2\theta + 4}}\right], \end{aligned} \tag{39}$$

since, on squaring G, we find $G^2 = 1 + AG \cos \theta$. Collecting terms:

$$K \sim T\sqrt{2\pi}\left(\frac{\sqrt{n}G}{e}\right)^{n-1} e^{n/2} \exp\left(-\frac{1}{G\sqrt{A^2\cos^2\theta+4}} - \frac{G}{\sqrt{A^2\cos^2\theta+4}} - \frac{A^2 n}{2}\cos^2\theta + \frac{n}{2}AG\cos\theta\right) \tag{40}$$

$$= T\sqrt{2\pi}\, n^{(n-1)/2} G^{n-1} e^{-n/2} \exp\left(-\frac{n}{2}A^2\cos^2\theta + \frac{n}{2}AG\cos\theta\right)$$

since a little algebra shows that the terms

$$1 - \frac{1}{G\sqrt{A^2\cos^2\theta+4}} - \frac{G}{\sqrt{A^2\cos^2\theta+4}}$$

in the exponential cancel to zero. The coefficient of the integral (34), using the asymptotic expression for $\Gamma[(n+1)/2]$, is asymptotic to

$$\frac{(n-1)e^{-(\sin^2\theta)(A^2n)/2}\sin\theta^{n-2}e^{(n+1)/2}}{2^{n/2}\sqrt{\pi}\left(\frac{n+1}{2}\right)^{n/2}\sqrt{2\pi}}. \tag{41}$$

Combining with the above and collecting terms (we find that $T = G/\sqrt{1+G^2}$):

$$-\frac{dQ}{d\theta} \sim \frac{n-1}{\sqrt{\pi n}}\frac{1}{\sqrt{1+G^2}\sin^2\theta}\left[G\sin\theta\exp\left(-\frac{A^2}{2}+\tfrac{1}{2}AG\cos\theta\right)\right]^n. \tag{42}$$

This is our desired asymptotic expression for the density $dQ/d\theta$.

As we have arranged it, the coefficient increases essentially as $\sqrt{n}$ and there is another term of the form $e^{-E_L(\theta)n}$, where

$$E_L(\theta) = \frac{A^2}{2} - \tfrac{1}{2}AG\cos\theta - \log(G\sin\theta).$$

It can be shown that if we use for θ the special value $\theta_0 = \cot^{-1}A$ (see Fig. 4) then $E_L(\theta_0) = 0$ and also $E'_L(\theta_0) = 0$. In fact, for this value

$$G(\theta_0) = \tfrac{1}{2}(A\cos\theta_0 + \sqrt{A^2\cos^2\theta_0+4}) = \frac{1}{2}\left(\frac{A^2}{\sqrt{A^2+1}} + \sqrt{\frac{A^4}{A^2+1}+4}\right) = \frac{1}{2}\left(\frac{A^2}{\sqrt{A^2+1}} + \frac{A^2+2}{\sqrt{A^2+1}}\right) = \csc\theta_0.$$

Hence the two terms in the logarithm cancel. Also

$$\frac{A^2}{2} - \tfrac{1}{2}AG\cos\theta_0 = \frac{A^2}{2} - \tfrac{1}{2}A\sqrt{A^2+1}\,\frac{A}{\sqrt{A^2+1}} = 0.$$

So $E_L(\theta_0) = 0$. We also have

$$E'_L(\theta) = \tfrac{1}{2}AG\sin\theta - \tfrac{1}{2}AG'\cos\theta - \frac{G'}{G} - \cot\theta. \qquad (43)$$

When evaluated, the term $-G'/G$ simplifies, after considerable algebra, to

$$\frac{A\sin\theta}{\sqrt{A^2\cos^2\theta+4}}.$$

Substituting this and the other terms we obtain

$$\begin{aligned} E'_L(\theta) = {} & \frac{A^2}{2}\sin\theta\cos\theta + \frac{A^3\cos^2\theta\sin\theta}{4\sqrt{A^2\cos^2\theta+4}} \\ & + \frac{A}{4}\frac{(A^2\cos^2\theta+4)}{\sqrt{A^2\cos^2\theta+4}}\sin\theta + \frac{A\sin\theta}{\sqrt{A^2\cos^2\theta+4}} - \cot\theta. \end{aligned} \qquad (44)$$

Adding and collecting terms, this simplifies to

$$\begin{aligned} E'_L(\theta) &= \frac{A}{2}\left(A\cos\theta + \sqrt{A^2\cos^2\theta+4}\right)\sin\theta - \cot\theta \\ &= AG\sin\theta - \cot\theta \\ &= \cot\theta\left[\frac{A^2}{2}\sin^2\theta + \frac{A}{2}\sin^2\theta\sqrt{A^2+\frac{4}{\cos^2\theta}} - 1\right]. \end{aligned} \qquad (45)$$

Notice that the bracketed expression is a monotone increasing function of θ ($0 \leqq \theta \leqq \pi/2$) ranging from -1 at $\theta = 0$ to ∞ at $\theta = \pi/2$. Also, as mentioned above, at θ_0, $G = \csc\theta_0$ and $A = \cot\theta_0$, so $E'_L(\theta_0) = 0$. It follows that $E'_L(\theta) < 0$ for $0 \leqq \theta < \theta_0$ and $E'_L(\theta) > 0$ for $\theta_0 \leqq \theta < \pi/2$.

From this, it follows that, in the range from some θ_1 to $\pi/2$ with $\theta_1 > \theta_0$, the minimum $E_L(\theta)$ will occur at the smallest value of θ in the range, that is, at θ_1. The exponential appearing in our estimate of $Q(\theta)$, namely, $e^{-E_L(\theta)n}$, will have its *maximum* at θ_1, for such a range. Indeed, for sufficiently large n, the maximum of the entire expression (45) must occur at θ_1, since the effect of the n in the exponent will eventually dominate anything due to the coefficient. For, if the coefficient is called $\alpha(\theta)$ with $y(\theta) = \alpha(\theta)\,e^{-nE_L(\theta)}$, then

$$y'(\theta) = e^{-nE_L(\theta)}[-\alpha(\theta)nE'_L(\theta) + \alpha'(\theta)], \qquad (46)$$

and, since $\alpha(\theta) > 0$, when n is sufficiently large $y'(\theta)$ will be negative and the only maximum will occur at θ_1. In the neighborhood of θ_1 the function goes down exponentially.

We may now find an asymptotic formula for the integral

$$Q(\theta) = \int_{\theta_1}^{\pi/2} \alpha(\theta) e^{-nE_L(\theta)} d\theta + Q(\pi/2) \tag{47}$$

by breaking the integral into two parts,

$$Q(\theta) = \int_{\theta_1}^{\theta_1 + n^{-2/3}} + \int_{\theta_1 + n^{-2/3}}^{\pi/2} + Q(\pi/2). \tag{48}$$

In the range of the first integral, $(1 - \epsilon)\alpha(\theta_1) \leqq \alpha(\theta) \leqq \alpha(\theta_1)(1 + \epsilon)$, and ϵ can be made as small as desired by taking n sufficiently large. This is because $\alpha(\theta)$ is continuous and nonvanishing in the range. Also, using a Taylor's series expansion with remainder,

$$e^{-nE_L(\theta)} = \exp\left[-nE_L(\theta_1) - n(\theta - \theta_1)E'_L(\theta_1) - n\frac{(\theta - \theta_1)^2}{2} E''_L(\theta^*) \right], \tag{49}$$

where θ^* is the interval θ_1 to θ. As n increases the maximum value of the remainder term is bounded by $n(n/2)^{-4/3} E''_{\max}$, and consequently approaches zero. Hence, our first integral is asymptotic to

$$\begin{aligned} &\alpha(\theta_1) \int_{\theta_1}^{\theta_1 + n^{-2/3}} \exp\left[-nE_L(\theta_1) - n(\theta - \theta_1)E'_L(\theta_1)\right] d\theta \\ &\quad = -\alpha(\theta_1) \exp\left[-nE_L(\theta_1)\right] \frac{\exp\left[-n(\theta - \theta_1)E'_L(\theta_1)\right]}{nE'_L(\theta_1)} \Bigg]_{\theta_1}^{\theta_1 + n^{-2/3}} \\ &\quad \sim \frac{\alpha(\theta_1) e^{-nE_L(\theta_1)}}{nE'_L(\theta_1)}. \end{aligned} \tag{50}$$

since, at large n, the upper limit term becomes small by comparison. The second integral from $\theta_1 + n^{-2/3}$ to $\pi/2$ can be dominated by the value of the integrand at $\theta_1 + n^{-2/3}$ multiplied by the range

$$\pi/2 - (\theta_1 + n^{-2/3}),$$

(since the integrand is monotone decreasing for large n). The value at $\theta_1 + n^{-2/3}$ is asymptotic, by the argument just given, to

$$\alpha(\theta_1) \exp\left[-nE_L(\theta_1) - n(n^{-2/3}) E'_L(\theta_1)\right].$$

This becomes small compared to the first integral [as does $Q(\pi/2) =$

$\Phi(-A)$ in (47)] and, consequently, on substituting for $\alpha(\theta_1)$ its value and writing θ for θ_1, *we obtain as an asymptotic expression for* $Q(\theta)$:

$$Q(\theta) \sim \frac{1}{\sqrt{n\pi}} \frac{1}{\sqrt{1 + G^2 \sin \theta}} \frac{\left[G \sin \theta \exp\left(-\frac{A^2}{2} + \frac{1}{2}AG \cos \theta\right)\right]^n}{(AG \sin^2 \theta - \cos \theta)} \qquad (51)$$

$$\left(\frac{\pi}{2} \geqq \theta > \theta_0 = \cot^{-1} A\right).$$

This expression gives an asymptotic lower bound for $P_{e\,\text{opt}}$, obtained by evaluating $Q(\theta)$ for the θ_1 such that $M\Omega(\theta_1) = \Omega(\pi)$.

Incidentally, the asymptotic expression (51) can be translated into an asymptotic expression for the noncentral t cumulative distribution by substitution of variables $\theta = \cot^{-1}(t/\sqrt{f})$ and $n - 1 = f$. This may be useful in other applications of the noncentral t-distribution.

VII. ASYMPTOTIC EXPRESSIONS FOR THE RANDOM CODE BOUND

We now wish to find similar asymptotic expressions for the *upper bound* on $P_{e\,\text{opt}}$ of (20) found by the random code method. Substituting the asymptotic expressions for $dQ(\theta)d\theta$ and for $\Omega(\theta)/\Omega(\pi)$ gives for an asymptotic upper bound the following:

$$Q(\theta_1) + e^{nR} \int_0^{\theta_1} \frac{\Gamma\left(\frac{n}{2} + 1\right)(\sin \theta)^{n-1}}{n\Gamma\left(\frac{n+1}{2}\right)\pi^{1/2} \cos \theta} \sqrt{\frac{n}{\pi}} \cdot \frac{\left[G \sin \theta \exp\left(-\frac{P}{2N} + \frac{1}{2}\sqrt{\frac{P}{N}}\, G \cos \theta\right)\right]^n}{\sqrt{1 + G^2 \sin^2 \theta}}\, d\theta. \qquad (52)$$

Thus we need to estimate the integral

$$W = \int_0^{\theta_1} \frac{1}{\cos \theta \sin^3 \theta \sqrt{1 + G^2}} \cdot \exp\left\{n\left(-\frac{P}{2N} + \frac{1}{2}\sqrt{\frac{P}{N}}\, G \cos \theta + \log G + 2 \log \sin \theta\right)\right\} d\theta. \qquad (53)$$

The situation is very similar to that in estimating $Q(\theta)$. Let the coefficient of n in the exponent be D. Note that $D = -E_L(\theta) + \log \sin \theta$. Hence its derivative reduces to

$$\frac{dD}{d\theta} = -AG \sin \theta + 2 \cot \theta. \qquad (54)$$

$dD/d\theta = 0$ has a unique root θ_c, $0 \leqq \theta_c \leqq \pi/2$ for any fixed $A > 0$. This follows from the same argument used in connection with (45), the only difference being a factor of 2 in the right member. Thus, for $\theta < \theta_c$, $dD/d\theta$ is positive and D is an increasing function of θ. Beyond this maximum, D is a decreasing function.

We may now divide the problem of estimating the integral W into cases according to the relative size of θ_c and θ_1.

Case 1: $\theta_1 < \theta_c$.

In this case the maximum of the exponent within the range of integration occurs at θ_1. Consequently, when n is sufficiently large, the maximum of the entire integrand occurs at θ_1. The asymptotic value can be estimated exactly as we estimated $Q(\theta)$ in a similar situation. The integral is divided into two parts, a part from $\theta_1 - n^{-2/3}$ to θ_1 and a second part from 0 to $\theta_1 - n^{-2/3}$. In the first part the integrand behaves asymptotically like:

$$\frac{1}{\cos\theta_1 \sin^3\theta_1 \sqrt{1 + G^2(\theta_1)}} \exp\left(n\left\{-\frac{P}{2N} + \frac{1}{2}\sqrt{\frac{P}{N}}\, G(\theta_1)\cos\theta_1 + \log G(\theta_1) + 2\log\sin\theta_1 - (\theta - \theta_1)[AG(\theta_1)\sin\theta_1 - 2\cot\theta_1]\right\}\right). \tag{55}$$

This integrates asymptotically to

$$\frac{\exp\left\{n\left[-\frac{P}{2N} + \frac{1}{2}\sqrt{\frac{P}{N}}\, G(\theta_1)\cos\theta_1 + \log G(\theta_1) + 2\log\sin\theta_1\right]\right\}}{\cos\theta_1 \sin^3\theta_1 \sqrt{1 + G^2(\theta_1)}\,[-AG(\theta_1)\sin\theta_1 + 2\cot\theta_1]n}. \tag{56}$$

The second integral becomes small in comparison to this, being dominated by an exponential with a larger negative exponent multiplied by the range $\theta_1 - n^{-2/3}$. With the coefficient

$$\frac{1}{\pi\sqrt{n}}\left[\frac{\Gamma\left(\frac{n}{2} + 1\right)}{\Gamma\left(\frac{n+1}{2}\right)}\right] e^{nR},$$

and using the fact that

$$\frac{\Gamma\left(\frac{n}{2} + 1\right)}{\Gamma\left(\frac{n+1}{2}\right)} \sim \sqrt{\frac{n}{2}},$$

our dominant term approaches

$$\frac{\left[G \sin \theta_1 \exp\left(-\frac{A^2}{2} + \tfrac{1}{2}AG \cos \theta_1\right)\right]^n}{\sqrt{n\pi}\sqrt{1+G^2}\sin \theta_1(2 \cos \theta_1 - AG \sin^2 \theta_1)}. \tag{57}$$

Combining this with the previously obtained asymptotic expression (51) for $Q(\theta_1)$ we obtain the following *asymptotic expression for the upper bound on* $P_{e\,\mathrm{opt}}$ *for* $\theta_1 < \theta_c$ *:*

$$\left(1 - \frac{\cos \theta_1 - AG \sin^2 \theta_1}{2 \cos \theta_1 - AG \sin^2 \theta_1}\right) \cdot \frac{\left[G \sin \theta_1 \exp\left(-\frac{A^2}{2} + \tfrac{1}{2}AG \cos \theta_1\right)\right]^n}{\sqrt{n\pi}\sqrt{1+G^2}\sin \theta_1(AG \sin^2 \theta_1 - \cos \theta_1)}. \tag{58}$$

Since our lower bound was asymptotic to the same expression without the parenthesis in front, *the two asymptotes differ only by the factor*

$$\left(1 - \frac{\cos \theta_1 - AG \sin^2 \theta_1}{2 \cos \theta_1 - AG \sin^2 \theta_1}\right)$$

independent of n. This factor increases as θ_1 increases from the value θ_0, corresponding to channel capacity, to the critical value θ_c, for which the denominator vanishes. Over this range the factor increases from 1 to ∞. In other words, for large n, $P_{e\,\mathrm{opt}}$ is determined to within a factor. Furthermore, the percentage uncertainty due to this factor is smaller at rates closer to channel capacity, approaching zero as the rate approaches capacity. It is quite interesting that these seemingly weak bounds can work out to give such sharp information for certain ranges of the variables.

Case 2: $\theta_1 > \theta_c$.

For θ_1 in this range the previous argument does not hold, since the maximum of the exponent is not at the end of the range of integration but rather interior to it. This unique maximum occurs at θ_c, the root of $2 \cos \theta_c - AG \sin^2 \theta_c = 0$. We divide the range of integration into three parts: 0 to $\theta_c - n^{-2/5}$, $\theta_c - n^{-2/5}$ to $\theta_c + n^{-2/5}$ and $\theta_c + n^{-2/5}$ to θ. Proceeding by very similar means, in the neighborhood of θ_c the exponential behaves as

$$\exp\left(-n\left\{E_L(\theta_c) + \frac{(\theta - \theta_c)^2}{2} E''_L(\theta_c) + O[(\theta - \theta_c)^3]\right\}\right).$$

The coefficient of the exponential approaches constancy in the small interval surrounding θ_c. Thus the integral (53) for this part is asymptotic to

$$\frac{1}{\cos\theta_c \sin^3\theta_c \sqrt{1+G^2}} \cdot \int \exp\left\{-n\left[E_L(\theta_c) + \frac{(\theta-\theta_c)^2}{2} E''_L(\theta_c)\right]\right\} d\theta \tag{59}$$
$$\sim \frac{1}{\cos\theta_c \sin^3\theta_c \sqrt{1+G^2}} \exp\left[-nE_L(\theta_c)\right] \frac{\sqrt{2\pi}}{\sqrt{nE''_L(\theta_c)}}.$$

The other two integrals become small by comparison when n is large, by essentially the same arguments as before. They may be dominated by the value of the integrand at the end of the range near θ_c multiplied by the range of integration. Altogether, then, the integral (52) is asymptotic to

$$\frac{1}{\sqrt{\pi n}\cos\theta_c \sin^3\theta_c \sqrt{1+G^2}\sqrt{E''_L(\theta_c)}} e^{-n[E_L(\theta_c)-R]}. \tag{60}$$

The other term in (52), namely, $Q(\theta_1)$, is asymptotically small compared to this, under the present case $\theta > \theta_c$, since the coefficient of n in the exponent for $Q(\theta)$ in (51) will be smaller. Thus, all told, *the random code bound is asymptotic to*

$$\frac{1}{\cos\theta_c \sin^3\theta_c \sqrt{n\pi E''_L(\theta_c)[1+G(\theta_c)^2]}} e^{-n[E_L(\theta_c)-R]} \tag{61}$$

for $\theta > \theta_c$ or for rates $R < R_c$ the rate corresponds to θ_c.

Incidentally, the rate R_c is very closely one-half bit less than channel capacity when $A \geqq 4$, and approaches this exactly as $A \to \infty$. For lower values of A the difference $C - R_c$ becomes smaller but the ratio $C/R_c \to 4$ as $A \to 0$.

VIII. THE FIRM UPPER BOUND ON $P_{e\text{ opt}}$

In this section we will find an upper bound, valid for all n, on the probability of error by manipulation of the upper bound (20). We first find an upper bound on $Q'(\theta)$. In Ref. 6 the integral (35) is transformed into $\bar{z}^{\nu} \exp\left(-\frac{1}{2}\bar{z}^2 + \bar{z}\sqrt{\nu+1}\, w\right)$ times the following integral (in their notation):

$$U = \int_{-\infty}^{\infty} \varphi_{\bar{z}}(y) \exp\left\{-\tfrac{1}{2} y^2 + \nu\left[\ln\left(1+\frac{y}{\bar{z}}\right) - \frac{y}{\bar{z}}\right]\right\} dy.$$

It is pointed out that the integrand here can be dominated by $e^{-y^2/2}$. This occurs in the paragraph in Ref. 6 containing Equation 2.6. Therefore, this integral can be dominated by $\sqrt{2\pi}$, and our integral in (34) involved in $dQ/d\theta$ is dominated as follows:

$$\int_0^\infty \exp\left[-\frac{(r - A\sqrt{n}\cos\theta)^2}{2}\right] r^{n-1}\,dr$$

$$= \left(\frac{\bar{z}}{e}\right)^{n-1} \exp\left(\frac{\bar{z}}{2}\right)^2 \exp\frac{-A^2 n}{2}\cos^2\theta U$$

$$\leqq \left(\frac{\bar{z}}{e}\right)^{n-1} \exp\left(\frac{\bar{z}}{2}\right)^2 \exp\frac{-A^2 n}{2}\cos^2\theta\sqrt{2\pi}.$$

We have

$$\bar{z} = \tfrac{1}{2}\sqrt{n}\,(A\cos\theta + \sqrt{A^2\cos^2\theta + 4 - 4/n}) \leqq \sqrt{n}\,G.$$

Replacing $\bar{z}$ by this larger quantity gives

$$\left(\frac{\sqrt{n}G}{e}\right)^{n-1} \exp\left(\frac{nG^2}{2} - \frac{A^2 n}{2}\cos^2\theta\right)\sqrt{2\pi}.$$

We have, then,

$$-\frac{dQ}{d\theta} \leqq \frac{(n-1)\exp\left(\frac{-A^2 n}{2}\sin^2\theta\right)(\sin\theta)^{n-2}}{2^{n/2}\sqrt{\pi}\,\Gamma\frac{n+1}{2}}\left(\frac{\sqrt{n}\,G}{e}\right)^{n-1} \cdot\exp\left(\frac{nG^2}{2} - \frac{A^2 n}{2}\cos^2\theta\right)\sqrt{2\pi}. \tag{62}$$

Replacing the gamma function by its Stirling expression

$$\left(\frac{n+1}{2}\right)^{n/2} \exp\left(\frac{n+1}{2}\right)\sqrt{2\pi}$$

(which is always too small), and replacing $[1 + (1/n)]^{n/2}$ by $\sqrt{2}$ (which is also too small) again increases the right member. After simplification, we get

$$-\frac{dQ}{d\theta} \leqq \frac{(n-1)(G\sin\theta)^n \exp\left[\left(\frac{n}{2}\right)(-A^2 + 1 + AG\cos\theta)\right]}{\sqrt{n}\,G\sin^2\theta\sqrt{2\pi}\exp\left(\frac{n-3}{2}\right)} \tag{63}$$

$$\leqq \frac{(n-1)e^{3/2}e^{-E_L(\theta)n}}{\sqrt{2\pi n}\,G\sin^2\theta}.$$

Notice that this differs from the asymptotic expression (42) only by a factor

$$\frac{e^{3/2}\sqrt{1+G^2}}{\sqrt{2}\,G} \leqq e^{3/2}$$

(since $G \geqq 1$). A firm upper bound can now be placed on $Q(\theta)$:

$$Q(\theta_1) = \int_{\theta_1}^{\pi/2} \frac{dQ}{d\theta}\,d\theta + Q\left(\frac{\pi}{2}\right).$$

We use the upper bound above for $dQ/d\theta$ in the integral. The coefficient of $-n$ in the exponent of e

$$E_L(\theta) = \tfrac{1}{2}(A^2 - AG\cos\theta) - \log G\sin\theta$$

is positive and monotone increasing with θ for $\theta > \theta_0$, as we have seen previously. Its derivative is

$$E'_L(\theta) = AG\sin\theta - \cot\theta.$$

As a function of θ this curve is as shown in Fig. 5, either rising monotonically from $-\infty$ at $\theta = 0$ to A at $\theta = \pi/2$, or with a single maximum. In any case, the curve is concave downward. To show this analytically, take the second derivative of E'_L. This consists of a sum of negative terms.

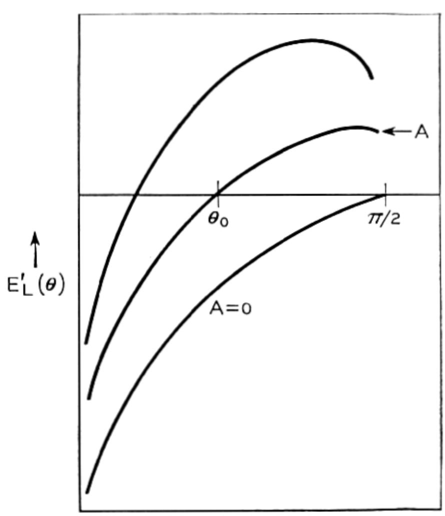

Fig. 5 — $E_L'(\theta)$ as a function of θ.

Returning to our upper bound on Q, the coefficient in (63) does not exceed

$$\frac{\dfrac{\sqrt{n}}{\sqrt{2\pi}}\, e^{3/2}}{\sin^2 \theta_1},$$

replacing $\sin \theta$ and G by $\sin \theta_1$ and 1, their minimum values in the range. We now wish to replace $e^{-nE_L(\theta)}$ by

$$\exp - n[E_L(\theta_1) + (\theta - \theta_1)h].$$

If h is chosen equal to the minimum $E'_L(\theta)$, this replacement will increase the integral and therefore give an upper bound. From the behavior of $E'_L(\theta)$ this minimum occurs at either θ_1 or $\pi/2$. Thus, we may take $h = \min\,[A,\, AG(\theta_1) \sin \theta_1 - \cot \theta_1]$. With this replacement the integral becomes a simple exponential and can be immediately integrated.

The term $Q(\pi/2)$ is, of course,

$$\Phi(-A\sqrt{n}) \leqq \frac{1}{\sqrt{2\pi n}\, A}\, e^{-A^2 n/2}.$$

If we continue the integral out to infinity instead of stopping at $\pi/2$, the extra part added will more than cover $Q(\pi/2)$. In fact, $E_L(\pi/2) = A^2/2$, so the extra contribution is at least

$$\frac{\sqrt{n}\, e^{3/2}}{An \sin^2 \theta_1 \sqrt{2\pi}}\, e^{-A^2 n/2},$$

if we integrate

$$\frac{\sqrt{n}\, e^{3/2}}{\sin^2 \theta_1 \sqrt{2\pi}}\, e^{-A^2 n/2 - n(\theta - \theta_1)A}$$

to ∞ instead of stopping at $\pi/2$. Since $e^{3/2}/\sin^2 \theta_1 \geqq 1$, we may omit the $Q(\pi/2)$ term in place of the extra part of the integral.

Consequently, we can *bound* $Q(\theta_1)$ *as follows*:

$$Q(\theta_1) \leqq \frac{e^{3/2} \exp \{(n/2)\,[AG(\theta_1) \cos \theta_1 - A^2 + 2 \log G \sin \theta_1]\}}{\sqrt{2\pi n}\, \sin^2 \theta_1 \min\,(A,\, AG(\theta_1) \sin \theta_1 - \cot \theta_1)}. \qquad (64)$$

In order to overbound $P_{e\,\mathrm{opt}}$ by (3) it is now necessary to overbound the term

$$\int_0^{\theta_1} \frac{\Omega(\theta)}{\Omega(\theta_1)}\, dQ(\theta).$$

This can be done by a process very similar to that just carried out for $\int dQ(\theta)$. First, we overbound $\Omega(\theta)/\Omega(\theta_1)$ using (21). We have

$$\frac{\Omega(\theta)}{\Omega(\theta_1)} = \frac{\int_0^\theta (\sin x)^{n-2}\, dx}{\int_0^{\theta_1} (\sin x)^{n-2}\, dx}$$

$$= \frac{\int_0^\theta (\sin x)^{n-2}\, dx}{\int_0^\theta (\sin x)^{n-2}\, dx + \int_\theta^{\theta_1} (\sin x)^{n-2}\, dx}$$

$$\leqq \frac{\int_0^\theta (\sin x)^{n-2} \cos x\, dx}{\int_0^\theta (\sin x)^{n-2} \cos x\, dx + \cos\theta \int_\theta^{\theta_1} (\sin x)^{n-2}\, dx}$$

$$\leqq \frac{\int_0^\theta (\sin x)^{n-2} \cos x\, dx}{\int_0^\theta (\sin x)^{n-2} \cos x\, dx + \int_\theta^{\theta_1} (\sin x)^{n-2} \cos x\, dx},$$

and, finally,

$$\frac{\Omega(\theta)}{\Omega(\theta_1)} \leqq \frac{(\sin\theta)^{n-1}}{(\sin\theta_1)^{n-1}}. \tag{65}$$

Here the third line follows since the first integral in the denominator is reduced by the same factor as the numerator and the second integral is reduced more, since $\cos\theta$ is decreasing. In the next line, the denominator is reduced still more by taking the cosine inside.

Using this inequality and also the upper bound (63) on $dQ/d\theta$, we have

$$\int_0^{\theta_1} \frac{\Omega(\theta)}{\Omega(\theta_1)}\, dQ(\theta) \leqq$$

$$\int_0^{\theta_1} \frac{(\sin\theta)^{n-1}}{(\sin\theta_1)^{n-1}} \frac{(n-1)e^{3/2}(G\sin\theta)^n e^{(n/2)(-A^2+A\cos\theta G)}}{\sqrt{2\pi n}G\sin^2\theta}\, d\theta \tag{66}$$

$$= \frac{(n-1)e^{3/2}}{\sqrt{2\pi n}\,(\sin\theta_1)^{n-1}} \int_0^{\theta_1} G^n (\sin\theta)^{2n-3} e^{(n/2)(-A^2+A\cos\theta G)}\, d\theta.$$

Near the point θ_1 the integrand here behaves like an exponential when n is large (provided $\theta_1 < \theta_c$), and it should be possible to find a firm upper bound of the form

$$\frac{k}{\sqrt{n}} e^{-E_L(\theta_1)n},$$

where k would not depend on n. This, however, leads to considerable complexity and we have settled for a cruder formulation as follows:

The integrand may be bounded by its maximum values. If $\theta_1 < \theta_c$, the maximum of the integrand will occur at θ_1, at least when n is large enough. In this case, the integral will certainly be bounded by

$$\theta_1 G^n(\theta_1)(\sin\theta_1)^{2n-3} e^{(n/2)[-A^2+A\cos\theta_1 G(\theta_1)]}.$$

The entire expression for $P_{e\,\text{opt}}$ may then be bounded by [adding in the bound (64) on $Q(\theta_1)$]

$$P_{e\,\text{opt}} \leqq \frac{\sqrt{n}e^{3/2}\,\theta_1 e^{-E_L(\theta_1)}}{\sqrt{2\pi}\sin^2\theta_1}\left\{1 + \frac{1}{n\theta_1 \min\,[A, AG(\theta_1)\sin\theta_1 - \cot\theta_1]}\right\}, \quad (67)$$

It must be remembered that (67) is valid only for $\theta_1 < \theta_c$ and if n is large enough to make the maximum of the integrand above occur at θ. For $\theta_1 > \theta_c$, bounds could also be constructed based on the maximum value of the integrand.

IX. A FIRM LOWER BOUND ON $P_{e\,\text{opt}}$

In this section we wish to find a lower bound on $P_{e\,\text{opt}}$ that is valid for all n. To do this we first find a lower bound on $Q'(\theta)$ and from this find a lower bound on $Q(\theta)$. The procedure is quite similar to that involved in finding the firm upper bound.

In Ref. 6, the integral (35) above was reduced to the evaluation of the following integral (Equation 2.5 of Ref. 6):

$$\begin{aligned}
&\int_{-\bar{z}}^{\infty}\left(1 + \frac{y}{\bar{z}}\right)^{\nu} \exp\left(-\tfrac{1}{2}y^2 - y\frac{\nu}{\bar{z}}\right) dy \\
&\qquad \geqq \int_0^{\infty} \exp\left\{-\tfrac{1}{2}y^2 + \nu\left[\ln\left(1 + \frac{y}{\bar{z}}\right) - \frac{y}{\bar{z}}\right]\right\} dy \\
&\qquad \geqq \int_0^{\infty} \exp\left[-\tfrac{1}{2}y^2 + \nu\left(\frac{-y^2}{2\bar{z}^2}\right)\right] dy \\
&\qquad = \int_0^{\infty} \exp\left[\frac{-y^2}{2}\left(1 + \frac{\nu}{\bar{z}^2}\right)\right] dy = \frac{1}{2}\frac{\sqrt{2\pi}}{\sqrt{1 + \frac{\nu}{\bar{z}^2}}} \\
&\qquad \geqq \frac{\sqrt{2\pi}}{2\sqrt{2}} = \frac{\sqrt{\pi}}{2}
\end{aligned}$$

Here we used the inequality

$$\ln\left(1+\frac{y}{\bar z}\right)-\frac{y}{\bar z}\geqq-\frac{y^2}{2\bar z^2}\qquad\text{for}\qquad\frac{y}{\bar z}>0,$$

and also the fact that $\nu/\bar z^2 \leqq 1$. This latter follows from Equation 2.3 of Ref. 6 on dividing through by $\bar z^2$.

Using this lower bound, we obtain from (34)

$$\frac{dQ}{d\theta}\geqq\frac{(n-1)\sin^{n-2}\theta\exp\left(\frac{-A^2n}{2}\right)}{2^{n/2}\sqrt{\pi}\,\Gamma\left(\frac{n+1}{2}\right)}\left(\frac{\bar z}{e}\right)^{n-1}\exp\left(\frac{\bar z^2}{2}\right)\frac{\sqrt{\pi}}{2}.\qquad(68)$$

Now $\bar z \geqq \sqrt{n-1}\;G$ and

$$\Gamma\left(\frac{n+1}{2}\right)<\left(\frac{n+1}{2}\right)^{n/2}e^{-(n+1)/2}\sqrt{2\pi}\exp\left[\frac{1}{6(n+1)}\right]$$

and, using the fact that

$$\left(\frac{n-1}{n+1}\right)^{n/2}\geqq\frac{1}{3}\qquad\text{for}\qquad n\geqq 2,$$

we obtain

$$\frac{dQ}{d\theta}\geqq\frac{1}{6\sqrt{2\pi}}\frac{\sqrt{n-1}\,e^{3/2}\,e^{-nE_L(\theta)}}{G\exp\left[\frac{G^2}{2}+\frac{1}{6(n+1)}\right]\sin^2\theta}\qquad\text{for } n\geqq 2.\qquad(69)$$

This is our lower bound on $dQ/d\theta$.

To obtain a lower bound on $Q(\theta)$ we may use the same device as before—here, however, replacing the coefficient by its minimum value in the range and the exponent by $-nE_L(\theta_1)-n(\theta-\theta_1)E'_{L\,\max}$:

$$\begin{aligned}E'_L&=AG\sin\theta-\cot\theta\\&\leqq AG\\&\leqq A(A+1).\end{aligned}$$

Similarly, in the coefficient, G can be dominated by $A+1$ and $\sin^2\theta$ by 1. Thus,

$$Q(\theta_1)\geqq$$

$$\int_{\theta_1}^{\pi/2}\frac{\sqrt{n-1}e^{3/2}e^{-nE_L(\theta_1)}e^{-n(\theta-\theta_1)A(A+1)}}{6\sqrt{2\pi}(A+1)\exp\left[\frac{(A+1)^2}{2}+\frac{1}{6(n+1)}\right]}\,d\theta+Q\left(\frac{\pi}{2}\right).\qquad(70)$$

Integrating and observing that the term due to the $\pi/2$ limit can be absorbed into the $Q(\pi/2) - \operatorname{erf} A$, we arrive at the lower bound:

$$Q(\theta_1) \geqq \frac{\sqrt{n-1}\, e^{3/2} e^{-nE_L(\theta_1)}}{6\sqrt{2\pi n}\,(A+1)^3 \exp\left[\dfrac{(A+1)^2}{2} + \dfrac{1}{6(n+1)}\right]}. \tag{71}$$

X. BEHAVIOR NEAR CHANNEL CAPACITY

As we have seen, near channel capacity the upper and lower asymptotic bounds are substantially the same. If in the asymptotic lower bound (42) we form a Taylor expansion for θ near θ_0, retaining terms up to $(\theta - \theta_0)^2$, we will obtain an expression applying to the neighborhood of channel capacity. Another approach is to return to the original noncentral t-distribution and use its normal approximation which will be good near the mean (see Ref. 5). Either approach gives, in this neighborhood, the approximations [since $E(\theta_0) = E'(\theta_0) = 0$]:

$$\begin{aligned} -\frac{dQ}{d\theta} &\doteq \frac{\sqrt{n}\,(1+A^2)}{\sqrt{\pi}\,\sqrt{2+A^2}} \exp\left[-n\,\frac{(A^2+1)^2}{A^2+2}\,(\theta-\theta_0)^2\right] \\ Q(\theta) &\doteq \Phi\left[(\theta_0-\theta)\,\frac{A^2+1}{\sqrt{A^2+2}}\,\sqrt{2n}\right], \end{aligned} \tag{72}$$

or, since near channel capacity, using $e^{-R} \doteq \sin\theta$,

$$\theta - \theta_0 \doteq A^{-1}(C-R)$$

$$\begin{aligned} P_{e\,\mathrm{opt}}\left(n, R, \sqrt{\frac{P}{N}}\right) &\doteq \Phi\left[\sqrt{2n}\,A^{-1}\,\frac{A^2+1}{\sqrt{A^2+2}}\,(R-C)\right] \\ &= \Phi\left[\frac{P+N}{\sqrt{P(P+2N)}}\,\sqrt{2n}\,(R-C)\right]. \end{aligned} \tag{73}$$

The reliability curve is approximated near C by

$$E(R) \doteq \frac{(P+N)^2}{P(P+2N)}\,(C-R)^2. \tag{74}$$

It is interesting that Rice[2] makes estimates of the behavior of what amounts to a lower bound on the exponent E near channel capacity. His exponent, translated into our notation, is

$$E^*(R) \doteq \frac{P+N}{2P}\,(C-R)^2,$$

a poorer value than (74); that is, it will take a larger block length to achieve the same probability of error. This difference is evidently due to the slight difference in the manner of construction of the random codes. Rice's codes are obtained by placing points according to an n-dimensional gaussian distribution, each coordinate having variance P. In our codes the points are placed at random on a sphere of precisely fixed radius $\sqrt{nP}$. These are very close to the same thing when n is large, since in Rice's situation the points will, with probability approaching 1, lie between the spheres of radii $\sqrt{nP}\,(1 - \epsilon)$ and $\sqrt{nP}\,(1 + \epsilon)$, (any $\epsilon > 0$). However, we are dealing with very small probability events in any case when we are estimating probability of error, and the points within the sphere are sufficiently important to affect the exponent E. In other words, the Rice type of code is sufficient to give codes that will have a probability of error approaching zero at rates arbitrarily near channel capacity. However, they will not do so at as rapid a rate (even in the exponent) as can be achieved. To achieve the best possible E it is evidently necessary to avoid having too many of the code points interior to the $\sqrt{nP}$ sphere.

At rates R greater than channel capacity we have $\theta_1 < \theta_0$. Since the Q distribution approaches normality with mean at θ_0 and variance $2n(A^2 + 1)^2/(A^2 + 2)$, we will have $Q(\theta_1)$ approaching 1 with increasing n for any fixed rate greater than C. Indeed, even if the rate R varies but remains always greater than C (perhaps approaching it from above with increasing n), we will still have $P_{\epsilon\,\mathrm{opt}} > \frac{1}{2} - \epsilon$ for any $\epsilon > 0$ and sufficiently large n.

XI. UPPER BOUND ON $P_{\epsilon\,\mathrm{opt}}$ BY METHOD OF EXHAUSTION

For low rates of transmission, where the upper and lower bounds diverge widely, we may obtain better estimates by other methods. For very low rates of transmission, the main contribution to the probability of error can be shown to be due to the code points that are nearest together and thus often confused with each other, rather than to the general average structure of the code. The important thing, at low rates, is to maximize the minimum distance between neighbors. Both the upper and lower bounds which we will derive for low rates are based on these considerations.

We will first show that, for $D \leqq \sqrt{2\,nP}$, it is possible to find at least

$$M_D = \left(\sin 2 \sin^{-1} \frac{D}{2\sqrt{nP}}\right)^{1-n}$$

points on the surface of an n sphere of radius $\sqrt{nP}$ such that no pair of them is separated by a distance less than D. (If M_D is not an integer, take the next larger integer.) The method used will be similar to one used by E. N. Gilbert for the binary symmetric channel.

Select any point on the sphere's surface for the first point. Delete from the surface all points within D of the selected point. In Fig. 6, x is the selected point and the area to be deleted is that cut out by the cone. This area is certainly less (if $D \leqq \sqrt{2nP}$) than the area of the hemisphere of radius H shown and, even more so, less than the area of the sphere of radius H. If this deletion does not exhaust the original sphere, select any point from those remaining and delete the points within D of this new point. This again will not take away more area than that of a sphere of radius H. Continue in this manner until no points remain. Note that each point chosen is at least D from each preceding point. Hence all interpoint distances are at least D. Furthermore, this can be continued at least as many times as the ratio of the surface of a sphere of radius $\sqrt{nP}$ to that of a sphere of radius H, since each deletion takes away not more than this much surface area. This ratio is clearly

$$(\sqrt{nP}/H)^{n-1}.$$

By simple geometry in Fig. 6, we see that H and D are related as follows:

$$\sin \theta = \frac{H}{\sqrt{nP}},$$

$$\sin \frac{\theta}{2} = \frac{D}{2\sqrt{nP}}.$$

Hence

$$H = \sqrt{nP} \sin 2 \sin^{-1} \frac{D}{2\sqrt{nP}}. \tag{75}$$

Substituting, *we can place at least*

$$M_D = \left(\sin 2 \sin^{-1} \frac{D}{2\sqrt{nP}}\right)^{-(n-1)}$$

points at distances at least D from each other, for any $D \leqq \sqrt{2nP}$.

If we have M_D points with minimum distance at least D, then the probability of error with optimal decoding will be less than or equal to

$$M_D \Phi\left(\frac{-D}{2\sqrt{N}}\right).$$

To show this we may add up pessimistically the probabilities of each point being received as each other point. Thus the probability of point 1 being moved closer to point 2 than to the original point 1 is not greater than $\Phi[-D/(2\sqrt{N})]$, that is, the probability of the point being moved in a certain direction at least $D/2$ (half the minimum separation). The contribution to errors due to this cause cannot, therefore, exceed $(1/M_D)\Phi[-D/(2\sqrt{N})]$, (the $1/M_D$ factor being the probability of message 1 being transmitted). A similar argument occurs for each (ordered) pair of points, a total of $M_D(M_D - 1)$ contributions of this kind. Consequently, the probability of error cannot exceed $(M_D - 1)\Phi[-D/(2\sqrt{N})]$ or, more simply, $M_D\Phi[-D/(2\sqrt{N})]$.

If we set

$$e^{nR} = M_D = \left(\sin 2 \sin^{-1} \frac{D}{2\sqrt{nP}}\right)^{-(n-1)}$$

then the rate R (in natural units) is

$$R = \left(1 - \frac{1}{n}\right) \log \left(\sin 2 \sin^{-1} \frac{D}{2\sqrt{nP}}\right)^{-1}$$

with

$$P_e \leqq e^{nR}\Phi\left(\frac{-D}{2\sqrt{N}}\right) \leqq e^{nR} \frac{\sqrt{2N}}{D\sqrt{\pi}} e^{-(D^2/8N)}, \tag{76}$$

using the well-known upper bound $\Phi(-x) \leqq (1/x\sqrt{2\pi})e^{-x^2/2}$. These are

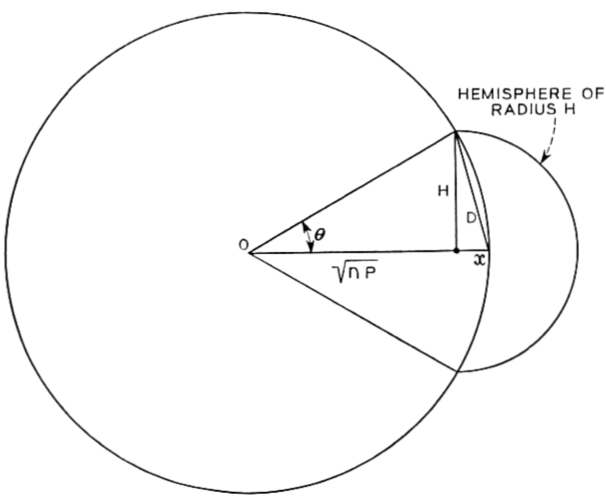

Fig. 6 — Geometry of sphere of radius $\sqrt{nP}$.

parametric equations in terms of D. It is more convenient to let

$$D = \lambda\sqrt{2nP}.$$

We then have

$$\begin{aligned} R &= \left(1 - \frac{1}{n}\right)\log\left(\sin 2\sin^{-1}\frac{\lambda}{\sqrt{2}}\right)^{-1}, \\ P_e &\leqq \frac{1}{\lambda\sqrt{\pi n\dfrac{P}{N}}}\, e^{n[R-(\lambda^2P)/(4N)]}. \end{aligned} \tag{77}$$

The asymptotic reliability, that is, the coefficient of $-n$ in the exponent of P_e, is given by $(\lambda^2 P/4N) - R$. This approaches

$$(\sin \tfrac{1}{2} \sin^{-1} e^R)^2 \frac{P}{2N} - R \qquad \text{as} \qquad n \to \infty.$$

Thus our asymptotic *lower* bound for reliability is (eliminating λ):

$$E \geqq (\sin \tfrac{1}{2} \sin^{-1} e^R)^2 \frac{P}{2N} - R. \tag{78}$$

As $R \to 0$ the right-hand expression approaches $P/(4N)$.

This lower bound on the exponent is plotted in the curves in Section XIV and it may be seen to give more information at low rates than the random code bound. It is possible, however, to improve the random coding procedure by what we have called an "expurgating" process. It then becomes the equal of the bound just derived and, in fact, is somewhat stronger over part of the range. We shall not go into this process in detail but only mention that the expurgating process consists of eliminating from the random code ensemble points which have too close neighbors, and working with the codes that then remain.

XII. LOWER BOUND ON P_e IN GAUSSIAN CHANNEL BY MINIMUM DISTANCE ARGUMENT

In a code of length n with M code words, let $m_{is}(i = 1, 2, \cdots, M; s = 1, 2, \cdots, n)$ be the sth coordinate of code word i. We are here assuming an average power limitation P, so that

$$\frac{1}{nM} \sum_{i,s} m_{is}^2 \leqq P. \tag{79}$$

We also assume an independent gaussian noise of power N added to each coordinate.

We now calculate the average squared distance between all the $M(M-1)/2$ pairs of points in n-space corresponding to the M code words. The squared distance from word i to word j is

$$\sum_s (m_{is} - m_{js})^2.$$

The average $\overline{D^2}$ between all pairs will then be

$$\overline{D^2} = \frac{1}{M(M-1)} \sum_{s,i,j} (m_{is} - m_{js})^2.$$

Note that each distance is counted twice in the sum and also that the extraneous terms included in the sum, where $i = j$, contribute zero to it. Squaring the terms in the sum,

$$\begin{aligned} \overline{D^2} &= \frac{1}{M(M-1)} \left(\sum_{i,j,s} m_{is}^2 - 2\sum_s \sum_{i,j} m_{is}m_{js} + \sum_{i,j,s} m_{js}^2\right) \\ &= \frac{1}{M(M-1)} \left[2M \sum_{i,s} m_{is}^2 - 2\sum_s \left(\sum_i m_{is}\right)^2\right] \\ &\leqq \frac{1}{M(M-1)}\, 2MPnM \\ \overline{D^2} &\leqq \frac{2nMP}{M-1}, \end{aligned} \tag{80}$$

where we obtain the third line by using the inequality on the average power (79) and by noting that the second term is necessarily non-positive.

If the *average* squared distance between pairs of points is

$$\leqq (2nMP)/(M-1),$$

there must exist a pair of points for whose distance this inequality holds. Each point in this pair is used $1/M$ of the time. The best detection for separating this pair (if no other points were present) would be by a hyperplane normal to and bisecting the joining line segment. Either point would then give rise to a probability of error equal to that of the noise carrying a point half this distance or more in a specified direction. We obtain, then, a contribution to the probability of error at least

$$\frac{1}{M} \cdot \Pr\left\{\text{noise in a certain direction} \geqq \frac{1}{2}\sqrt{\frac{2nMP}{M-1}}\right\}$$

$$= \frac{1}{M}\,\Phi\left[-\sqrt{\frac{nMP}{(M-1)2N}}\right].$$

This we may assign to the first of the two points in question, and the errors we have counted are those when this message is sent and is received closer to the second message (and should therefore be detected as the second or some other message).

Now delete this first message from the set of code points and consider the remaining $M - 1$ points. By the same argument there must exist among these a pair whose distance is less than or equal to

$$\sqrt{\frac{2nP(M-1)}{(M-2)}}$$

This pair leads to a contribution to probability of error, due to the first of these being displaced until nearer the second, of an amount

$$\frac{1}{M}\Phi\left[-\sqrt{\frac{(M-1)nP}{(M-2)2N}}\right].$$

This same argument is continued, deleting points and adding contributions to the error, until only two points are left. Thus we obtain a lower bound on $P_{e\,\mathrm{opt}}$ as follows:

$$P_{e\,\mathrm{opt}} \geqq \frac{1}{M}\left[\Phi\left(-\sqrt{\frac{nP}{2N}\frac{M}{M-1}}\right) + \Phi\left(-\sqrt{\frac{nP}{2N}\frac{M-1}{M-2}}\right) + \cdots + \Phi\left(-\sqrt{\frac{nP}{2N}\frac{2}{1}}\right)\right]. \tag{81}$$

To simplify this bound somewhat, one may take only the first $M/2$ terms [or $(M + 1)/2$ if M is odd]. Since they are decreasing, each term would be reduced by replacing it with the last term taken. Thus we may reduce the bound by these operations and obtain

$$P_{e\,\mathrm{opt}} \geqq \frac{1}{2}\Phi\left(-\sqrt{\frac{M}{M-2}\frac{nP}{2N}}\right). \tag{82}$$

For any rate $R > 0$, as n increases the term $M/(M - 2)$ approaches 1 and the bound, then, behaves about as

$$\frac{1}{2}\Phi\left(-\sqrt{\frac{nP}{2N}}\right).$$

This is asymptotic to

$$\frac{1}{2\sqrt{\frac{\pi nP}{N}}}e^{-(nP)/(4N)}.$$

It follows that the reliability $E \leqq P/(4N) = A^2/4$. This is the same value as the lower bound for E when $R \to 0$.

XIII. ERROR BOUNDS AND OTHER CONDITIONS ON THE SIGNAL POINTS

Up to now we have (except in the last section) assumed that all signal points were required to lie on the surface of the sphere, i.e., have a mean square value $\sqrt{nP}$. Consider now the problem of estimating $P'_{e\,\mathrm{opt}}(M, n, \sqrt{P/N})$, where the signal points are only required to lie on or within the spherical surface. Clearly, since this relaxes the conditions on the code, it can only improve, i.e., decrease the probability of error for the best code. Thus $P'_{e\,\mathrm{opt}} \leqq P_{e\,\mathrm{opt}}$.

On the other hand, we will show that

$$P'_{e\,\mathrm{opt}}\left(M, n, \sqrt{\frac{P}{N}}\right) \geqq P_{e\,\mathrm{opt}}\left(M, n + 1, \sqrt{\frac{P}{N}}\right). \qquad (83)$$

In fact, suppose we have a code of length n, all points on or within the

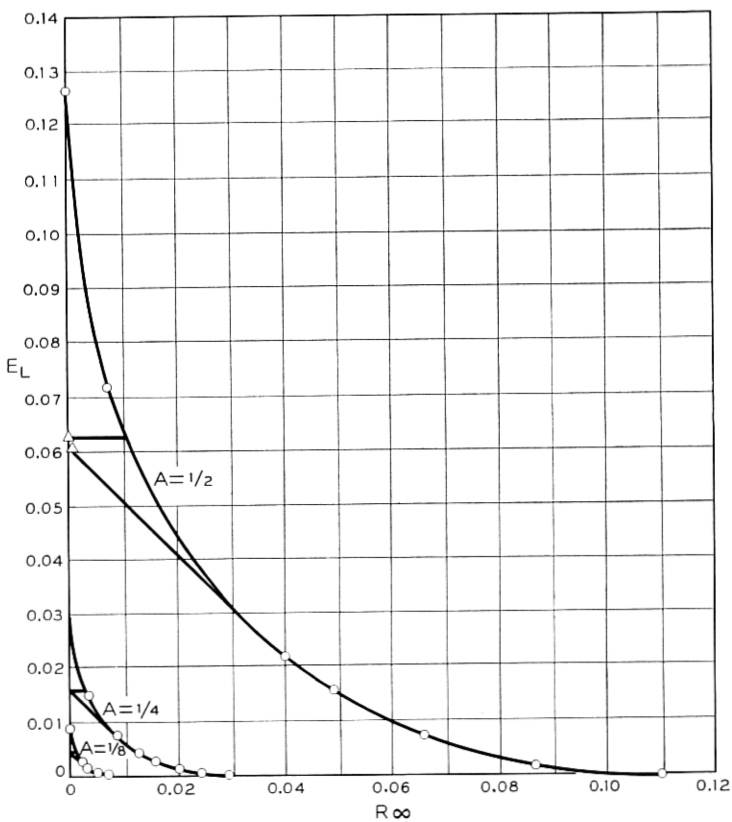

Fig. 7 — Curves showing E_L vs. R for $A = \frac{1}{8}, \frac{1}{4}$ and $\frac{1}{2}$.

n sphere. To each code word add a further coordinate of such value that in the $n + 1$ space the point thus formed lies *exactly on* the $n + 1$ sphere surface. If the first n coordinates of a point have values $x_1, x_2, \cdots, x_n$ with

$$\sum_{i=1}^{n} x_i^2 \leqq nP,$$

the added coordinate will have the value

$$x_{n+1} = \sqrt{(n+1)P - \sum_{i=1}^{n} x_i^2}.$$

This gives a derived code of the first type (all points *on* the $n + 1$ sphere surface) with M words of length $n + 1$ at signal-to-noise ratio P/N. The probability of error for the given code is at least as great as that of the derived code, since the added coordinate can only improve

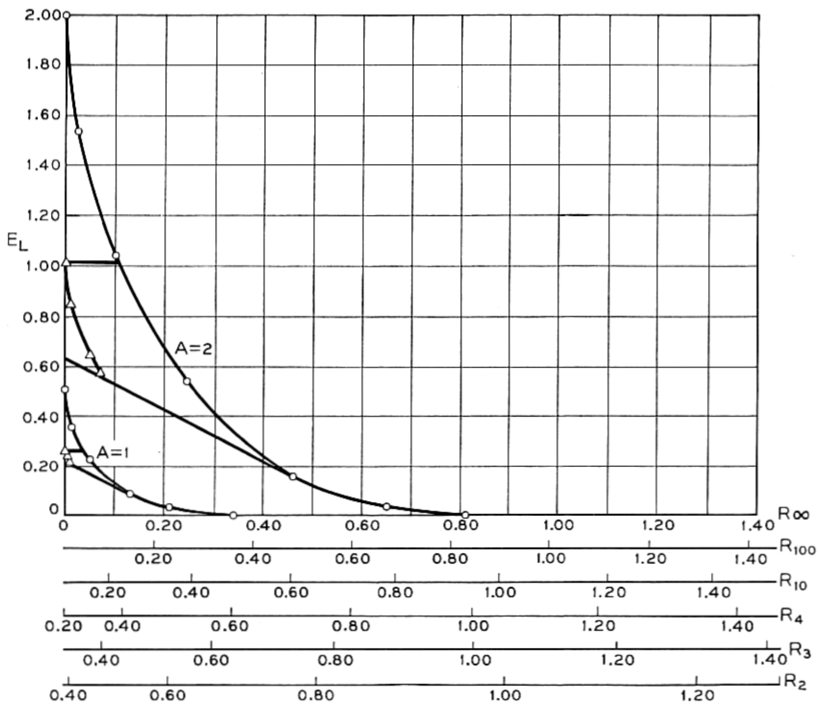

Fig. 8 — Curves showing E_L vs. different values of R for $A = 1$ and 2.

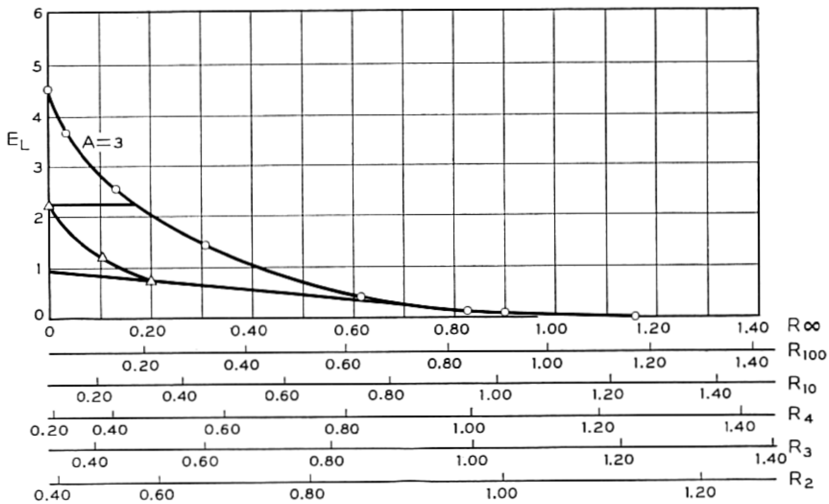

Fig. 9 — Curves showing E_L vs. different values of R for $A = 3$.

the decoding process. One might, for example, decode ignoring the last coordinate and then have the same probability of error. Using it in the best way would, in general, improve the situation.

The probability of error for the derived code of length $n + 1$ must be greater than or equal to that of the optimal code of the length $n + 1$ with all points on the surface. Consequently we have (83). Since $P_{e\,\mathrm{opt}}(M, n, \sqrt{P/N})$ varies essentially exponentially with n when n is large, the effect of replacing n by $n + 1$ is essentially that of a constant multiplier. Thus, our upper bounds on $P_{e\,\mathrm{opt}}$ are not changed and our lower bounds are multiplied by a quantity which does not depend much on n when n is large. The asymptotic reliability curves consequently will be the same. Thus the E curves we have plotted may be applied in either case.

Now consider the third type of condition on the points, namely, that the *average* squared distance from the origin of the set of points be less than or equal to nP. This again is a weakening of the previous conditions and hence the optimal probability of error, $P''_{e\,\mathrm{opt}}$, is less than or equal to that of the previous cases:

$$P''_{e\,\mathrm{opt}}\left(M, n, \frac{P}{N}\right) \leqq P'_{e\,\mathrm{opt}}\left(M, n, \frac{P}{N}\right) \leqq P_{e\,\mathrm{opt}}\left(M, n, \frac{P}{N}\right). \quad (84)$$

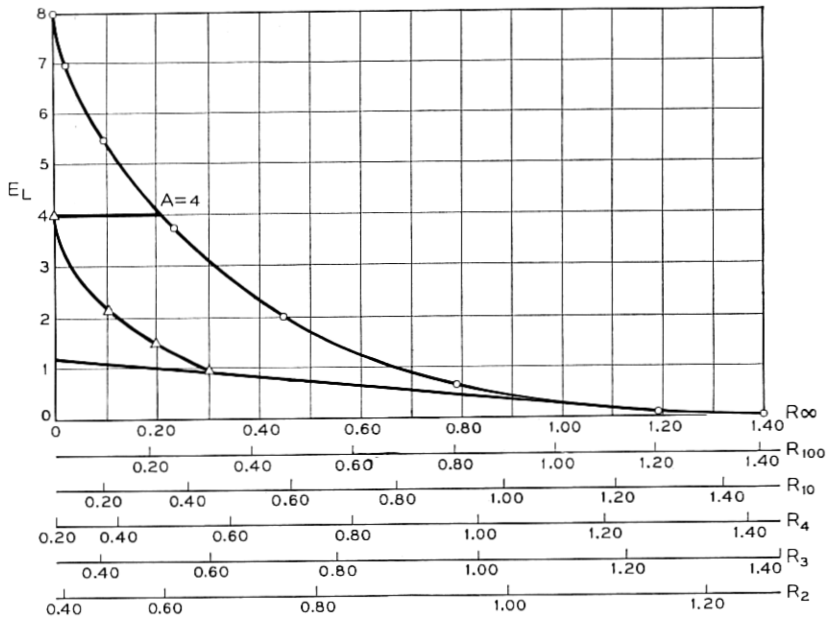

Fig. 10 — Curves showing E_L vs. different values of R for $A = 4$.

Our upper bounds on probability of error (and, consequently, lower bounds on reliability) can be used as they stand.

Lower bounds on $P''_{e\ \mathrm{opt}}$ may be obtained as follows. If we have M points whose mean square distance from the origin does not exceed nP, then for any $\alpha (0 < \alpha \leqq 1)$ at least αM of the points are within a sphere of squared radius $nP/(1 - \alpha)$. [For, if more than $(1 - \alpha)M$ of them were outside the sphere, these alone would contribute more than

$$(1 - \alpha)MnP/(1 - \alpha)$$

to the total squared distance, and the mean would then necessarily be greater than nP.] Given an optimal code under the third condition, we can construct from it, by taking αM points within the sphere of radius $\sqrt{nP/1 - \alpha}$, a code satisfying the second condition with this smaller number of points and larger radius. The probability of error for the new code cannot exceed $1/\alpha$ times that of the original code. (Each new code word is used $1/\alpha$ times as much; when used, its probability of error is at least as good as previously.) Thus:

$$P''_{e\,\mathrm{opt}}\left(M,\,n,\,\sqrt{\frac{P}{N}}\right) \geqq \frac{1}{\alpha}\,P'_{e\,\mathrm{opt}}\left(\alpha M,\,n,\,\sqrt{\frac{P}{(1-\alpha)N}}\right)$$

$$\geqq \frac{1}{\alpha}\,P_{e\,\mathrm{opt}}\left(\alpha M,\,n+1,\,\sqrt{\frac{P}{(1-\alpha)N}}\right).$$

XIV. CURVES FOR ASYMPTOTIC BOUNDS

Curves have been calculated to facilitate evaluation of the exponents in these asymptotic bounds. The basic curves range over values of

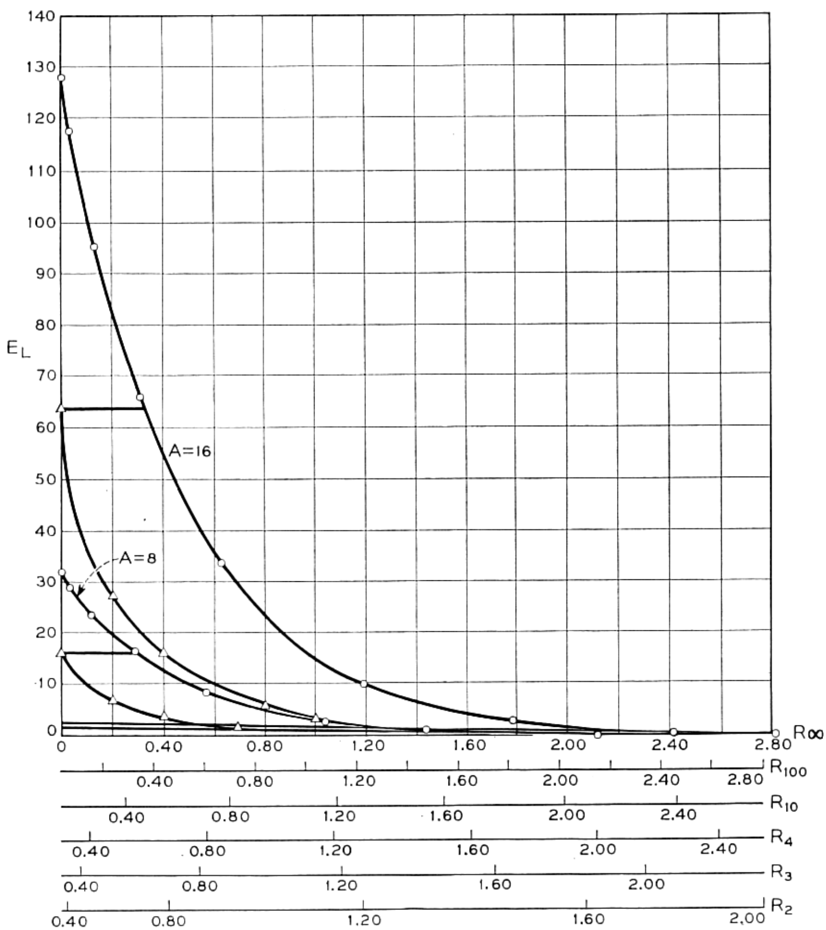

Fig. 11 — Curves showing E_L vs. different values of R for $A = 8$ and 16.

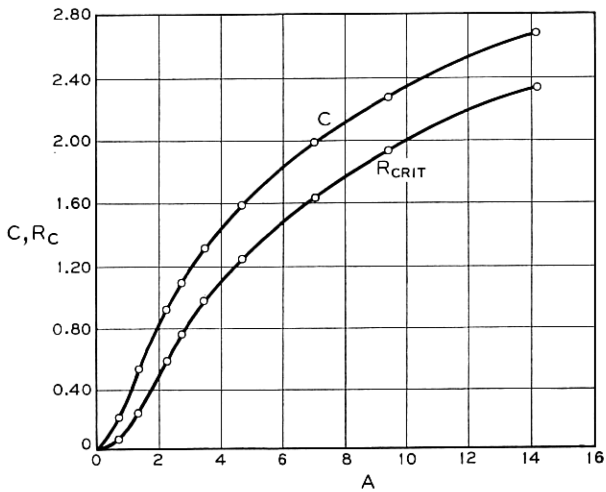

Fig. 12 — Channel capacity, C, and critical rate, R_c, as functions of θ.

$A = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2, 3, 4, 8, 16$. Figs. 7 through 11 give the coefficients of n and E_L as functions of the rate R. Since E_L strictly is a function of θ, and the relation between θ and R depends somewhat on n, a number of slightly different R scales are required at the bottom of the curve. This, however, was considered a better means of presenting the data than the use of auxiliary curves to relate R and θ. These same curves give the coefficient of n in the upper bounds (the straight line part together with the curve to the right of the straight line segment). The point of tangency is the critical R (or critical θ). In other words, the curve and the curve plus straight line, read against the $n = \infty$ scale, give upper and lower bounds on the reliability measure. The upper and lower bounds on E for low R are also included in these curves. The upper bound is the horizontal line segment running out from $R = 0$, $E = A^2/4$. The lower bound is the curved line running down from this point to the tangent line. Thus, the reliability E lies in the four-sided figure defined by these lines to the left of R_c. It is equal to the curve to the right of R_c. Fig. 12 gives channel capacity C and the critical rate R_c as functions of θ. For A very small, the $E_L(R)$ curve approaches a limiting form. In fact, if $\theta = (\pi/2) - \epsilon$, with ϵ small, to a close approximation by obvious expansions we find

$$E_L(R) \doteq \frac{A^2}{2} - A\epsilon + \frac{\epsilon^2}{2} \quad \text{and} \quad R \doteq \frac{\epsilon^2}{2}.$$

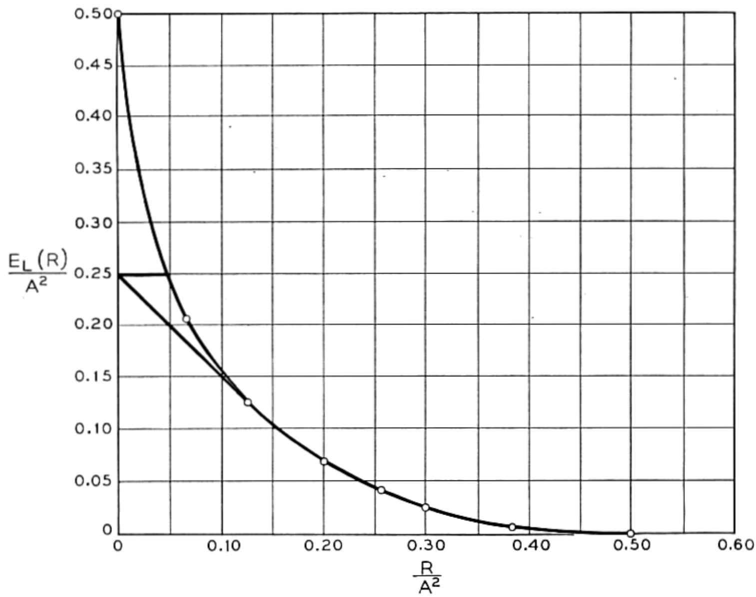

Fig. 13 — Plots of $E_L(R)/A^2$ against R/A^2.

Eliminating ϵ, we obtain

$$\frac{E_L(R)}{A^2} \doteq \frac{1}{2} - \sqrt{\frac{2R}{A^2}}.$$

Fig. 13 plots $E_L(R)/A^2$ against R/A^2.

XV. ACKNOWLEDGMENTS

I am grateful to several people for help in preparing this paper. Mrs. Judy Frankman computed the curves of E_L and other members of the Center for Advanced Study in the Behavioral Sciences were helpful in many ways. The referee made several valuable suggestions which have been incorporated in the paper. Finally, I am particularly indebted to my wife Betty for checking much of the algebra involved in the asymptotic bounds.

REFERENCES

1. Shannon, C. E., Communication in the Presence of Noise, Proc. I.R.E., **37,** January 1949, p. 10.

2. Rice, S. O., Communication in the Presence of Noise — Probability of Error for Two Encoding Schemes, B.S.T.J., **29,** January 1950, p. 60.
3. Elias, P., in *Information Theory* (Cherry, C., ed.), Academic Press, New York, 1956.
4. Shannon, C. E., Certain Results in Coding Theory for Noisy Channels, Inform. and Cont., **1,** September 1957, p. 6.
5. Johnson, N. L. and Welch, B. L., Applications of the Noncentral t-Distribution, Biometrika, **31,** 1939, p. 362.
6. David, H. T. and Kruskal, W. H., The WAGR Sequential t-Test Reaches a Decision with Probability One, Ann. Math. Stat., **27,** September 1956, p. 797