Poisson, Shannon, and the Radio Amateur*

J. P. COSTAS†, SENIOR MEMBER, IRE

Summary—**Congested band operation as found in the amateur service presents an interesting problem in analysis which can only be solved by statistical methods. Consideration is given to the relative merits of two currently popular modulation techniques, SSB and DSB. It is found that in spite of the bandwidth economy of SSB this system can claim no over-all advantage with respect to DSB for this service. It is further shown that there are definite advantages to the use of very broadband techniques in the amateur service.**

The results obtained from the analysis of the radio amateur service are significant, for they challenge the intuitively obvious and universally accepted thesis that congestion in the radio frequency spectrum can only be relieved by the use of progressively smaller transmission bandwidths obtained by appropriate coding and modulation techniques. In order to study the general problem of spectrum utilization, some basic results of information theory are required Some of the significant work of Shannon is reviewed with special emphasis on his channel capacity formula. It is shown that this famous formula, in spite of its deep philosophical significance, cannot be used meaningfully in the analysis and design of practical, present day communications systems. A more suitable channel capacity formula is derived for the practical case.

The analytical results thus obtained are used to show that broadband techniques have definite merit for both civil and military applications. Furthermore, such techniques will result in far more efficient spectrum utilization in many applications than any practical narrow-band, frequency-channelized approach. Thus broad-band techniques can, in many cases, increase the number of available "channels." With regard to military communications it is shown that the ability of a communication system to resist jamming varies in direct proportion to the transmission bandwidth for a given data rate. Thus narrow-band techniques lead progressively to more expensive communications systems and less expensive jammers. It is concluded that in the military field broad-band techniques are not only desirable but also often mandatory.

I. Introduction

MOST common usage of the radio frequency spectrum involves operation at specified frequencies as assigned by the appropriate regulatory agencies in the various countries. In contrast, the radio amateur service is assigned various bands of frequencies and properly licensed stations are permitted to operate at any frequency within these bands. This freedom of choice of frequency is necessitated by the obviously impossible administrative problem of assigning specific frequencies to specific stations and, furthermore, the available bandwidths fall short by several orders of magnitude of providing exclusive channels to each authorized station. Thus, as one might suspect, the situation in the amateur bands is a chaotic one in terms of mutual interference. There is very little tendency to "channelize" for several reasons. The crowded conditions normally leave no empty spaces in frequency so that a station starting operation has no choice but to transmit "in between" two strong stations or on top of a weaker station. Furthermore, at the higher HF frequencies, the ionospheric "skip" makes it impossible to choose a good operating frequency by listening, since the signal situation will be radically different between two points spaced many miles apart. Thus, the very nature of the amateur service would lead one to expect that any meaningful analysis of this problem must be based on a statistical approach.

A mathematical study of amateur radio communications can be of use in other important areas. Consider, for example, military communications where allocation of frequencies cannot possibly prevent interference due to the use of the same frequencies by the opposing forces. It is not hard to imagine that under such conditions each operator will shift frequency and take other appropriate action in order to get his message through. Thus, in a combat area we might well expect to find the very same chaos in the communications services that we observe in the amateur bands today. Certainly in such situations interference cannot be eliminated by allocation; interference will exist and we must simply learn to live with it. We are not speaking here of intentional jamming but rather of the casual interference which is inevitable when two opposing military forces (which today depend heavily on radio) attempt to operate independently and use the same electromagnetic spectrum. The problem of intentional jamming will be treated in detail in Section VI.

In the analysis of the radio amateur problem which follows, three modes of operation are compared. It is first assumed that all stations employ suppressed-carrier single-sideband (SSB). Then exclusive use of suppressed-carrier AM (DSB) is assumed. Finally, a frequency diversity system is examined in which each station transmits a large number of identical signals at randomly selected frequencies in the band. Intuitively we might suspect that SSB would be superior to DSB because of the two-to-one difference in signal bandwidths. The frequency diversity system is intuitively ridiculous because it apparently "wastes" bandwidth rather indiscriminantly. As we shall see, intuition is a poor guide in these matters. The feeling that we should always try to "conserve bandwidth" is no doubt caused by an environment in which it has been standard practice to share the RF spectrum on a frequency basis. Our emotions do not alter the fact that bandwidth is but one dimension of a multidimensional situation.

* Original manuscript received by the IRE, April 21, 1959; revised manuscript received, June 13, 1959.
† General Electric Co., Syracuse, N. Y.

II. Congested Band Analysis

SSB Case

We shall first consider the case of exclusive use of SSB. The spectral situation is shown in Fig. 1 as it might

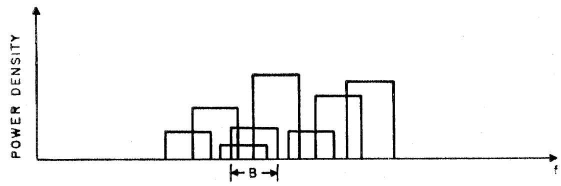

Fig. 1—Power density spectra—SSB case.

appear to a particular receiver. Each signal occupies a bandwidth B (equal to the baseband bandwidth for SSB), has a location in frequency independent of all other signal locations, and has an amplitude of power density independent of all other signal amplitudes. The signal amplitudes will have a probability distribution which will be specified at a later time. While the frequency locations of the various signals are distributed at random, it can be said that, on the average, there are a given number of signals per given unit of bandwidth. Thus, we may specify the density of loading of the band by a quantity k which represents the average number of signals per unit bandwidth. It happens that we shall need to know the probability of having a given number of signals ν falling in a bandwidth B. This, of course, is given for the conditions specified by the celebrated distribution of Poisson as

$$P(\nu, B) = \frac{(kB)^\nu}{\nu!} e^{-kB}, \tag{1}$$

where $P(\nu, B)$ is the probability of having ν signals in the bandwidth B if there are k signals per unit bandwidth on the average.

The choice of the distribution function for the signal power densities is somewhat arbitrary and, as far as the final results are concerned, apparently not particularly critical. It is physically reasonable and mathematically convenient to choose the chi-squared distribution[1]

$$p_\nu(x) = \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} \qquad (x \geq 0), \tag{2}$$

where $p_\nu(x)$ is the probability density function of the spectral amplitude which results from the summation of ν independent signals. For $\nu = 1$ the distribution has a mean of unity. This specifies that the average signal strength at the receiver is unity which results in no loss of generality for this application.

For convenience only, we shall assume that we are receiving a signal of average strength and want to find the probability that the Signal-to-Noise Ratio at the receiver output will equal or exceed a specified value. For SSB operation, the SNR at RF is the same as the SNR at the receiver output. We shall estimate the effective noise level at the receiver input by noting the interference level at the center of the pass band. We shall now determine the probability that the interference level will be less than or equal to J, which means that the SNR at the receiver output will be equal to or greater than $1/J$, since the desired signal is assumed to be of average strength of unity. Let $P_{\text{SSB}}(\text{SNR} \geq 1/J)$ be this probability. Then

$$P_{\text{SSB}}(\text{SNR} \geq 1/J) = P(0, B) + P(1, B)\int_0^J p_1(x)dx + P(2, B)\int_0^J p_2(x)dx + P(3, B)\int_0^J p_3(x)dx + \cdots, \tag{3}$$

which states that the event will occur if there are no signals in B, if there is one signal in B with amplitude less than J, if there are two signals in B the sum of whose amplitudes is less than J, etc. It should be clear that if an interfering signal is to contribute to the measurement of interference, its lowest frequency must fall somewhere within a frequency band extending from the center of the pass band to B cycles below. It is to this event that the terms $P(\nu, B)$ in (3) refer. Substituting (1) and (2) into (3) one obtains

$$P_{\text{SSB}}(\text{SNR} \geq 1/J) = e^{-kB}\left[1 + \sum_{\nu=1}^{\infty} \frac{(kB)^\nu}{\nu!}\int_0^J \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)}\,dx\right]. \tag{4}$$

Evaluation of (4) for a fixed J and variable k will give the probability of exceeding a certain receiver output SNR as a function of band loading. For example, for $J=1$ the expression gives the probability of exceeding a 0-db SNR when receiving a signal of average strength, or of exceeding a +3-db SNR when receiving a signal of twice (power) average strength, etc. Fortunately, the integral function in (4) is tabulated[2] and the series converges rather rapidly, so that the numerical work involved in evaluating (4) is not too difficult.

DSB Case

As might be suspected, the analysis of the case involving exclusive use of DSB is quite similar to the SSB analysis. There are two important differences to be noted. First, since all transmitted signals have twice the baseband bandwidth it is to be expected for a given band loading there will be more interfering signals involved than in the case of SSB. In the DSB analysis then, we will be concerned with the probability of having ν interfering signals in a bandwidth $2B$, using the same estimate of effective receiver input noise level as before.

[1] H. Cramer, "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J., ch. 18; 1946.

[2] C. D. Hodgman, "Mathematical Tables," Chemical Rubber Publishing Co., Cleveland, Ohio, p. 257; 1946.

Thus the Poisson distribution $P(\nu, 2B)$ must be used in the equation equivalent to (3) for the DSB analysis. This represents a loss caused by increased transmission bandwidth; there is a compensating gain as will be seen. The second difference between the SSB and DSB analysis involves the relationship between the predetector and postdetector SNR's. In SSB these two ratios are the same. In DSB the postdetector SNR is 3 db better than the predetector value. This difference arises because of the coherent addition of upper and lower sideband components of the signal and incoherent addition of the corresponding interference components in the synchronous detector. Thus, for identical output SNR's the interference power density will be two times as great relative to desired signal density in DSB as compared to SSB. Consequently, in the equation equivalent to (3) the upper limit on all integrals must be changed from J to $2J$ in order that J have the same meaning in both cases.

When the two changes discussed above are made, the probability of exceeding an output SNR of $1/J$ for a desired signal of mean strength (unity) becomes

$$P_{\text{DSB}}(\text{SNR} \geq 1/J) = e^{-2kB}\left[1 + \sum_{\nu=1}^{\infty} \frac{(2kB)^\nu}{\nu!} \int_0^{2J} \frac{x^{\nu/2-1}e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)}\, dx\right]. \quad (5)$$

A comparison of (4) and (5) shows that the increased bandwidth of DSB has in some ways been detrimental ($2kB$ in place of kB in the Poisson distribution), and in other ways beneficial ($2J$ in place of J in the integral expression). As later calculations show, the increased bandwidth of DSB does not affect the relative congested band performance as compared to SSB in any significant manner. We might begin to suspect that the efficient use of broader bandwidths in a congested operating band is not necessarily a bad idea. The broader bandwidth signals will increase the tendency of frequency overlap and tend, in a sense, to cause more interference. This is obvious. *What is not so obvious is the fact that the increased bandwidth gives to the receiving system an increased ability to discriminate between the desired signal and the interference.* In order to investigate further the effects of increasing transmission bandwidth, a rather simple form of broad-band technique will now be analyzed.

Frequency Diversity Case

For this example we shall use the SSB mode of transmission (although the DSB mode would yield identical results), in a somewhat unusual manner. Each station will transmit not one but M (where M is a large number) identical signals at randomly chosen frequencies in the congested band. The receiver must know these frequency locations so that all M signals may be received, detected, and added coherently to produce the receiver output signal. With each station transmitting M identical signals, the interference spectrum amplitude will, with nearly unit probability, be very nearly equal to a constant value at all frequencies for sufficiently large M. This value may be determined quite easily by inspection.

Consider first the normal SSB situation without diversity. The received signals are distributed in amplitude of power density about a mean of unity. Thus, the average received power is B watts per station. Since there are k stations per cycle on the average, the mean interference power density will be kB watts per cycle. Going from one transmission to M transmissions per station (assuming the power of each station is now split evenly between the M signals) does not alter the value of the average interference power density. In the diversity case this *average* value will be very nearly the *actual* value of interference density level which will exist at all frequencies and at all times. The diversity receiver output SNR may now be easily calculated.

Each of the M signals will have a power B/M (for the average signal strength case) and the noise power accepted in receiving each of the M signals will be kB^2. The RF SNR at each of the M frequencies will be $1/MkB$ and coherent addition of M such signals will yield an output SNR of $1/kB$. So then

$$(\text{SNR})_{\text{Div}} = \frac{1}{kB} \quad (6)$$

on a power basis for a desired signal of mean strength. Note that in (6) we are able to specify the precise SNR, while in the SSB and DSB cases of (4) and (5) we can only predict the probability or the percentage time the SNR will exceed a given value.

III. Results and Discussion—Congested Band

The results represented by (4)–(6) may be interpreted in many different ways. For the purposes of this discussion let us assume that voice communications is involved and that message reception will be considered successful if the receiver output SNR equals or exceeds unity or 0 db. Keep in mind that this is not a commerical service but rather a service where the operator is willing to exert some effort in order to understand what is being said. Thus, the 0-db choice is probably reasonable with regard to sentence intelligibility where the interference is of an incoherent nature. The three equations will then be used to calculate the circuit reliability for signals at 0, +3, +6, and +9 db relative to mean signal strength as a function of kB, the band loading expressed in average number of stations per audio bandwidth. The resulting graphs are shown in Figs. 2 through 5. Turning first to Fig. 2, which assumes a received signal of mean strength, we note that the circuit reliability drops rather rapidly with band loading for both SSB and DSB. SSB shows some advantage, but of a small amount, at loadings which result in a reasonable reliability percentage. An estimate of the increased number of users for the same performance which results from SSB use may be obtained by drawing a line horizonally from any given

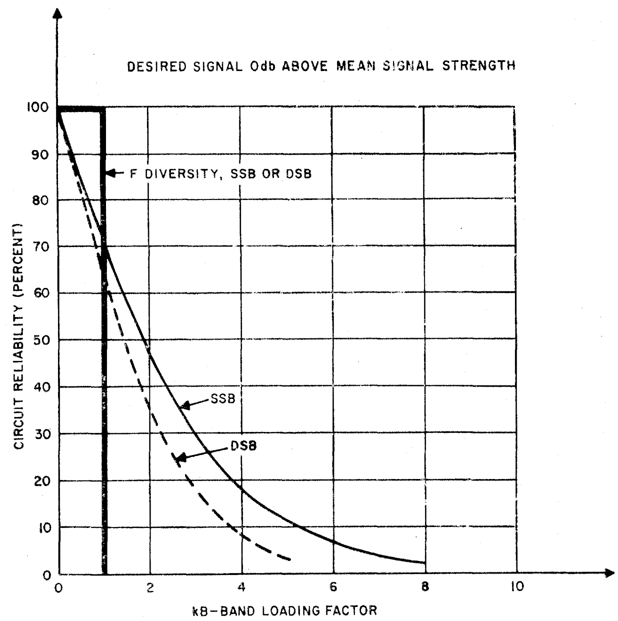

Fig. 2—Per cent circuit reliability vs band loading.

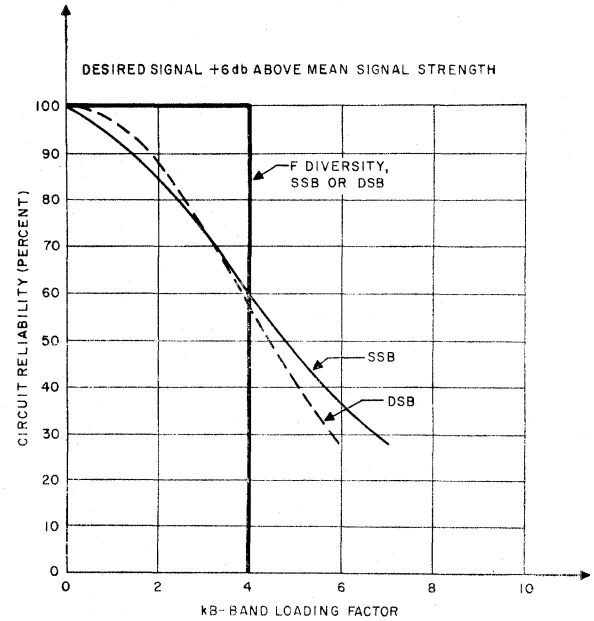

Fig. 4—Per cent circuit reliability vs band loading.

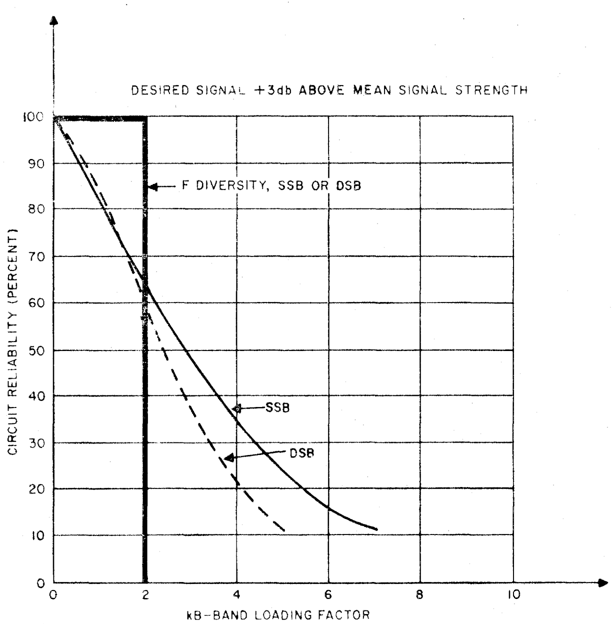

Fig. 3—Per cent circuit reliability vs band loading.

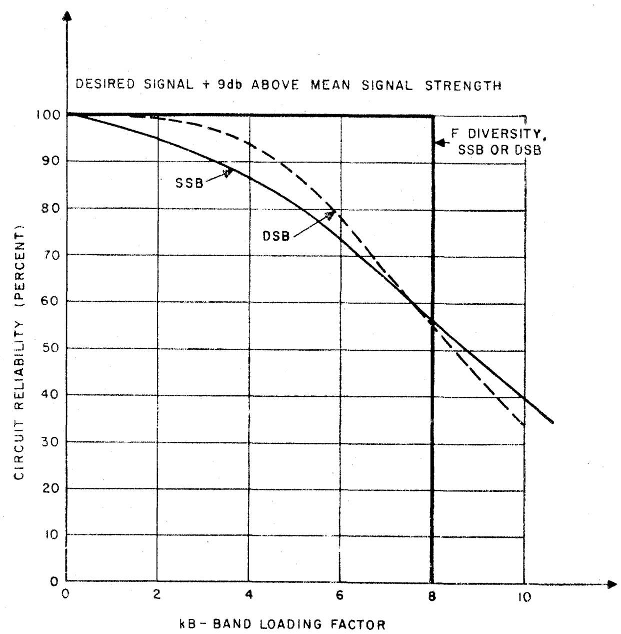

Fig. 5—Per cent circuit reliability vs band loading.

ordinate value and noting the kB values at which this line interesects the SSB and DSB curves. The two-to-one increase in loading which we might at first expect from SSB certainly does not materialize, except at values of circuit reliability which are so low as to be meaningless. Thus, the randomness of band occupancy has a significant effect on performance, and any intuitive conclusions based on orderly channel assignments are subject to considerable error. Note in particular that the circuit reliability for SSB at $kB=1$ is 70 per cent. At this loading there are enough channels to satisfy all needs, and 100 per cent reliability could be had if some organization could be obtained. About the only conclusions to be drawn from Fig. 2 are that SSB and DSB give nearly the same performance and that it is usually hopeless to try to communicate with a station whose signal strength is only average at times when the band is crowded. This last conclusion will come as no surprise to the experienced operator.

As the strength of the desired signal increases above the mean value the situation improves rather rapidly, as shown by Figs. 3–5. The SSB and DSB curves now "cross over" and both curves tend to stay at higher reliability values as kB is increased, which is to be expected. Note in Fig. 5 that DSB shows a slight advantage over SSB for the lower loading values and the cross-over occurs when the reliability factor is 63 per cent. In total these results show the futility of claiming any advantage for either SSB or DSB in this service. If one is insistent upon claiming an advantage, the specific conditions under which the comparisons are made must be given.

In our attempt to determine the sensitivity of the calculated results with respect to the choice of the amplitude distribution function, an exponential distribution was tried in place of the chi-squared. The exponential distribution gave more weight to signals above the mean than did the chi-squared. However, the final results were very nearly the same. A further calculation involving a delta-function distribution (all interfering signals of the same strength) showed no significant differences. Thus, one is led to believe that the results obtained are not particularly sensitive to the choice of any reasonable distribution function for the signal strengths.

The performance of the frequency diversity system shows up in a rather unusual manner in the graphs. This is due in part to the way in which we chose to interpret the results, and in part to the fact that in this case the interference is not random but constant. In the narrow-band cases the interference level changes considerably in short periods of time because of the random appearances and disappearances of signals close to the operating frequency. In the broad-band case, the interference observed is the net result of *nearly all* the stations on the band so that the actions of *any one* station have a negligible effect on the interference level at the output of an appropriate broad-band receiving system. Thus, for a given loading, the interference level stays fixed and only the signal strengths of the various stations to which the receiver is "tuned" will be found to vary. Some signals will be sufficiently above the noise to be understood all of the time, while others will be below the noise and will not be heard at all. We have made a rather interesting trade in going from narrow- to broad-band operation. In narrow-band operation, we can copy a strong signal most of the time and a weak signal just part of the time. In broad-band operation, we can copy a strong signal all of the time but a weak signal cannot be copied at all. The reason for the shape of the frequency diversity curves should now be clear, and the nature of the "trade-off" may be evaluated by an examination of Figs. 2–5.

Amateur band operation with broad-band systems will prove to be somewhat different in certain respects. There will be fewer stations with which contact may be established (since the weaker signals which were formerly heard intermittently will now not be heard at all), but once contact is established the conversation can be expected to continue without interruption for a considerable period of time. Since the amateur is not normally concerned with communicating with a specific person, the exchange of some freedom of choice of possible contacts for reliability of communications will probably be welcomed.

In the case of military communications, the problem is more difficult, since specific messages must be transmitted to specific stations. If the signal strengths are weak, the narrow-band approach certainly offers no solution since, as we have seen, the circuit reliability will be poor. The message will have to be repeated over and over again before it is received with any reasonable degree of completeness and accuracy. Thus, under such adverse conditions *we have been forced to lower the data rate* because the necessity for repetition has increased the time required for the transmission of a given message. Broad-band operation under the same adverse conditions will suffer the same fate, but to a lesser degree. The data rate will have to be lowered (this can be done without decreasing the bandwidth) but since the interference level will be fixed at some average value we can lower the rate by just the amount necessary to keep the error rate below the acceptable maximum. With narrow-band operation, practical considerations will no doubt force us to reduce the data rate to a value determined by the *maximum* interference level. Thus, for congested-band operation, broad-band systems appear to offer a more orderly approach to the problem and a potentially higher average traffic volume than narrow-band systems.

Nothing that has been said so far should be construed as meaning that broad-band systems will always give us the traffic volume we would like to have, or feel we must have to support operations. As the congestion becomes worse it will be impossible to avoid reducing the data rate per circuit. The important point here is that the broad-band philosophy *accepts interference as a fact of life* and an attempt is made to do the best that is possible under the circumstances. The narrow-band philosophy essentially denies the existence of interference since there is an implied assumption that the narrow-band signals can be placed in non-overlapping frequency bands and thereby prevent interference. It is perhaps redundant to state that the realities of most practical military situations almost completely destroy the validity of such reasoning.

At this point we shall leave the problem of the radio amateur and turn our attention to other communications areas. We have seen that the operating environment of the amateur is not unique to his service but that in other services, especially the military, conditions in actual practice will quite often degenerate to the congested situation of the amateur service. Under such conditions we have shown the necessity for a statistical approach to the problem. It has been further demonstrated that the efficient use of additional transmission bandwidth does not constitute a "waste" in the basic sense of the word. The policy of "conserving bandwidth" is not based on sound physical principles but is based rather on a very common but still myopic view of communications. Such a policy will, in many situations, conserve only the opportunity to communicate as efficiently as might otherwise be possible. Even worse, this point of view quite often leads to the design of systems which have little or no true military capability because of extreme sensitivity to intentional interference. These and other matters will be discussed in Sections VI and VII in more detail. First, it is necessary to derive some rather simple results from information theory.

IV. Information Theory

Consider the problem of data transmission by electrical means. We transmit pulses over a noisy circuit and the pulses, together with the noise, are received and interpreted. Errors in interpretation of the message occur because of this noise. If the error rate is too high and the transmitter power is fixed, we have traditionally lowered the data rate in order to reduce the errors. This has always worked and the reason given was very simple. A lower data rate means that the pulse lengths can be increased, which in turn allows narrower bandwidths to be used, thereby reducing the amount of noise accepted by the receiver. Thus, it became axiomatic that lower error rates could be obtained only by corresponding decreases in bandwidth and data rate. To almost everyone in the communications art the validity of this axiom was unquestioned since there was a great deal of experience in support and none in contradiction. It remained for Shannon to show that systems could be constructed, in theory at least, which would behave quite differently from what our previous experience would lead us to expect. First of all, he showed that the data rate could be held at a constant value (provided this value were below a certain maximum) and at the same time the error rate could be reduced to arbitrarily small values. As for the general belief that one should always use the minimum possible bandwidth in order to reduce the noise accepted by the receiver, Shannon showed that in the ideal case, with a white-noise background, the system bandwidth should be increased to the point where the accepted noise power is at least equal to the signal power.[3] This new theory presented a radically different picture of the limiting behavior of communications systems.

A very superficial study of Shannon-type systems will now be made in the belief that many readers, who are not specialists in information theory, might find a practical discussion of this topic interesting and perhaps useful. Fig. 6 shows a form of communications system suggested by Shannon's work. The channel has a bandwidth W and average (white) noise power N. The transmitter is limited to an average power P. Consider a white-noise generator having a bandwidth W. We record M different samples of the generator output, each sample having a duration of T seconds. These waveforms are now designated $f_1(t)$, $f_2(t)$, $f_3(t)$, $\cdots$, $f_k(t)$, $\cdots$, $f_M(t)$ and are made available as transmitted symbols, as indicated in the figure. Copies of each of the M waveforms are made and placed at the receiver in the corresponding operator units $O_1, O_2, \cdots, O_k, \cdots, O_M$. In operation, one of the M waveforms (say the kth one) is selected for transmission. Waveform $f_k(t)$ plus channel noise is received by each of the operator units. The operator units subtract the waveform stored within each unit from the received signal, square this difference, integrate the square for T seconds (which is the duration of the symbols), and indicate this mean-square value as shown. If T is sufficiently large, each meter (except for the kth one) will with almost unit probability read very nearly a value corresponding to $2P+N$, which is the average power of the difference voltage in each case. The kth meter will give a reading corresponding to very nearly N (again with almost unit probability), since the $f_k(t)$ portion of the received signal is completely removed in the subtraction process and only the channel noise remains. Thus, by noting which meter has the lowest reading we can identify which of the M symbols was transmitted. Of course, because of the channel noise we will make an occasional error and identify the wrong symbol.

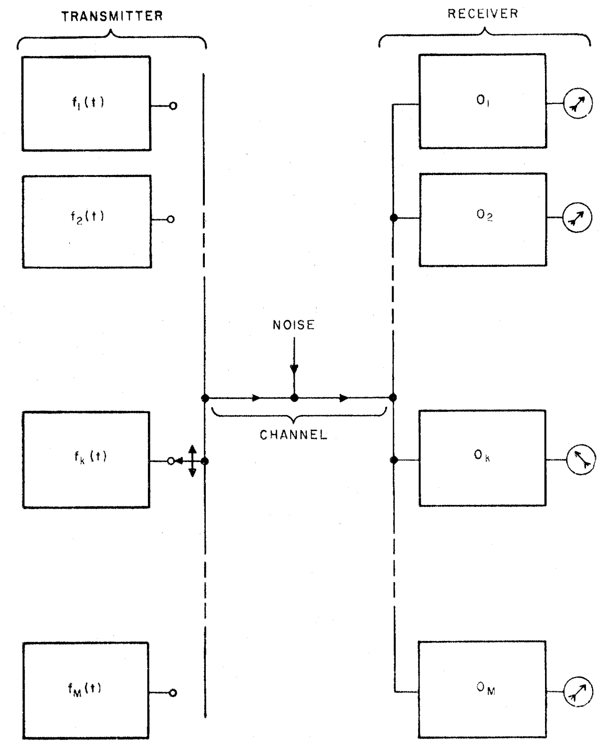

Fig. 6—An ideal communications system.

Before investigating the problem of errors we should examine the relationships between data rate R, symbol duration T, and number of symbols M. Assume that in each T seconds of time the system receives S binary digits (0, 1) to transmit. R will then be S/T bits per second. Since our symbol length is T, we must be prepared to indicate a choice of one out of 2^S possibilities with each symbol transmitted, since this is the number of different sequences of S binary digits. Then clearly

[3] C. E. Shannon, "Communication in the presence of noise," Proc. IRE, vol. 37, pp. 10–21; January, 1949.

$M = 2^S$ and

$$R = \frac{S}{T} = \frac{\log_2 M}{T} \text{ bits per second.} \tag{7}$$

Note that if the symbol length T is increased, the number of symbols M used must increase *exponentially* with T in order to keep a constant data rate. Thus, if T is *doubled*, M will have to be *squared* for the same data rate. In general terms

$$M = A^T \tag{8}$$

and

$$R = \log_2 A. \tag{9}$$

Returning again to Fig. 6, assume the kth symbol has been transmitted. Thus, we look to see if the kth meter gives the lowest reading. If this is so, there is no error. If any one of the other meters gives a lower reading, an error in selection will occur. The probability that any meter will read less than the kth one can be made progressively smaller by increasing T, which increases the integration time in the operator units. However, this is only part of the story. As T is increased to lower the probability of any one meter indicating lower than the kth, the number of such comparisons needed rises according to (8) in order that the data rate remain fixed. Thus, we have two conflicting trends as T is increased. The probability of error per comparison drops, but the number of comparisons necessary to arrive at a selection rises with increasing T. Shannon shows that we can always reduce the over-all probability of error in selection to as small a value as we may choose by letting T become large, *provided* that M does not increase with T faster than

$$M = \left(1 + \frac{P}{N}\right)^{TW}. \tag{10}$$

This maximum permissible rate of increase of M with T determines the maximum data rate which can be supported with an arbitrarily small error rate. This maximum rate is known as the channel capacity C and is obtained by substituting (10) into (7) to obtain

$$C = W \log_2 (1 + P/N). \tag{11}$$

Of course, we do not have to send data at the rate given by (11). We may send slower, and enjoy arbitrarily low, error rates. We may even send faster than C, but then we must accept a certain irreducible error rate.

As remarkable as (11) may be, the engineer concerned with practical system design needs more information than has been given thus far. We now know that multisymbol systems of the type shown in Fig. 6 are capable, practical considerations aside, of making the most efficient possible use of the communications channel. There are two engineering constraints which must be considered carefully. First, there is an inherent delay of $2T$ seconds involved in data transmission because a T-second length sample of input binary data must be available before choice of transmitted symbol may be made, and another T seconds is required for processing at the receiver before identification may be made. What will be the order of magnitude of this transmission delay? Secondly, how many different symbols M will be required in a given situation? This last consideration is of special importance because it determines, rather directly, system complexity. We might suspect that any attempt to operate at or very near the rate C would require intolerably large T and M since this rate represents a limiting condition. Similarly, large T and M would be expected to result at operating rates lower than C if the error rate is specified at a very small value. What we really need to know is the behavior of T and M for a practical error rate, say 10^{-5}, as the data rate is varied from zero to 100 per cent of capacity. Rice, in an excellent paper,[4] gives us a good indication of the orders of magnitude involved. Rice assumed an SNR of 10 and an error rate of 10^{-5}. He then determined the number of bits per symbol S which would be necessary for various values of the ratio of actual data rate to channel capacity. The results are shown plotted in Fig. 7. Notice

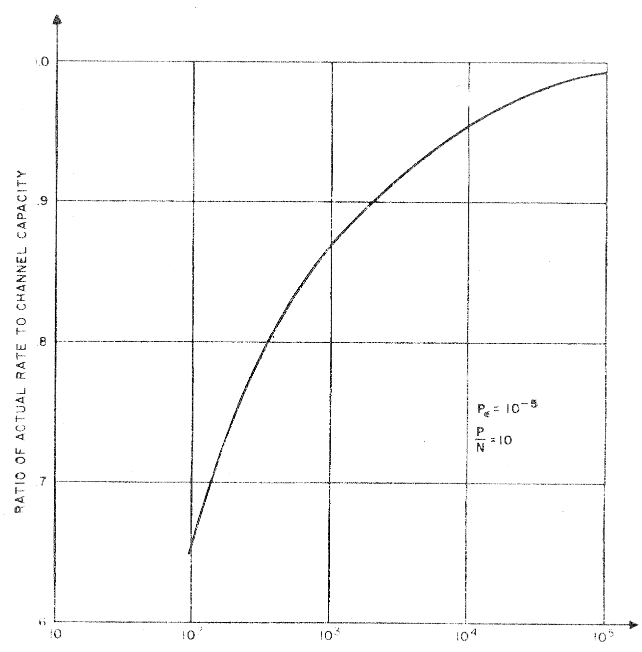

Fig. 7—Curve from Rice showing approach to capacity.

that the numbers S of bits per symbol are quite large, and keep in mind that the number of symbols M is 2^S. We need no numerical examples to conclude that the number of symbols needed will be fantastically large and that it is completely impractical to attempt to build systems which operate at rates close to the Shannon capacity under the conditions assumed above. (An

[4] S. O. Rice, "Communication in the presence of noise," *Bell Sys. Tech. J.*, vol. 29, pp. 60–93; January, 1950.

interesting piece of work by Stutt[5] shows that the situation is not quite so unreasonable if the SNR is low and the symbol waveforms are chosen systematically rather than at random.)

In brief retrospect, we (as communications engineers) have been shown by Shannon that there is an upper limit to what we can do no matter how hard we may try or how ingenious we may be. That it may be extremely difficult to achieve or even approach this upper limit in practice can hardly be blamed on Shannon. He has located the top of our mountain; the problem of reaching the peak is ours, not his.

V. A Practical System of High Efficiency

It is quite clear that any analysis of a communications problem which uses the capacity formula without careful qualification may give results of doubtful practical value. If a system of high efficiency and of reasonable complexity could be found, perhaps problem analysis could be carried out with results which would be significant in practice. Consider once more the system of Fig. 6, but now let there be only two symbols used, $f_1(t)$ and $f_2(t)$. Shannon's idea of using noise-like symbols is quite intriguing. This will be retained except that $f_2(t)$ will be the negative of $f_1(t)$ instead of being chosen at random as before. Thus, $f_1(t)$ is now transmitted for mark (or binary 1) and $-f_1(t)$ for space (or binary 0). For obvious reasons we shall refer to this two-symbol system as the binary system.

In the analysis of this binary system it is convenient to recall one form of the sampling theorem which states that a time function of T-seconds duration and of W-cycles bandwidth is completely specified by $2TW$ equally-spaced sample values of the funtion. Thus, we will represent the function $f_1(t)$ by the sequence of numbers $\{x_1, x_2, \cdots, x_{2TW}\}$, which are the values of the function at the sampling times. The function $f_1(t)$ will be noise-like except that we shall adjust the function so that we obtain the exact relationship

$$\frac{1}{2TW}\sum_{1}^{2TW} x_j^2 = P, \tag{12}$$

where P is the average transmitter power. In a like manner the channel noise, which has an average power N, will be represented by the sequence of numbers $\{n_1, n_2, \cdots, n_{2TW}\}$, where the n_j are independent normal variables with zero mean and variance N. If one performs the operations described for Fig. 6 one obtains the following for the probability of error P_ϵ:

$$P_\epsilon = \text{Prob.}\left[\frac{1}{2TW}\sum_{1}^{2TW} x_j n_j < -P\right]. \tag{13}$$

The summation term may be shown to be Gaussian with zero mean and variance $PN/2TW$. If operating conditions yield a low error probability, then

$$P_\epsilon = \frac{e^{-\gamma}}{2\sqrt{\pi}(\gamma)^{1/2}}, \tag{14}$$

where

$$\gamma = \frac{P}{N}TW. \tag{15}$$

A plot of $\log_{10} P_\epsilon$ as a function of γ is shown in Fig. 8.

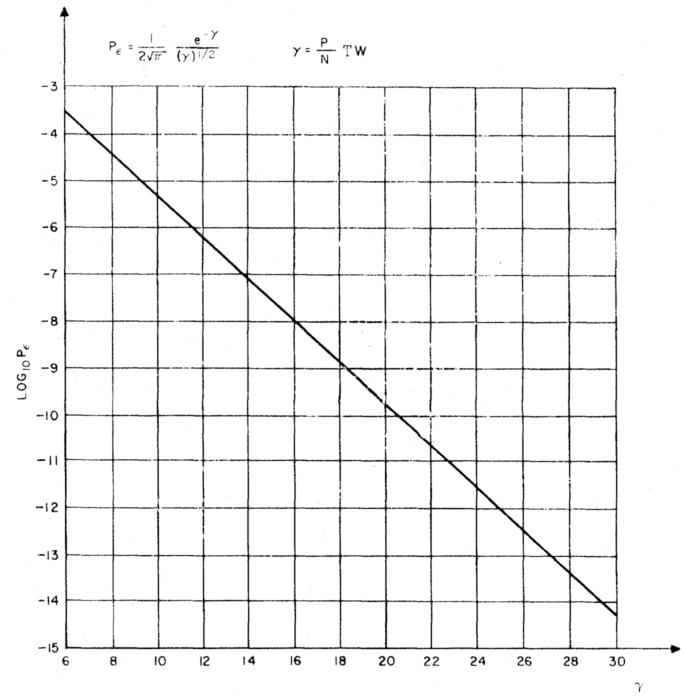

Fig. 8—Plot of $\log_{10} P_\epsilon$ vs γ for the binary system.

Note that once the system error probability is fixed, the relationship between SNR, bandwidth, and data rate ($1/T$ bits per second) is immediately determined.

We might inquire now as to how good our binary system is. It is certainly as good as any two-symbol system can be. Better results can be obtained only by increasing the number of symbols. The gain in doing this, however, does not generally appear to be worth the effort. For example, Stutt[5] shows that for a P/N of 1/10 and error probabilities in the neighborhood of 10^{-4} to 10^{-6}, the most efficient symbol choice requires the use of about 100 symbols in order to increase the data rate over binary by a factor of five. Note, however, that at a fixed error rate the data rate of the binary system may be made 5 times as large by inceasing transmitter power by 7 db. Thus, we must evaluate the relative costs of a 7-db transmitter power increase vs the increase in symbols from two (actually one in terms of equipment complexity) to 100. We must conclude, therefore, that our binary system performance represents about the best that can be done in practice. Better results may be obtained by using more symbols but the rate of improve-

[5] C. A. Stutt, "Regular Polyhedron Codes," Research Laboratory, General Electric Co., Schenectady, N. Y., Tech. Rept. No. 59-RL-2202; February, 1959.

ment will generally prove to be low.[6]

This places us in a good position to derive expressions for the channel capacity in the practical case from (14) and (15). Before doing so we should understand that for high SNR's these equations may yield rates in bits per second far in excess of the bandwidth in cycles per second. There is a mathematical limitation which prevents this. This limitation will not be discussed except to mention the fact that in theory the binary system is limited to a maximum rate $2W$ regardless of SNR.[7] In practice it is usually quite difficult to achieve even W as a rate; we shall choose this as our limiting value.

Since the rate R is $1/T$, the channel capacity C_p in the practical case may be obtained from (15) as

$$C_p = \frac{W}{\gamma} \frac{P}{N}, \tag{16}$$

$$C_p \leq W, \tag{16a}$$

where γ is fixed by the desired error probability according to (14). For an error probability of 10^{-5}, γ is approximately 9.2.

Admittedly, the result (16) is not as elegant as (11). Keep in mind, however, that the concise nature of the capacity formula (11) is made possible by a limiting process in mathematics which cannot be duplicated in practice. A valid objection could also be raised to the application of the term "capacity" to the rate indicated by (15) and expressed in (16). From purely theoretical considerations, such an objection is certainly justified. From a practical point of view, (16) does, in a sense, qualify as a capacity since the performance indicated may only be approached by the most efficient use of modulation and processing techniques. It is quite doubtful that there exist at present any operating systems which perform as well as (16) indicates is obtainable. The main point to remember is that for many years to come (16) will represent a sensible, realizable (but not easily realizable) design goal for the communications engineer; the capacity formula (11) can never serve this purpose. As processing and storage techniques improve, it is to be expected that at some future time multisymbol systems may be built whose performance will exceed that indicated by (16). This does not in any way lessen the utility of (16) as a frame of reference.

VI. Jamming

From the work of the previous section, we now derive some rather simple results which are well-known to information theory specialists the world over, but whose significance is apparently not appreciated by many engineers, at least in this country.

Consider first the performance of the binary system in a white-noise background having a density of n_0 watts per cycle. The effective noise power N will then be

$$N = n_0 W, \tag{17}$$

and (15) will read in this case

$$\gamma = \frac{PT}{n_0} = \frac{E}{n_0}, \tag{18}$$

where E is the energy per transmitted symbol. We now have derived the well-known result that for binary systems of this type operating against flat channel noise, the error probability is independent of bandwidth and is a function only of energy per symbol and noise power density. Thus, for fixed average signal power and fixed data rate, the error probability does not change as the system bandwidth is increased. It is clear that as the bandwidth increases, the noise power accepted by the receiver increases, and for large bandwidths the received noise power becomes quite large compared to received signal power. Thus, systems of this type can operate with SNR's far below unity, or put another way, these systems can be made to operate satisfactorily in the presence of very large amounts of noise power. One might begin to suspect that a broad-band system would be fairly immune to intentional jamming, since in normal operation it is contending (satisfactorily) with such large amounts of natural noise that the additional noise contributed by the jammer would be insignificant by comparison. That this is precisely the case will be made more definite in what follows.

Consider a binary communications system designed to operate in a white-noise background of power density n_0 watts per cycle. Let practical considerations demand that the error probability be kept at or below a critical value P_{e0} corresponding to a γ value of γ_0. Then the channel capacity will be from (16):

$$C_p = \frac{W}{\gamma_0} \frac{P}{n_0 W} \tag{19}$$

as far as natural noise is concerned. For the sake of argument, we shall choose to operate at a data rate R corresponding to one-half capacity. Then,

$$R = \frac{1}{T} = \frac{C_p}{2} = \frac{P}{2\gamma_0 n_0}. \tag{20}$$

Consider now the appearance of a jamming signal of average power J in the channel and let us investigate the effect of J on γ, since this factor must be kept above the assumed critical value of γ_0. The noise term N in (15) will now be

$$N = n_0 W + J, \tag{21}$$

and when (20) and (21) are substituted into (15) we obtain

$$\gamma = \gamma_0 \frac{2n_0}{n_0 + J/W}. \tag{22}$$

[6] This, like all generalizations, will have exceptions. One can conceive of situations in which the multisymbol system would have sensible application. In such cases the work of Stutt, *ibid.*, should prove quite useful.

[7] See discussion of sampling theorem which precedes (12).

This last equation tells a very interesting story. First of all, note the appearance of the J/W term in the denominator. This indicates that the effectiveness of a given average jamming power varies *inversely* as system bandwidth. The broader the bandwidth the less effective will be the resultant jamming. In particular, in order to disrupt the circuit ($\gamma=\gamma_0$) one needs an amount of average jamming power J_0 equal to

$$J_0 = n_0 W \tag{23}$$

under the conditions specified. Thus, the relationships between bandwidth and jamming power become quite clear and may be summarized as follows: *If the most efficient system design is assumed for a fixed data rate in each case, the necessary power required to jam the ciruit varies in direct proportion to system bandwidth. The broader the bandwidth the more difficult it will be to jam the circuit. Conversely, the narrower the bandwidth the easier it becomes to jam the circuit.*

It should be quite clear that if intentional jamming is a consideration, one must of necessity choose a broad-band technique. The narrow-band approach can only lead to eventual disaster.

VII. The Question of Channels

The well-known, but not necessarily sufficiently appreciated, relationship between jamming immunity and system bandwidth discussed above leads to a natural concern over loss of channels if broad-band techniques are employed, as obviously they must be in many applications. It is the purpose of this section to discuss the general problem of "channels" somewhat more thoroughly than before, through use of the practical channel capacity formula (16).

Consider the following problem. Communications service must be provided which requires that a total of K stations be permited to transmit messages at *any* time. Let α be the average fraction of time each station is active. The average signal strength (power) at a particular receiver will be denoted by $\overline{P}$ and it is assumed that Ω cycles of total bandwidth are allocated to this service. Background or natural noise will be ignored. Thus:

Ω = Total bandwidth allocated to service.

K = Number of stations, each of which must be permitted to transmit at any time.

α = Average fraction of time each station is actually transmitting.

$\overline{P}$ = Mean signal power at a receiving site (one station).

C_{pN}, C_{pB} = Practical channel capacity per circuit in narrow- and broad-band operation, respectively.

We now wish to inquire as to the relative merits of narrow- and broad-band techniques for this service.

First let us assume an environment in which *all* stations are under the complete control of a central authority. Under this special condition, frequency division will result in circuit bandwidths of Ω/K and, since there will be no interference and background noise is ignored, the capacity per circuit will be, using (16a),

$$C_{pN} = \frac{\Omega}{K} \cdot \tag{24}$$

By comparison, the broad-band approach would yield a circuit bandwidth of Ω and a noise power N of $\alpha K\overline{P}$, which, if an average strength signal were being received, would result in a capacity per circuit of

$$C_{pB} = \frac{\Omega}{\gamma\alpha K}, \tag{25}$$

using (16). Comparing (25) and (24) we see that if such a well-disciplined environment can be found, the narrow-band system would be superior *provided* that the duty cycle factor α is kept high. For example, if $\alpha=1$ (each station transmitting continuously) the narrow-band system appears to offer about a ten-to-one data rate advantage (for $\gamma=10$). If, due to operational considerations, the average duty cycle is low (say, 10 per cent or even 1 per cent or less as may quite often be the case), then the broad-band system, even under such ideal conditions, becomes superior.

The reasons for this are quite clear. If the duty cycle is low, the narrow-band system wastes spectrum since most of the allocated channels in Ω will be idle at any time. This cannot be avoided since each of the K stations must have access to communications at any time. The broad-band system takes immediate advantage of a low-duty cycle since this keeps the "noise" level at low values and increases the per-circuit capacity. The narrow-band approach guarantees complete elimination of interference between stations (orthogonality, as the specialist would say), while in the broad-band case each station appears as noise to the others. Thus, at high duty cycles the narrow-band system is superior because it avoids this "noise" problem completely. We must conclude then that the narrow-band system has sensible application under very special conditions (such as in radio broadcasting), but that even where complete control of all transmitters is possible, the broad-band system can easily prove to be the more efficient user of spectrum.

We shall now consider the same communications service problem as before, except that we shall abandon any hope of a disciplined use of the bandwidth Ω. In most military applications, a congested band assumption is much more realistic for several reasons. Certainly two opposing military forces will have planned their spectrum usage independently. Under such conditions interference will be the expected rather than the unusual event. If narrow-band systems have been chosen, it is quite likely that each operator will shift frequency when severe interference is encountered in an attempt to maintain service. This is only the natural and sensible thing to do. Furthermore, under conditions where signals propagate over distances of many thousands of miles, interference will no doubt be quite common even between stations that are a part of the same military force. It seems unrealistic to expect that interference can be prevented by administrative means when the total number of users is large and when the geographic distances between groups of users is great. It must be presumed then that, in spite of careful allocation attempts, the narrow-band approach will not prevent interference and that congested operating conditions will certainly prevail.

Consider the problem that an operator faces when trying to clear messages in a congested band using narrow-band systems. As we have shown, the SNR in such a case is a statistical quantity varying from very good at one time to hopelessly poor minutes or even seconds later. If the data rate is set too high (based on those times when the SNR is good), much of the traffic will be lost and repetition will be necessary. In order to know what messages or parts of messages were lost, a return link is required, but this will also suffer from interference. Such operation is quite inefficient and it would soon be discovered that the data rate would have to be determined by the *least favorable* SNR anticipated during the operating period. Thus, the per-circuit channel capacity in this case may be approximated roughly by

$$C_{pN} = \left(\frac{P}{N}\right)_{\min} \frac{\Omega}{\gamma K}, \tag{26}$$

using (16).

The assumption of congestion does not alter the performance of the broad-band system so that (25) still holds. A rough estimate of the relative performance of these two approaches to the problem of congested band operation may now be obtained by taking the ratio of (25) to (26):

$$\frac{C_{pB}}{C_{pN}} = \frac{1}{\alpha(P/N)_{\min}}. \tag{27}$$

As rough as this approximation may be it still seems rather certain that in a congested band the broad-band system will normally far outperform the narrow-band system.

Eq. (25) may then be taken as the average capacity per circuit of a congested band. In slightly modified form we have:

$$C_{pB} = \frac{1}{\gamma k} \text{ bits per second,} \tag{28}$$

where k is the average number of actual users per cycle of bandwidth and γ is determined by the required error probability according to (14).

It might be mentioned at this point that with broad-band operation the DSB and SSB methods of modulation give identical results for the same transmission bandwidth. As a practical matter, the DSB system offers a two-to-one increase in transmission bandwidth in the modulation process over and above the bandwidth increase obtained by coding processes at baseband. This may sound strange to engineers accustomed to design work aimed at conserving bandwidth. It is still true that there are practical difficulties involved in designing equipment which uses more bandwidth *efficiently* and that the bandwidth doubling which may be obtained in the modulation process with DSB will prove quite helpful in general.

VIII. Conclusions

Since the invention, many years ago, of the frequency-selective filter, it has been common practice to share the inherent capacity of the RF spectrum among users on the basis of frequency allocations. As the number of users increased, methods were found for reducing transmission bandwidths so that new services could be accommodated in the existing spectrum. Extrapolating the past into the future has led to the natural attempt to continue this evolutionary process of seeking methods for the further narrowing of transmission bandwidths, thus providing service for the increasing user population.

This philosophy of spectrum usage is based on a particular course of development which the radio art happened to take, rather than on any fundamental physical principles. The inherent communication capacity of the spectrum can be shared in ways other than by frequency allocation and for many applications the frequency division approach represents a very poor choice indeed. In the field of military communications in particular, the tendency to follow the trends of the past quite often leads to systems having negligible military capability although good intentions may be to the contrary.

This is not to say that broad-band systems have been completely ignored in the past. It could safely be said, however, that the magnitude of the effort thus far expended on the broad-band approach is far out of proportion to the importance of this technique.

Acknowledgment

The author wishes to express appreciation to his many colleagues at the General Electric Company who assisted in the work upon which this paper is based. Special thanks are due to J. C. Kovarik who contributed much original material for the congested-band analysis and to S. Applebaum and Dr. H. D. Friedman for their constructive discussions of information theory problems.