Quantum Effects in Communications Systems*

J. P. GORDON†

Summary—**The information capacity of various communications systems is considered. Quantum effects are taken fully into account. The entropy of an electromagnetic wave having the quantum statistical properties of white noise in a single transmission mode is found, and from it the information efficiency of various possible systems may be derived. The receiving systems considered include amplifiers, heterodyne and homodyne converters and quantum counters. In the limit of high signal or noise power (compared to $h\nu B$, where h is Planck's constant and ν and B are, respectively, the center frequency and bandwidth of the channel) the information efficiency of an amplifier can approach unity. In the limit of low powers the amplifier becomes inefficient, while the efficiency of the quantum counter can approach unity. The amount of information that can be incorporated in a wave drops off rather rapidly when the power drops below $h\nu B$.**

I. Introduction

With the advent of the possibility of broad-band communications at frequencies in the infrared and optical range, it has become important to investigate the effects of the quantization of radiation on the capacity of electromagnetic waves to transmit information. Unlike the situation prevailing in the microwave range, where thermal noise generally provides an ultimate limit to our ability to transmit information, in the infrared and optical range this limit is provided by what may be called quantum noise.

Our work stems principally from the classic work of Shannon[1] on discrete and continuous information channels. Gabor[2,3] introduced the concept of quantization into electromagnetic communication channels and coined the term "quantum noise." In consideration of the problem of field measurements by a receiver, he used an electron beam probe. The shot noise in the beam influenced his results in an important and, in the light of present knowledge, unnecessary way. Stern[4,5] has considered information rates in "photon channels." His conclusion[5] that the information efficiency of a linear amplifier can be no greater than 50 per cent conflicts with the results presented here. The major difference may be traced to the fact that he takes no account of the information that may be stored in the signal phase; and phase information approaches 50 per cent of the total possible information in the large signal-to-noise case where the quantum theory and the classical theory approach one another. Lasher[6,7] has also obtained expressions for information capacity based on quantum mechanical principles. His results agree qualitatively with ours; the quantitative differences presumably arise from the approximate methods which he used. We[8] have previously discussed some of the ideas which are utilized in this paper. In other recent work the important question of the statistical properties of quantum noise in linear amplifiers has been studied.[9,10,11]

Our ruminations will be limited to waves existing in a transmission system for which only a single transmission mode of the field is utilized. That is, the polarization and distribution of the field over any plane perpendicular to the direction of propagation are considered invariant. This situation is typical of transmisssion in a coaxial line or in a waveguide. It will also very likely be true for long-distance broad-band optical communication systems. A possible departure from such a single-mode system would involve the use of the two orthogonal field polarizations to provide two independent channels.

During the course of passage from transmitter to receiver, the signal is presumed to suffer a large attenuation and, in general, to be supplemented by some amount of additive white[12] noise power. At the receiver

* Received March 22, 1962; revised manuscript received June 7, 1962.

† Bell Telephone Laboratories, Murray Hill, N. J.

1 C. E. Shannon and W. Weaver, "The Mathematical Theory of Communication," University of Illinois Press, Urbana, Ill.; 1949.

2 D. Gabor, "Communication theory and physics," *Phil. Mag.*, vol. 41, pp. 1161–1187; 1950.

3 D. Gabor, "Lectures on Communication Theory," Res. Lab. of Electronics, M.I.T., Cambridge, Mass., Tech. Rept. No. 238; April 3, 1952.

4 T. E. Stern, "Some quantum effects in information channels," IRE Trans. on Information Theory, vol. IT-6, pp. 435–440; September, 1960.

5 T. E. Stern, "Information rates in photon channels and photon amplifiers," 1960 IRE International Convention Record, pt. 4, pp. 182–188.

6 G. J. Lasher, "A quantum statistical treatment of the channel capacity problem of information theory," in "Advances in Quantum Electronics," J. R. Singer, Ed., Columbia University Press, New York, N. Y., pp. 520–536; 1961.

7 G. J. Lasher, "Channel capacity of optical frequencies," presented at the NATO-SADTC Symp. on Technical and Military Applications of Laser Techniques, The Hague, Netherlands; April, 1962.

8 J. P. Gordon, "Information capacity of a communications channel in the presence of quantum effects," in "Advances in Quantum Electronics," J. R. Singer, Ed., Columbia University Press, New York, N. Y., pp. 509–519; 1961.

9 W. H. Wells, "Quantum formalism adapted to radiation in a coherent field," *Ann. Phys.* (*N. Y.*), vol. 12, pp. 1–40; January, 1961.

10 J. P. Gordon, W. H. Louisell, and L. R. Walier, "Quantum fluctuations and noise in parametric processes. II," to be published.

11 J. P. Gordon, W. H. Louisell, and L. R. Walker, "Quantum statistics of maser amplifiers and attenuators," to be published.

12 Since we are concerned with a very broad range of frequencies, neither thermal noise nor quantum noise is truly "white," as this would imply a uniform spectral density. Rather, the noise is "colored"; its spectral density is generally a function of frequency. Since, however, the systems we consider are all narrow band, in the sense that the bandwidth is always much smaller than the carrier frequency, the noise may be considered to be white within the bandwidth. Cases in which the spectral density of the noise varies appreciably across the band may be treated by dividing the band up into smaller segments, and treating each such segment as an independent channel.

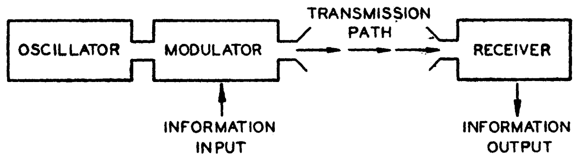

Fig. 1—Typical communication system.

as much as possible of the information remaining in the received wave is extracted. The receiver may incorporate an amplifier at the carrier frequency or it may not. We will investigate both of these cases. Fig. 1 shows a typical communications channel such as we have described.

So long as the electromagnetic waves may be described classically, *i.e.*, without quantization, Shannon[1] has shown that the information capacity C of a signal of average power S in the presence of additive white noise power N in a channel of bandwidth B is given by

$$C = B \log\left(1 + \frac{S}{N}\right) \tag{1}$$

If the logarithm is taken to the base 2, C is in units of bits per second. To realize this capacity the signal must be modulated in such a way as to have also the statistical randomness of white noise.

In deriving (1) Shannon noted that information and entropy were closely allied quantities. In fact he identified information as *prescribed* entropy. He was able to show that the entropy rate R of a continuous wave, having the statistical properties of narrow-band white noise and average power P, could be expressed as

$$R = B \log\left(\frac{P}{P_0}\right) \tag{2}$$

where the constant P_0 is arbitrary. To obtain (1) he subtracted the entropy rate for the noise alone from the entropy rate for the combined signal and noise. The latter also has the statistical properties of white noise when both signal and noise have these properties independently. Thus the constant is cancelled out and

$$C = B \log\left(\frac{S+N}{P_0}\right) - B \log\left(\frac{N}{P_0}\right) = B \log\left(1 + \frac{S}{N}\right).$$

C is the additional entropy occasioned by the presence of the signal. Since the signal is completely prescribed, the added entropy is prescribed entropy, or information.

Eq. (1) says that the information capacity approaches infinity as the signal-to-noise ratio approaches infinity. This is because as the noise decreases we can make more and more accurate measurements of the state of the signal field. However, the uncertainty principle of quantum mechanics tells us that in fact we cannot measure a field to arbitrary accuracy, and so as $N \to 0$, fundamental quantum limitations on information capacity make their appearance.

II. Entropy of White Noise

The fact that an electromagnetic wave is quantized allows us to obtain an absolute value for its entropy without the arbitrary constant of (2). Consider the wave in a transmission line traveling toward the receiver. Assume that the wave velocity is v and that there is no dispersion. Then, in time t, the receiver measures the field which had previously occupied a length $L = vt$ of the line. To describe this field we can expand it into a series of orthogonal modes, and then measure the state of excitation of each mode as well as possible. A commonly used expansion is a spatial Fourier series. For this expansion the qth mode varies with distance and time according to the exponential factor

$$\exp\left[jq\frac{2\pi}{L}(z - vt)\right].$$

The condition for orthogonality of the modes is that the different values of q differ by integers. It is also clear from the above expression that the mode q has frequency qv/L. Thus the frequency separation between adjacent modes is $\Delta\nu = v/L$. In a bandwidth B there are $B/\Delta\nu = BL/v$ orthogonal modes. Since $L = vt$ we see that in time t the receiver measures the state of excitation of Bt such modes. The rate of arrival of independent field modes at the receiver is therefore B.

The complete description of the field requires measurement of the state of excitation of each mode. Classically this would involve independent simultaneous measurements of the amplitude and phase of each mode, or equivalently simultaneous measurement of the electric and magnetic fields associated with each mode. Thus, classically, we make $2B$ independent measurements per second to identify the wave. In quantum mechanics the measurements of electric and magnetic fields are not independent, so we must consider that we make only B independent measurements per second, each measurement specifying the state of one particular field mode.

Now we know that a white noise wave must have the most random possible excitation of the various modes consistent with the average power in the wave. This allows us to calculate the entropy of such a wave. Let us specify the state of each mode by assigning to it exactly m photons, *i.e.*, an excitation energy $mh\nu$. From statistical mechanics[13] we know that the entropy per mode for a large number of modes is given by the expression

$$H = -\sum_m p(m) \log p(m)$$

where $p(m)$ is the probability that a mode will contain just m photons. The average energy per mode is given by

$$\bar{E} = h\nu\bar{m} = h\nu \sum_m mp(m)$$

and of course since $p(m)$ is a probability, the $p(m)$'s must fulfill the requirement that

$$\sum_m p(m) = 1.$$

To find the most random possible excitation consistent with a given average power, we must maximize H by varying the probabilities $p(m)$ while keeping $\Sigma p(m)$ and $\Sigma mp(m)$ constant. This is a simple problem in the calculus of variations. The set of $p(m)$ which maximize H are

$$p(m) = \frac{1}{1+\bar{m}}\left(\frac{\bar{m}}{1+\bar{m}}\right)^m.$$

The average power P in this wave is

$$P = EB = \bar{m}h\nu B$$

since $E = \bar{m}h\nu$ is the average energy per mode and B modes per second are incident on the receiver. This exponential probability distribution for the excitation of the modes is consistent with the exponential power distribution which we know is characteristic of white noise. The entropy per mode for white noise is thus

$$\begin{aligned} H &= -\sum p(m) \log p(m) \\ &= \sum p(m)\left[\log(1+\bar{m}) + m \log\left(\frac{1+\bar{m}}{\bar{m}}\right)\right] \\ &= \log(1+\bar{m}) + \bar{m}\log\left(1+\frac{1}{\bar{m}}\right). \qquad (3) \end{aligned}$$

Since $\bar{m} = P/h\nu B$ where P is the average power in the wave, we may express the entropy per mode as

$$H = \log\left(1+\frac{P}{h\nu B}\right) + \frac{P}{h\nu B}\log\left(1+\frac{h\nu B}{P}\right).$$

One may object that the specification of the excitation of each mode in terms of exact numbers of photons is not the only possible way. However, the number of distinguishable excitations within an energy range from E to $E+\Delta E$ should be independent of the quantities used for the field specification, and so we are free to choose the most convenient specification, as we have done. Finally we note that the rate of arrival of entropy at the receiver for a white noise wave is

$$R = HB = B\log\left(1+\frac{P}{h\nu B}\right) + \frac{P}{h\nu}\log\left(1+\frac{h\nu B}{P}\right). \qquad (4)$$

Eq. (4) is the quantum equivalent of (2).

Of the terms in (4) the first has a form quite similar to the classical expression and predominates when the average number of photons per mode is large compared to unity. We can call it the mode entropy. It is equal to the rate of arrival of modes, B, times the logarithm of $\bar{m}+1$, which may be thought of rather loosely as the number of frequently occurring mode occupation numbers in a typical noise wave. By mode occupation number we mean the number of photons in the mode.

The second term of (4) is of fundamental quantum origin. It is the predominant term at power levels less than $h\nu B$ where the mean occupation number $\bar{m}$ becomes less than unity. We can call it the photon entropy. It is equal to the rate of arrival of photons, $P/h\nu$, times the logarithm of the number of frequently occurring intervals (*i.e.*, modes) for each photon. We shall see that at least part of this entropy can take the form of information which is recoverable if we use a photocell or some other energy-sensitive device as a receiver.

If we approach classical theory by the frequently used artifice of supposing that h becomes very small, it may be seen that (4) approaches (2) with the arbitrary constant evaluated as

$$P_0 = h\nu B/e$$

where e is the Naperian base for natural logarithms. Since the arbitrary constant contains h, it is clear that it could not be determined from a classical description.

III. Entropy and Information

In Section II we found an absolute expression for the entropy of white noise, utilizing a particular quantum mechanical description of the possible excitations of the field modes. It is not obvious, however, that all of this entropy can be prescribed as a signal, and so constitute information. This is not to say that we cannot modulate a CW carrier wave in such a way as to give the resulting wave the statistical properties of white noise in the prescribed bandwidth B, but rather that there is very likely some part of the resulting entropy which is essentially irretrievable as information. We must confess that we do not know at present the answer to this problem. In any event the entropy of the wave is certainly an upper limit to the amount of information it may contain, and as such it is a useful quantity.

[13] R. C. Tolman, "The principles of statistical mechanics," Oxford University Press, Oxford, England, 1938. See also Shannon and Weaver.[1]

IV. Information Capacity in the Presence of Additive Noise

Suppose we have a signal with average power S accompanied by additive white noise with average power N. Following the ideas of Shannon we note that the information in the wave can be no greater than the entropy of the combination of signal plus noise less the uninformative entropy of the noise alone. The entropy of the combined signal and noise is maximized when the total wave has the statistics of white noise. Quantum mechanically as well as classically, this implies that the signal alone should also have the characteristics of white noise. The entropy rate for the combined wave is then given by (4) with $P=S+N$, while the entropy rate for noise alone has $P=N$. The upper limit to the information in the wave, which we will label C_{wave}, for a signal of average power S in the presence of white noise of average power N is thus given by

$$C_{\text{wave}} = R_{(P=S+N)} - R_{(P=N)}$$

or

$$C_{\text{wave}} = B \log\left(1 + \frac{S}{N + h\nu B}\right) + \frac{S+N}{h\nu}\log\left(1 + \frac{h\nu B}{S+N}\right) - \frac{N}{h\nu}\log\left(1 + \frac{h\nu B}{N}\right). \quad (5)$$

For a bandwidth of 10^9 cps and an additive noise power N taken as arising from a black body at 290°K, *i.e.*,

$$N = h\nu B\left[\exp\left(\frac{h\nu}{290k}\right) - 1\right]^{-1},$$

the information limit, C_{wave}, is plotted in Fig. 2 as a function of frequency for power levels ranging from 10^{-7} to 10^{-13} watt.

A. Classical Limit

If the noise power N is considerably greater than $h\nu B$, we have a situation where a classical description of the wave should be adequate. Expansion of (5) to first order in the small quantities $h\nu B/N$ and $(h\nu B)/(S+N)$ yields

$$C_{\text{wave}} = B\left[\log\left(1 + \frac{S}{N}\right) - \frac{h\nu BS}{2N(S+N)}\log e \cdots\right].$$

Under the assumed condition $N \gg h\nu B$, the second term is always much smaller than the first, independent of the value of S/N, so the classical description which results in (2) is quite good.

If there is no additive noise, but the signal is much larger than $h\nu B$, we find

$$C_{\text{wave}} = B\left[\log\left(1 + \frac{S}{h\nu B}\right) + \log e + \cdots\right].$$

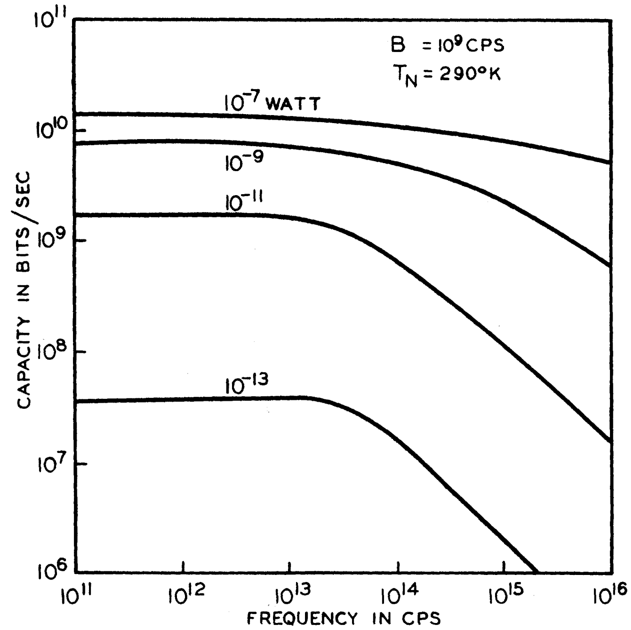

Fig. 2—Upper limit to the information that may be incorporated into an electromagnetic wave in a single transmission mode. Thermal noise, as originating from a black body at 290°K, is assumed to accompany the wave.

In the limit of very high signal power this expression is nearly the same as one would obtain from the classical expression (1), by assuming the presence of an equivalent "zero-point" noise power $h\nu B/e$. Note, however, that this equivalence is not exact.

V. Information Capacity After Transmission

As our transmitted signal travels toward the receiver, it is attenuated and usually some noise power is added to it. If we assume that the added noise is white then the information capacity of the received wave is limited by (5) where S is the received signal power and N the added noise power.

VI. Information Capacity After Coherent Amplification

Suppose now that the first element of the receiver is an amplifier at the carrier frequency. This could be a maser, a nondegenerate parametric amplifier or any other type of linear amplifier. Assume that the amplifier has high gain. There is always internal white noise generated in such an amplifier which, referred to the input, may be described by an effective input noise, N_{eff}. In the case of the maser this noise is known to be

$$N_{\text{eff}} = Kh\nu B,$$

where $K = n_2/(n_2 - n_1)$ and n_2 and n_1 are, respectively, the densities of upper-state and lower-state atoms in the active medium. In terms of a negative temperature of the active medium T_m, we have

$$K = \left[1 - \exp\left(-\frac{h\nu}{k|T_m|}\right)\right]^{-1}.$$

For the parametric amplifier N_{eff} may be written in a similar way, with K also greater than or equal to unity.[14] After much amplification the additive noise, given by the gain times the sum of the incident noise plus the effective input noise,[15] is *always* much greater than $h\nu B$ and so the classical formula applies for the information capacity. We find, therefore, that after much amplification the information capacity of the wave is reduced to

$$C_{\mathrm{amplifier}} = B \log \left(1 + \frac{S}{N + Kh\nu B}\right) \qquad (6)$$

where S is the incident signal, N is the incident noise and $K \geq 1$. Thus the best possible amplifier, for which $K = 1$, retains only the first term in the incident wave information limit, (5). We now can define the information efficiency of an amplifier as $C_{\mathrm{amplifier}}/C_{\mathrm{wave}}$. For the interesting case of a perfect amplifier this is plotted for various values of signal strength in Fig. 3. The incident noise is assumed the same as for Fig. 2.

After much amplification we may assume that all of the information remaining in the wave can be extracted, so (6) also gives the information capacity of a system using a high gain coherent amplifier at the carrier frequency as the first element of the receiver.[16] For small N the efficiency drops off for signal levels less than about $h\nu B$, indicating a substantial loss of information in this region for such a system.

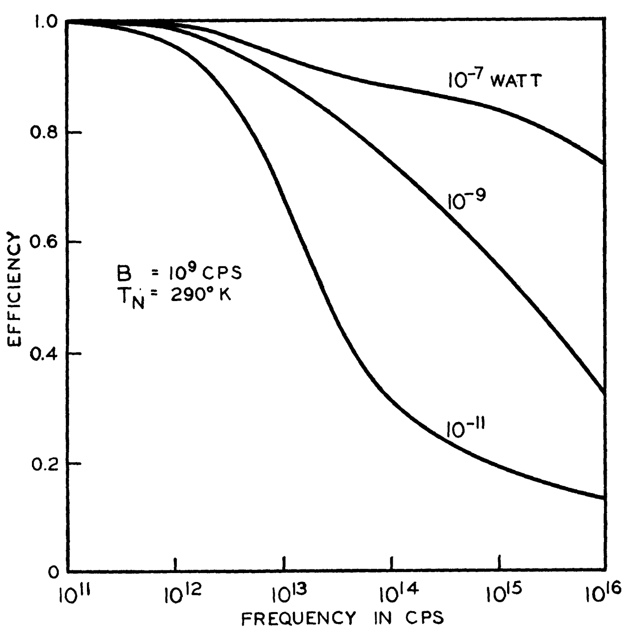

Fig. 3—Information efficiency for an ideal amplifier of high gain. Because of spontaneous emission, the ideal amplifier has an effective input noise power of $h\nu B$, which is responsible for the lowering of its efficiency at high frequencies.

VII. The Heterodyne Receiver

Instead of amplifying the wave we might immediately make use of a photoelectric device in a heterodyne receiver.[17,18] To do this we might let the signal and power from a CW local oscillator fall simultaneously on a photosensitive element.

Then the photocurrent is proportional to the instantaneous power P_{inst} incident on the element. If the quantum efficiency of the photosensitive device is ϵ, the current is given by

$$I = \frac{\epsilon P_{\mathrm{inst}}}{h\nu} q$$

where q is the electronic charge. Let the signal frequency be ω_{sig} and the local oscillator frequency be ω_{local}. If the local oscillator power is much greater than the signal power, the instantaneous power will have the form

$$P_{\mathrm{inst}} \approx P_{\mathrm{local}} + 2\sqrt{P_{\mathrm{sig}} P_{\mathrm{local}}} \cos (\omega_{\mathrm{sig}} - \omega_{\mathrm{local}})t + \cdots$$

where P_{sig} is the instantaneous input signal power and P_{local} is the local oscillator power. The photocurrent thus consists of a dc component

$$I_0 = \frac{\epsilon q}{h\nu} P_{\mathrm{local}}$$

and a signal current at the intermediate frequency whose mean square is

$$\overline{I_{\mathrm{sig}}^2} = 2\left(\frac{\epsilon q}{h\nu}\right)^2 S P_{\mathrm{local}}$$

where S is the average input signal power. Because of the dc current there will be shot noise, whose mean square is

$$\overline{I_N^2} = 2qI_0B = 2\left(\frac{\epsilon}{h\nu}\right) q^2 P_{\mathrm{local}} B.$$

[14] W. H. Louisell, A. Yariv, and A. E. Siegman, "Quantum fluctuations and noise in parametric processes. I," *Phys. Rev.*, vol. 124, pp. 1646–1654; December, 1961.

[15] It has now been established[9,10,11] that the simple addition (voltagewise) of the amplified effective input noise to the classically amplified signal and real input noise accounts for all fluctuations in the output wave. That is, if the signal wave leaving the transmitter has the form $v_s \cos(\omega_s t + \phi_s)$, then the amplified wave has the form $(G/L)^{1/2} v_s \cos(\omega t + \phi_s) + G^{1/2} v_n \cos(\omega t + \phi_n)$, where G and L are the gain and loss of the amplifier and attenuator, respectively, and where the added term in the amplified wave is the fluctuating white noise voltage. This is rigorously true no matter how small, in terms of quanta per mode, the signal may be at the amplifier input. We are of course assuming that the gain and loss are not subject to fluctuations caused by such things as variations in the density of attenuating or amplifying particles, variations in pumping of a parametric amplifier, etc.

[16] It might appear that we are departing somewhat from common usage here by speaking of the information capacity of a system using a specific receiver. The reason for it is that in quantum mechanics the properties of the measuring apparatus (*i.e.*, the receiver) inevitably influence to some extent the quantities to be measured. Thus, while we can obtain from entropy considerations an upper limit to the capacity of *any* system, from which we may derive "efficiencies" for particular systems, this upper limit cannot be termed a capacity. It would seem that we cannot obtain any expression which might properly be called a channel capacity unless we include as an essential part of the channel such elements of the receiver as are necessary to insure that subsequent measurement can be performed with no further appreciable reaction back on the channel itself.

[17] A. Javan and R. Kompfner, private communication.

[18] B. M. Oliver, "Signal-to-noise ratios in photoelectric mixing," Proc. IRE, vol. 49, pp. 1960–1961; December, 1961.

The ratio $\overline{I_{sig}^2}/\overline{I_N^2}$ is the signal-to-noise ratio at the IF, which comes out to be simply $\epsilon S/h\nu B$. This implies an information capacity for the IF signal of

$$C_{\text{heterodyne}} = B \log\left(1 + \epsilon\frac{S}{h\nu B}\right).$$

It is not difficult to include the effect of incident additive noise coming in with the signal. This simply reduces the signal-to-noise ratio at the IF to

$$\frac{S}{N + \frac{1}{\epsilon} h\nu B}$$

and the information capacity to

$$C_{\text{heterodyne}} = B \log\left[1 + \frac{S}{N + \frac{1}{\epsilon} h\nu B}\right].$$

The information capacity of a system using a heterodyne receiver thus has the same form as that of a system using a coherent amplifier, with K replaced by ϵ^{-1}.

VIII. The Homodyne Receiver

It was pointed out by B. M. Oliver[18,19] that the homodyne receiver has quite interesting properties. In this case we confine the modulation to *amplitude* modulation, along with an allowed phase shift of π, and then use a local oscillator in the receiver which has exactly the same frequency and phase as the signal carrier. Since $\cos(\omega_{sig} - \omega_{local})t$ is then always equal to ± 1, the instantaneous power incident on the photocell is

$$P = P_{local} + 2\sqrt{P_{sig}}\sqrt{P_{local}} + \cdots,$$

where the quantity $\sqrt{P_{sig}}$ may range through positive and negative values according to the modulation amplitude and phase. The dc component of the photocurrent is again

$$I_0 = \frac{\epsilon q}{h\nu} P_{local}.$$

For this case, however, the signal current is at baseband and has bandwidth $B/2$, where B is the high-frequency band used for transmission. The mean-square shot current at baseband is therefore

$$\overline{I_N^2} = 2qI_0(B/2) = \left(\frac{\epsilon}{h\nu}\right) q^2 P_{local} B,$$

while the mean-square signal current is now

$$\overline{I_{sig}^2} = 4\left(\frac{\epsilon q}{h\nu}\right)^2 S P_{local}$$

where again S is the average signal power, *i.e.*, the average of P_{sig}. The signal-to-noise ratio is therefore

$$\overline{I_{sig}^2}/\overline{I_N^2} = 4\frac{\epsilon S}{h\nu B},$$

and so the information capacity of the baseband signal is

$$C_{\text{homodyne}} = \frac{B}{2} \log\left(1 + 4\frac{\epsilon S}{h\nu B}\right).$$

As in the heterodyne case we may include incident noise without too much difficulty. The result is

$$C_{\text{homodyne}} = \frac{B}{2} \log\left[1 + \frac{2S}{N + \frac{1}{2\epsilon} h\nu B}\right]$$

where N is the average received noise in the high-frequency band B.

Oliver pointed out that in this case the equivalent input quantum noise is only half as large as that occurring in the heterodyne receiver or in the equivalent maser. At first sight this is somewhat curious. In fact it simply indicates that perhaps one cannot always deduce the effects of quantum noise simply on the assumption of some fixed equivalent input noise which is the same in all situations. In no case is the capacity of a system using a homodyne receiver greater than the capacity limit, (5), of a wave of average power S in the presence of the average incident noise N. Such a result would be truly surprising. In Fig. 4 the information efficiency for

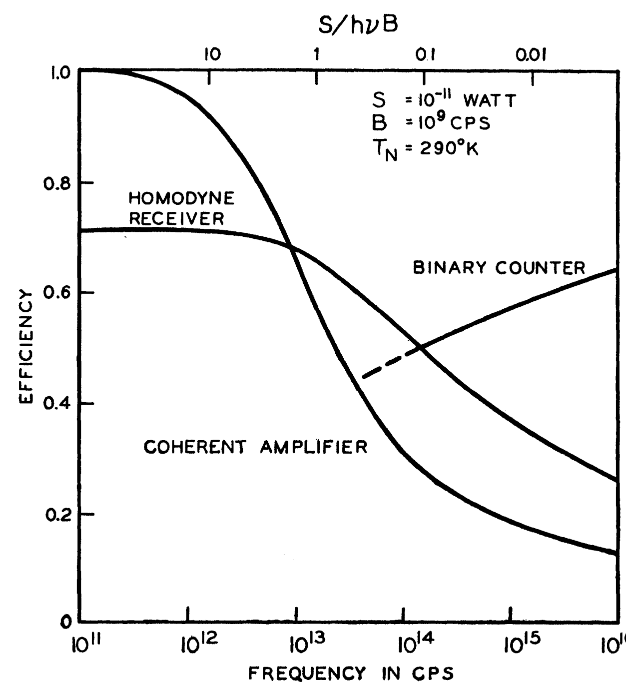

Fig. 4—Information efficiency for various receivers for an average received signal power of 10^{-11} w, a bandwidth of 10^9 cps, and an external noise temperature of 290°K. Note that at the higher frequencies the coherent amplifier is not as good as the other types of receivers.

[19] B. M. Oliver, "Comments on 'Noise in photoelectric mixing,'" Proc. IRE (*Correspondence*), vol. 50, pp. 1545–1546; June, 1962.

an ideal ($\epsilon = 1$) homodyne system is also plotted against frequency, for a signal power of 10^{-11} w and an external noise temperature of 290°K. For comparison the information efficiency of an ideal amplifier is plotted also, as well as that of an ideal detector using a binary quantum counter (see Section IX-A, and note that for frequencies of 10^{14} cps or greater the external noise may be completely neglected).

IX. The Quantum Counter

Instead of using any of the aforementioned receivers, we might simply allow the signal to fall on some photoelectric device and count the photoelectrons as they are produced. If we could do this with unity quantum efficiency and with perfect discrimination between different numbers of photoelectrons, we would surely have an ideal power-sensitive receiver. The information capacity for this general case can in principle be found since the probability distribution for the various numbers of received photons resulting from the transmission of some known number of photons has been computed.[20] Unfortunately, attempts to calculate the information capacity of a communication system using such a receiver encounter rather great computational difficulties. Nevertheless in some simple cases the problem can be solved approximately. When $S/h\nu B$ is either much larger or much smaller than unity, we may obtain approximately correct values for the capacity.

A. The Binary Counter

For the case $S \ll h\nu B$, the average number of photons per independent field mode is much smaller than unity, so that only the two events, no photon received or one photon received, have appreciable probabilities. Consider, then, the following communication system. The transmitted signal consists of a series of pulses, each of duration $1/B$ and of constant amplitude. The pulses occur in a statistically random sequence with the probability Q of sending a pulse in any particular time interval. A typical transmitted message would then appear as in Fig. 5. The average power in the signal is Q times the pulse power, or if the energy in each pulse is E the average power is QEB. The receiver measures the number of received photons in each time interval $1/B$; thus it makes B measurements per second, which is consistent with the notion that there are B independent field modes received per second. If the receiver simply distinguishes between no photons received or some photons received, we will have a system which should do nearly as well as possible when the average number of photons received per interval is much smaller than unity but of course is rather inefficient for larger average numbers of photons. This system has the advantage that one can compute its information capacity exactly, and we shall now proceed to do this.

Fig. 6 shows the communications channel under consideration. In each time interval $1/B$ the transmitter either emits a pulse or it does not. The probability of occurrence of a pulse in any particular time interval is Q. If the receiver detects at least one photon in any time interval, it records a 1; if not, it records a 0. To simplify matters, let us assume that the quantum efficiency of the receiver is unity, and at first let us assume that there is no noise in the channel. In this case if the transmitter does not send a pulse, the receiver definitely records a 0. This is indicated in Fig. 6. On the other hand if the transmitter sends a pulse, the receiver does not definitely record a 1. There is a finite probability that no photons reach the receiver even when the pulse is sent. This probability is known, however. So long as the number of photons in the transmitted pulse is reasonably well known, the probability distribution $q(m)$ for the various numbers m of photons received after large transmission loss is a Poisson distribution, from which

$$q(m) = \frac{s^m}{m!} e^{-s}.$$

Here the average or expected number of received photons in the pulse is labeled s. Thus the probability of receiving no photons is e^{-s}, and the probability of receiving at least one is of course $1 - e^{-s}$. These probabilities are also indicated on Fig. 6.

Now to compute information capacity we must use some further results of Shannon's work.[1] He showed that the information I per symbol (*i.e.*, time interval)

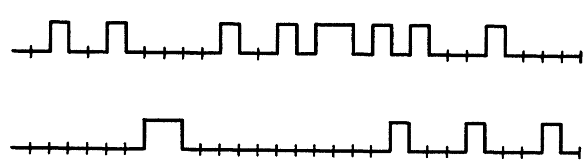

Fig. 5—Typical sequence of pulses in a message suitable for a binary communication system. The statistical probability for the occurrence of a pulse is 0.25 in this message.

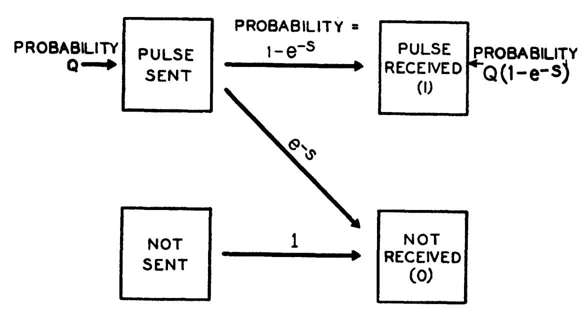

Fig. 6—Schematic diagram for a noiseless binary channel. The various probabilities necessary for the solution of the information problem are indicated on the diagram.

[20] K. Shimoda, H. Takahasi, and C. H. Townes, "Fluctuations in the amplification of quanta with applications to maser amplifiers," *J. Phys. Soc. Japan*, vol. 12, pp. 686–700; June, 1957.

for such a discrete communication channel is given by

$$I = H(y) - H_x(y)$$

where $H(y)$ is the entropy per symbol of the received message, given by

$$H(y) = -\sum p(y) \log p(y)$$

summed over the probabilities $p(y)$ for the possible received symbols y, while $H_x(y)$ is the conditional entropy of the received message, given by

$$H_x(y) = -\sum_x p(x) \sum_y p_x(y) \log p_x(y).$$

Here the quantities

$$\left[-\sum_y p_x(y) \log p_x(y)\right]$$

are the entropy per symbol of the received message when the transmitted symbol (x) is known, and $H_x(y)$ is this entropy averaged over the probability distribution $p(x)$ for transmitted symbols. Thus $H(y)$ is the total received entropy, while $H_x(y)$ is that part of the received entropy which does not contain information.

With the help of these formulas we are able to compute the information capacity of the channel. For a probability Q of sending a pulse, the total probabilities for receiving a 1 or a 0 are

$$p(1) = Q(1 - e^{-s}); \qquad p(0) = 1 - Q(1 - e^{-s})$$

while the conditional probabilities are

$$P_{\text{pulse}}(0) = e^{-s}; \; P_{\text{pulse}}(1) = 1 - e^{-s}, \; P_{\text{no pulse}}(0) = 1, \; P_{\text{no pulse}}(1) = 0.$$

The received entropy is then

$$H(y) = -Q(1 - e^{-s}) \log [Q(1 - e^{-s}] - [1 - Q(1 - e^{-s})] \log [1 - Q(1 - e^{-s})]$$

and the conditional entropy is

$$H_x(y) = -Q[e^{-s} \log e^{-s} + (1 - e^{-s}) \log (1 - e^{-s})] - (1 - Q)[0].$$

Subtracting the two, we find

$$I = -Q(1 - e^{-s}) \log Q - [1 - Q(1 - e^{-s})] \cdot \log [1 - Q(1 - e^{-s})] + Qe^{-s} \log e^{-s}.$$

To find the maximum information per symbol we must maximize I with respect to Q, under the constraint that the average power remain constant. Now since s is the average number of received photons per pulse, and Q the probability of sending a pulse, the average number of photons per time interval is Qs. This is the quantity which must remain constant and was called $\bar{m}$ in Section I. If we therefore substitute $Q = \bar{m}/s$, where $\bar{m}$ is a constant, into I, differentiate with respect to s and set the result equal to 0, we obtain the condition for maximum I. This is

$$\log_e \left[\frac{s}{\bar{m}} + e^{-s} - 1\right] = \frac{s}{\left(\dfrac{e^s}{s+1} - 1\right)} \cdot$$

To find $I_{\max}$ this transcendental equation must be solved for s, assuming some value of $\bar{m}$, and then the result used to evaluate I. In Fig. 7, s is plotted against $\bar{m}$. It may be seen that s does not drop off very rapidly for small $\bar{m}$. Finally $I_{\max}$ can then be calculated, and the information capacity of this system

$$C_{\text{binary}} = I_{\max} B$$

may be compared to the information limit for a noiseless wave of the same average power (*i.e.*, $\bar{m}h\nu B$) at the receiver input. The efficiency of the system

$$C_{\text{binary}}/C_{\text{wave}}$$

is plotted in Fig. 4. It may be seen to approach unity slowly at small signal levels. One can in fact show that at very very small signal levels, *i.e.*, for $\log_e(1/\bar{m}) \gg 1$, the information per symbol approaches

$$I_{\max} \to \bar{m} \log \frac{1}{\bar{m}} \cdot$$

This may be compared to H of (3).

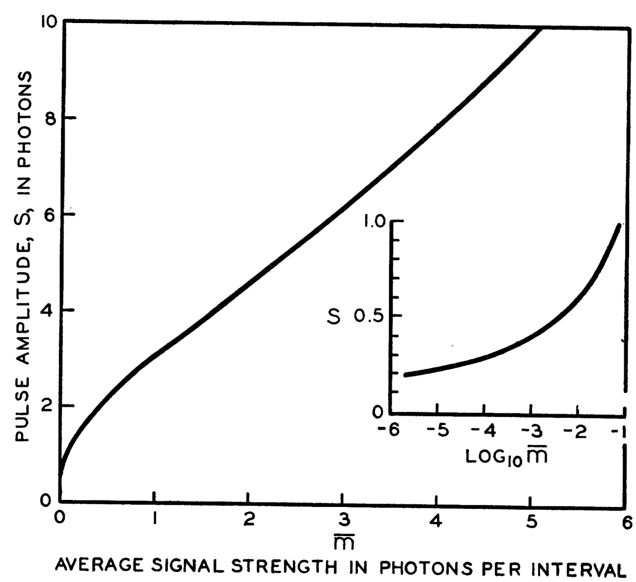

Fig. 7—Optimized average received pulse amplitude for the noiseless binary channel as a function of the average number of received photons per available time interval. The probability of sending a pulse is given by $Q = \bar{m}/s$.

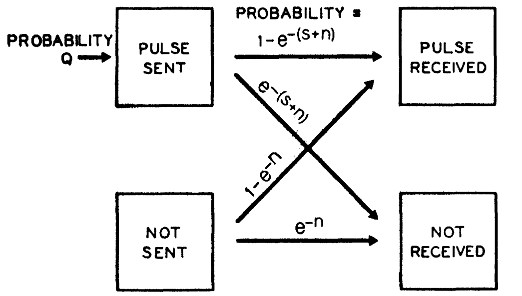

Fig. 8—Schematic diagram for a noisy binary channel.

From this fairly simple example the mathematical complexity of the quantum counter system is reasonably evident. However, the case of the binary counter with noise is also simple enough to be calculated. There are two cases of possible interest involving noise. The first is when the noise in the counter results from noise power in the transmission mode accompanying the signal. It is then important to consider the effects of interference between noise and signal. However, there seems little use in calculating information capacities for this case since, for a given noise temperature, the number of noise photons per interval $1/B$, as a function of frequency, is for the greater part of the frequency range either greater than one—when $h\nu < kT$, or much less than one—when $h\nu > kT$. In the former case the binary counter is clearly not the most efficient receiver, in the latter the noise may be ignored so long as it is much less than the signal.

The second case involving noise in the counter is when the noise and signal are statistically independent. This would occur if the noise results from dark current in the photodetector, or from the effects of stray light incident in the receiver from modes other than the transmission mode or at frequencies outside of the useful band. If we assume that the noise photoelectrons arise from a large number of statistically independent causes, then the probability distribution of noise photoelectrons is also a Poisson. From this the conditional probabilities given in Fig. 8 follow. The results of a calculation of I_{max} for these probabilities, based on the equations of the previous section, are given in Fig. 9. For comparison, I_{max} for the noiseless case, is plotted there also.

B. The Quantum Counter When $S \gg h\nu B$

In the previous subsection we considered a particular communication system using a quantum counter for which the capacity could be calculated exactly, but which approximates an ideal receiver only when $S + N \ll h\nu B$. We can also obtain an approximate result for an ideal quantum counter system which is valid at high power levels. We assume again that the transmitter sends out a sequence of pulses, each of duration $1/B$, but with varying amplitudes. We suppose that the receiver tells the exact number of photons it receives in each such time interval, and we may assume that in the great majority of intervals it receives a reasonably large number of photons.

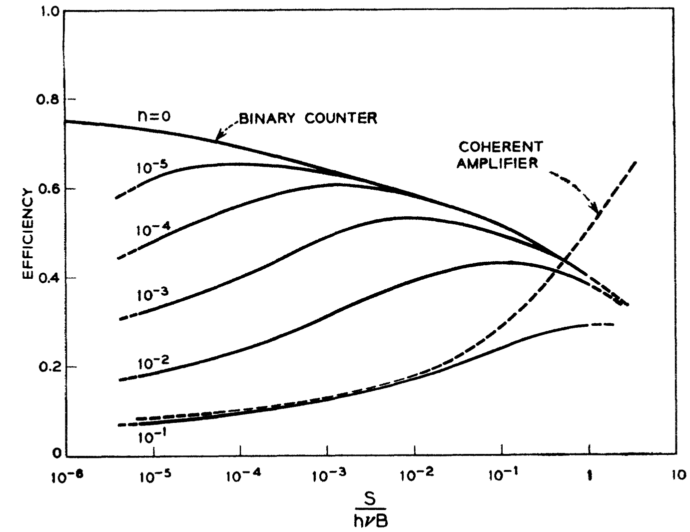

Fig. 9—Information efficiency of a binary counter perturbed by various amounts of incoherent noise. The numbers n represent the average number of noise photoelectrons per pulse interval B^{-1}. Note that the efficiency drops off when $nh\nu B > S$, and that for n greater than 0.1, the efficiency of the binary counter is always less than that of the coherent amplifier.

The calculation is carried out in the Appendix and gives the results that an exponential probability distribution for the energy of the transmitted pulses is approximately optimum, and that for this distribution and no noise the information capacity of the counter system is [see (13)]

$$C_{\text{counter}} = \tfrac{1}{2} C_{\text{wave}} - B[\log \sqrt{2\pi} + 0.289 \log e]. \qquad (7)$$

In the limit of very high power the constant term can be neglected, and the quantum counter then extracts half of the information in the wave. It is likely that when the wave capacity is small enough so that the second term begins to be significant, the exponential distribution is no longer optimum.

Having gone this far we perhaps should go on to add noise to the wave and again calculate the information capacity. In fact one can do so using similar approximate methods. Again one finds that in the limit of high power the counter system achieves half the capacity of the wave. The calculation is much like that in the Appendix, and in order not to bore the reader excessively, we shall omit it here.

The fact that a system using an energy-sensitive receiver has a capacity no greater than half of the wave capacity in the limit of high power (*i.e.*, high signal-to-noise ratio) is just what we might expect. In this limit the classical theory should give an adequate description of physical phenomena. Classically, when the signal-to-noise ratio is high, then equal amounts of information may be obtained from measurements of amplitude and measurements of phase; and the energy sensitive receiver automatically rejects all phase information.

X. Summary

We have found an expression for the absolute rate at which entropy is carried by an electromagnetic wave having the statistical properties of white noise, in a transmission medium which supports a single transmission mode. Further, we have found an upper limit for the information capacity of a wave consisting of signal with average power S in the presence of white noise with average power N. We have investigated the information capacity of a number of communication systems. The results of this investigation may be summarized as follows:

1) When the received signal or noise power is much larger than $h\nu B$, where ν is the center frequency of the wave and B its bandwidth, a receiver using an ideal coherent amplifier or an ideal heterodyne converter can extract essentially all the information that can be incorporated in the wave, while an ideal energy-sensitive receiver is limited to about half the capacity of the wave. The ideal homodyne converter is intermediate between these two.
2) When the total received power is much less than $h\nu B$, a binary quantum counter can extract essentially all the information that can be incorporated in the wave, while the other types of receivers become increasingly less efficient.
3) For a given power and bandwidth, the upper limit to the information which can be incorporated in an electromagnetic wave begins to drop off fairly rapidly when ν increases beyond P/hB. Viewed from another angle, for a given frequency and bandwidth there is a kind of threshold for received power below which the information capacity of a communications channel drops off rapidly. When external noise is absent, this power level is about $h\nu B$.

Appendix

The Quantum Counter When $S \gg h\nu B$

Our first step will be to calculate the conditional entropy $H_x(y)$. If the transmitter sends a pulse of M photons in a particular interval with any reasonably small uncertainty and there is no additive noise, the probability distribution for received photons is, as before, known to be a Poisson; that is, the probability of reception of m photons when M were sent is

$$P_M(m) = \frac{(\bar{m})^m e^{-\bar{m}}}{m!}\cdot$$

Here $\bar{m}$ is the expected number of received photons. $\bar{m}$ is of course MT, where T is the transmission coefficient of the transmission line. By supposition, T is much less than unity. The conditional entropy of the received signal, $H_M(m)$, may then be written as

$$H_M(m) = -\sum_M p(M) \sum_m P_M(m) \log P_M(m).$$

Now

$$\log P_M(m) = m \log \bar{m} - \bar{m} \log e - \log (m!)$$

and since we assume m is large in the great majority of instances we may use Stirling's approximation for $\bar{m}!$, which is

$$\log m! \approx (m + \tfrac{1}{2}) \log m - m \log e + \log \sqrt{2\pi}.$$

Using this relation we find

$$\log P_M(m) \approx \log \bar{m} - (m + \tfrac{1}{2}) \log m - (m - \bar{m}) \log e - \log \sqrt{2\pi},$$

whence

$$\sum_m P_M(m) \log P_M(m) = \bar{m} \log \bar{m} - \log \sqrt{2\pi} - \sum_m P_M(m)(m + \tfrac{1}{2}) \log m.$$

It remains to evaluate the last term. To do this we expand $\log m$ in a power series in $(m-\bar{m})/\bar{m}$, according to the prescription

$$\log m = \log \bar{m} + \log \left[1 + \frac{m - \bar{m}}{\bar{m}}\right] = \log \bar{m} + \left[\frac{m - \bar{m}}{\bar{m}} - \frac{1}{2}\frac{(m - \bar{m})^2}{\bar{m}^2} + \cdots\right] \log e.$$

We can then make the necessary summation in terms of the moments of the Poisson distribution $P_M(m)$, for which we know that

$$\sum_m P_M(m)(m - \bar{m}) = 0, \qquad \sum_m P_M(m)(m - \bar{m})^2 = \bar{m}$$
$$\sum_m P_M(m)(m - \bar{m})^3 = \bar{m}, \qquad \text{etc.}$$

Doing this we find that

$$\sum_m P_M(m)(m + \tfrac{1}{2}) \log m = (\bar{m} + \tfrac{1}{2}) \log \bar{m} + \tfrac{1}{2} \log e + O(1/\bar{m}).$$

Substituting this in (8) we find the relation

$$-\sum_m P_M(m) \log P_M(m) = \tfrac{1}{2} \log (2\pi e \bar{m}) + O\left(\frac{1}{\bar{m}}\right),$$

and so the conditional entropy is given very nearly by

$$H_M(m) = \frac{1}{2} \sum_M p(M) \log (2\pi e \bar{m}).$$

Since M is exceedingly large over most of its significant range, we can replace this summation by an integral, thus

$$H_M(m) \approx \frac{1}{2} \int_0^\infty dM p(M) \log (2\pi e \bar{m})$$

and finally since $\bar{m}$ is a known function of M, we have

$$H_M(m) \approx \frac{1}{2} \int_0^\infty d\bar{m} p(\bar{m}) \log (2\pi e \bar{m}) \qquad (9)$$

where $p(\bar{m})$ is the probability distribution of the expected value of the number of photons incident on the receiver and $p(\bar{m})d\bar{m}=p(M)dM$.

Now let us ask how the conditional entropy varies with the choice of the expected signal distribution $p(\bar{m})$. We must know this in order to maximize the information content of the signal. As before we can expand $\log(2\pi e\bar{m})$ in a power series in $\bar{m}-\bar{\bar{m}}$, where $\bar{\bar{m}}$ is the average value of $\bar{m}$ over the distribution $p(\bar{m})$. Thus

$$\log 2\pi e\bar{m} = \log 2\pi e\bar{\bar{m}} + \log\left(1+\frac{\bar{m}-\bar{\bar{m}}}{\bar{\bar{m}}}\right)$$
$$= \log 2\pi e\bar{\bar{m}} + \log e \cdot\left[\left(\frac{\bar{m}-\bar{\bar{m}}}{\bar{\bar{m}}}\right)-\frac{1}{2}\left(\frac{\bar{m}-\bar{\bar{m}}}{\bar{\bar{m}}}\right)^2+\frac{1}{3}\left(\frac{\bar{m}-\bar{\bar{m}}}{\bar{\bar{m}}}\right)^3-\cdots\right]$$

where

$$\bar{\bar{m}} = \int_0^\infty \bar{m}p(\bar{m})d\bar{m}.$$

Substituting this in (9), we find

$$H_M(m) \cong \tfrac{1}{2}\log 2\pi e\bar{\bar{m}} + \tfrac{1}{2}(\log e) \cdot\int_0^\infty\left[-\frac{(\bar{m}-\bar{\bar{m}})^2}{2\bar{\bar{m}}^2}+\frac{1}{3}\frac{(\bar{m}-\bar{\bar{m}})^3}{\bar{\bar{m}}^3}-\cdots\right]p(\bar{m})d\bar{m}. \quad (10)$$

It is clear from (10) that if $\bar{\bar{m}}$ is large, and if the distribution $p(\bar{m})$ is any reasonably sharp distribution, the conditional entropy is very nearly given by the first term alone; *i.e.*,

$$H_M(m) \approx \tfrac{1}{2}\log(2\pi e\bar{\bar{m}})$$

and thus is dependent only on the average signal power. Thus the problem of maximizing the information in the signal subject to a given average power (therefore, a given value of $\bar{\bar{m}}$) reduces simply to the problem of maximizing the received entropy. This we already know how to do. It requires an exponential probability distribution for received photons. For this distribution the received entropy is given by (3) with $\bar{\bar{m}}$ replacing $\bar{m}$, and for large $\bar{\bar{m}}$ may be approximated by

$$H(m) \cong \log\bar{\bar{m}} + \log e = \log(e\bar{\bar{m}}). \quad (11)$$

The most straightforward way to obtain an exponential probability distribution at the receiver is to generate an exponential probability distribution at the transmitter. Our final task is then to check whether our result for the conditional entropy at the receiver is valid for this very broad distribution as well as for narrow ones. For the exponential distribution with a reasonably large average, the probability distribution for the expected number of received photons may be assumed to be very nearly continuous and given by

$$p(\bar{m}) = \frac{1}{\bar{\bar{m}}}\exp\left(-\frac{\bar{m}}{\bar{\bar{m}}}\right).$$

For this distribution the series expansion (10), for the conditional probability converges embarrassingly slowly, so we must go back to the integral form (9). Doing this we obtain for the conditional entropy

$$H_M(m) \approx \frac{1}{2}\int_0^\infty d\bar{m}\left(\frac{1}{\bar{\bar{m}}}\right)\exp\left(-\frac{\bar{m}}{\bar{\bar{m}}}\right)\log(2\pi e\bar{m}).$$

Substituting $x=\bar{m}/\bar{\bar{m}}$ we find

$$H_M(m) \approx \frac{1}{2}\log(2\pi e\bar{\bar{m}}) + \frac{1}{2}\int_0^\infty dx\exp(-x)\log x.$$

The integral evaluates to 0.577 log e, so that

$$H_M(m) = \tfrac{1}{2}\log(2\pi e\bar{\bar{m}}) + 0.289\log e. \quad (12)$$

Comparison of this result with (10) shows that by going to the broad exponential distribution we have slightly increased the conditional entropy, but probably not enough to invalidate the conclusion that for large $\bar{\bar{m}}$ the exponential distribution is the optimum one.

Finally, we obtain for the information per symbol obtained by the ideal quantum counter,

$$I = H(m) - H_M(m)$$

where $H_M(m)$ is given by (12) and $H(m)$ by (11). We can express this as

$$I = \tfrac{1}{2}H(m) - (\tfrac{1}{2}\log 2\pi + 0.289\log e).$$

Since $BH(m)$ for thise case is just the information capacity of the wave, C_{wave}, we find for the information capacity of the ideal quantum counter, at high power levels, the expression

$$C_{\text{counter}} = BI \approx \tfrac{1}{2}C_{\text{wave}} - B(\tfrac{1}{2}\log 2\pi + 0.289\log e). \quad (13)$$

Acknowledgment

The author is pleased to acknowledge the able assistance of Mrs. C. A. Lambert in carrying out the necessary numerical computations; and many thanks to J. R. Pierce, for his continuing interest in the whole subject.