

Chapter 12

Classical Communication Systems

In optical communication systems, a weak optical signal attenuated by fiber loss is amplified by an optical linear amplifier before the signal is detected by a photodetector and electronically regenerated. An optical amplifier adds extra noise on the signal and thus the signal-to-noise (S/N) ratio of the optical signal wave is degraded. However, an electronic regeneration circuit has much higher thermal noise than the noise generated in an optical amplifier; thus, the use of an optical amplifier improves the overall S/N ratio.

The use of an optical amplifier as an analog repeater is a very useful technique for expanding an electronic regenerative terminal repeater spacing. The S/N ratio degradation due to optical amplifier noise becomes less and less important when more optical amplifiers are added into the system. The distinct difference between an optical preamplifier and optical repeater amplifier is discussed in this chapter. An alternative technique of optical receiver design, such as avalanche photodiode direct detection or coherent homodyne/heterodyne detection, is also reviewed.

We will conclude this chapter with the channel capacity and the fundamental limit of optical communication systems.

12.1 Regeneration in Digital Communication Systems

In digital communication systems, the information is encoded on the binary states of either an amplitude, phase, or frequency of the carrier wave. These modulation schemes are called amplitude-shift-keying (ASK), phase-shift-keying (PSK) and frequency-shift-keying (FSK), respectively. A transmission line such as an optical fiber attenuates the signal power through its loss and broadens the signal pulse through its dispersion. Once the signal is weakened, distorted and buried in the background noise, the information cannot be retrieved. One of the most important advantages of digital communication systems is that, before information is completely lost due to the transmission line loss and dispersion, the signal is detected and a clean pulse is regenerated.

How often it is necessary to regenerate the signal is an important factor for evaluating the system because the cost and maintenance of the system critically depend on the number of such regenerative repeaters. The signal-to-noise (S/N) ratio is gradually degraded by the transmission loss but the S/N ratio can be improved after regeneration. The penalty for the improvement of the original S/N ratio is a finite rate of erroneous

regeneration. Even though the transmitted signal was state “0,” a decision circuit may report state “1” and so the signal pulse of state “1” is transmitted to the next section. A definite advantage of the regeneration in digital communication systems is that the error rate increases only linearly with the number of regenerative repeaters, because the bit error at each regenerative repeater is accumulated only additively. For instance, if each regenerative repeater has a bit error rate of $P_e = 10^{-9}$ while the total bit error rate of the system should be less than $P_e = 10^{-6}$, one can install 10^3 regenerative repeaters in the whole system. The total system length for $P_e = 10^{-6}$ can thus be increased 10^3 times longer than the transmission length for $P_e = 10^{-9}$. Without a regenerative repeater, the bit error rate increases exponentially with transmission loss, and thus there is no substantial difference between the system length for $P_e = 10^{-9}$ and that for $P_e = 10^{-6}$. In fact, the required optical signal powers to achieve $P_e = 10^{-6}$ and $P_e = 10^{-9}$ differ by only ~ 2 dB, which corresponds to only ~ 10 Km of transmission distance.

12.2 Direct Detection in PCM-IM Communication Systems

12.2.1 Fundamental Limit

Consider a communication system in which information is carried by the presence or absence of photons per pulse. This is called pulse code modulation-intensity modulation (PCM-IM) signal. As shown already in Chapter 2 and Chapter 9, the optical pulse from a laser transmitter does not necessarily contain the fixed number of photons, but, rather, the photon number fluctuates from pulse to pulse due to the intensity noise of a laser and the random deletion noise in a fiber. According to the Burgess variance theorem, the laser noise is attenuated by the square of fiber loss, L^2 , so the dominant noise after large attenuation becomes a partition noise, which is Poissonian. Therefore, the on-pulse has Poissonian photon statistics with an average photon number $\langle n \rangle$,

$$p_{\text{on}}(n) = \frac{e^{-\langle n \rangle} \langle n \rangle^n}{n!} \quad , \quad (12.1)$$

while, the off-pulse always has zero photons, as shown in Fig. ?? . If a receiver is an ideal single photon counter without a dark count, the best detection strategy is to set a decision threshold at “0.5 photons”. That is, if no photons are received, assume an off-pulse was sent; if more than one photon is received, assume an on-pulse was sent. In such an ideal photon counting communication system, there is still a finite probability of error expected in this decision process because the on-pulse has a finite probability of having no photons due to Poissonian photon statistics. If one desires an error rate P_e smaller than 10^{-9} , which is a required bit error rate in a commercial system, one needs the average photon number of the on-pulse to be larger than

$$\langle n \rangle > 21 \quad \longrightarrow \quad P_e \equiv p_{\text{on}}(n=0) = e^{-\langle n \rangle} \leq 10^{-9} \quad . \quad (12.2)$$

This is the ultimate limit on the receiver sensitivity in the PCM-IM optical communication systems.

However, a practical optical receiver cannot achieve this ultimate sensitivity because there are finite detector dark count and huge thermal noise generated in the electronic

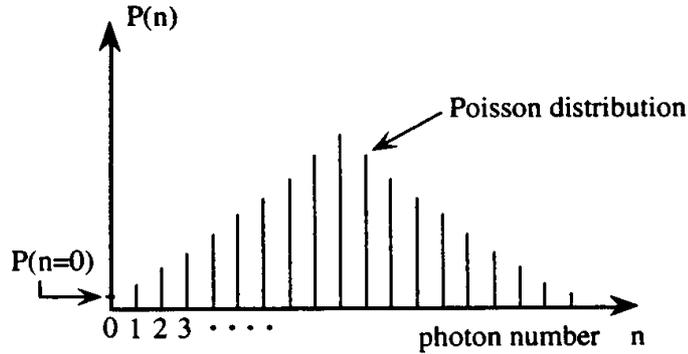


Figure 12.1: The Poissonian photon number distribution of an on-pulse.

amplifier following a photodiode. A pre-amplifier which adds extra noise to the amplified signal can still improve the overall S/N ratio because of the thermal noise of the electronic amplifier.

12.2.2 Digital Receiver System Considerations

A typical digital receiver system block diagram is shown in Fig. 12.2(a); associated waveforms are shown in Fig. 12.2(b). For now, assume a non-return-to-zero (NRZ) signal format; if a bit is a “one,” the signal is high during the entire bit interval; if a bit is zero, the signal is low during the entire bit interval. The receiver amplifier takes the photocurrent signal input, i_{ph} , and gives an output voltage, v_0 , of the same waveshape; however, this output voltage also includes noise due to the detector and the receiver amplifier. The average noise shown on v_0 is much less than the signal; however, occasionally a random fluctuation is bigger than the signal and can turn a 1 into a 0 or a 0 into a 1, causing an error. Typical systems require a bit-error rate (BER) of 10^{-9} , or one error per billion bits. The digital-receiver sensitivity is then the input optical signal power required for a signal-to-noise ratio high enough to get a BER of 10^{-9} .

The receiver output noise can be reduced by filtering; the receiver bandwidth is usually wider than necessary to pass the photocurrent signal; the extra bandwidth contains extra noise. This extra noise is removed by the channel filter, improving the signal-to-noise ratio at the input to the digital decision circuit. The total noise power on the channel filter output signal, v_f is the spectral noise power density $S(\omega)$ of the amplifier plus that of the photodetector integrated over the filter bandwidth; the lower the filter cutoff frequency, f_c the lower the noise bandwidth and the less noise at the input to the decision circuit.

The digital decision circuit recovers the digital bit stream from v_f . There are two types of decision circuits. The asynchronous decision circuit is presently cheaper and has been used in datalinks. The synchronous decision circuit, in which a clock is recovered from the signal and is used to sample v_f at the center of each zero or one, is more sensitive because it allows the use of a narrower channel-filter bandwidth, which removes more noise. In time, the synchronous circuit will be included in the receiver IC and will thus

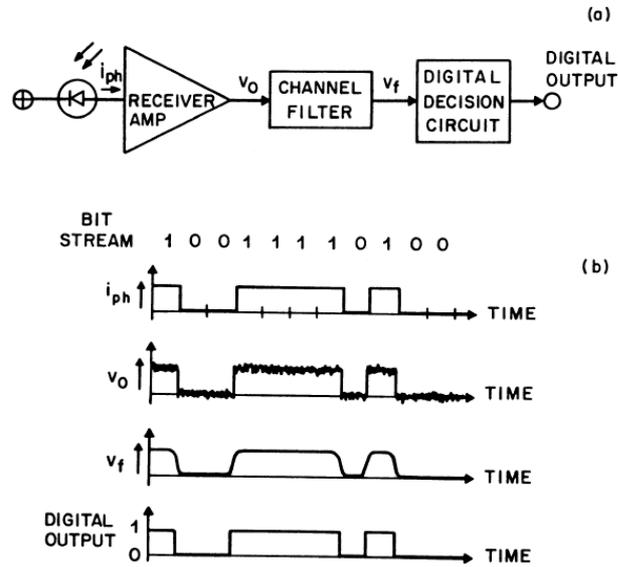


Figure 12.2: Optical receiver system: (a) block diagram, (b) typical waveforms.

cost essentially nothing.

Note that the receiver amplifier typically includes automatic gain control (AGC) so that the peak-to-peak signal at the decision circuit input is held constant, independent of the input optical signal level. This ensures proper decision circuit operation. This AGC function has typically been implemented in a postamplifier (not shown) after the input receiver amplifier in the new high-sensitivity, wide-dynamic-range amplifiers AGC is provided in the input amplifier as well.

The asynchronous decision circuit (Fig. 12.3) is just a discriminator, It gives a logic zero output when the filter output, v_f is less than the decision threshold voltage, V_T ; is gives a logic one output when v_f is greater than V_T . (Typically, V_T is midway between the “zero” and “one” analog signal levels.) Each digital transition (zero to one or one to zero) in the output bit stream occurs at the moment the signal v_f passes through V_T . Any noise as the signal transition passes through V_T will cause a transition timing error. These timing errors are minimized by maintaining fast rise and fall times on the signal v_f ; this means a wide channel-filter bandwidth, hence high noise, thus reducing the achievable sensitivity.

The channel-filter bandwidth for an NRZ asynchronous receiver would typically be about twice the bit rate B , giving a risetime of about 0.17 times the bit interval. (The 10% to 90% risetime is $\tau_r \approx 0.35/BW$).

The synchronous decision circuit (Fig. 12.4(a)) samples the filtered signal, v_f , at the center of each bit interval, as indicated by the arrows in Fig. 12.4(b), and decides whether it represents a zero ($v_f < V_T$) or a one ($v_f > V_T$). The circuit then puts out a standard length one or zero pulse exactly one bit-interval long. The sampling circuit is triggered at the center of every bit by the clock recovery circuit, which reconstructs the transmitter’s

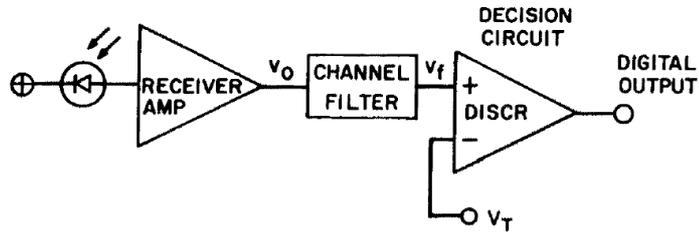


Figure 12.3: Asynchronous decision circuit optical receiver system.

digital clock signal (timing) from the received signal. The decision circuit output bit intervals are each one cycle of the recovered clock. Since the signal is sampled only in the center of each bit interval, the rise and fall times can be very slow; this means a narrow channel-filter bandwidth and hence low noise. Thus, the synchronous decision circuit gives the best digital-receiver sensitivities.

The channel-filter bandwidth for a synchronous NRZ receiver is typically set at 0.56 times the bit rate thus removing the high-frequency part of the NRZ signal spectrum. The resultant waveform, v_f , is shown in Fig. 12.4(b); the risetime is now about 0.6 times the bit interval; the waveform is still satisfactory because the time between sampling points is one bit interval. The noise bandwidth is now only 28% of that of the asynchronous receiver example; the receiver sensitivity is increased accordingly.

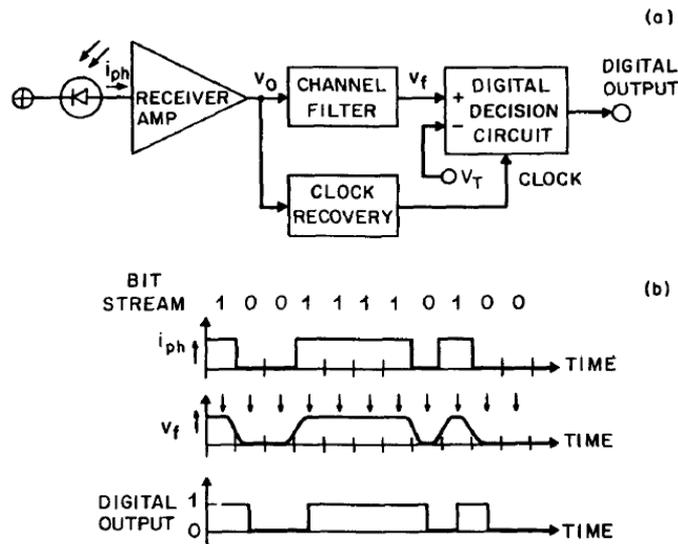


Figure 12.4: Synchronous decision circuit optical receiver system: (a) block diagram, (b) typical waveforms.

Typical optical receiver noise power spectra contain a term independent of frequency, which dominates at low frequencies and a term proportional to the frequency squared, which dominates at high frequencies. The frequency-independent noise term is due to the device leakage currents and the circuit; the frequency-squared term is primarily due to the input FET. In high-sensitivity designs, the frequency-independent term can essentially be eliminated over most of the bandwidth. Therefore, the frequency-squared noise term is dominant; integrating this term over the signal bandwidth gives a total electrical noise power proportional to the receiver bandwidth cubed. Since the synchronous receiver example has only 28% of the noise bandwidth of the asynchronous receiver, for example, the electrical noise power is only $0.28^3 = 0.022$ times as much; the noise voltage is $0.022 = 0.15$ times as much. Since the signal voltage is proportional to the optical power, the minimum optical signal is 0.15 times smaller. This is an 8.3-dB optical sensitivity advantage for the synchronous detection, example.

Until now, transmission systems have used synchronous decision circuits for maximum sensitivity; asynchronous circuits are used in some very-short-haul, low sensitivity datalinks. However, since a linear channel and synchronous detector circuit can ultimately be very economically integrated on the same IC, asynchronous systems will no longer be cheaper. This means that almost all transmission, loop, and local-area-network receivers plus many datalink receivers will use synchronous detection for better sensitivity.

So far, this discussion has assumed an NRZ optical data transmission format. Other formats such as return-to-zero (RZ), biphasic or Manchester, and block codes have been considered. Both RZ and biphasic approximately double the receiver noise bandwidth, for an approximately 4.5-dB optical sensitivity penalty; they therefore are not preferred. Block-coding schemes in NRZ format have been used with the integrating receivers. These involve a smaller sensitivity penalty. The biphasic and block-coding schemes reduce the low-frequency content of the signal, which can be helpful in ac-coupled systems or in integrating receivers; in addition, they provide more frequent signal transitions to help maintain synchronization of the clock recovery circuit. At present, the NRZ format is preferred for sensitivity; however, it may be used either with self-synchronizing scrambling or with block coding; both can be realized in IC form on the same chip.

12.2.3 pin-Photodiode Receiver Noise and Sensitivity Calculations

This section discusses noise and sensitivity calculations for lightwave receivers using (non-multiplying) pin photodiodes and FET input stages, and lays the noise-theory foundation for the receiver circuit designs of the following Sections.

The receiver-circuit noise expressions are for equivalent RMS input noise currents; the real noise sources in the amplifier and detector are replaced by a single equivalent noise current source at the input to a noise-free equivalent amplifier. This is convenient because the signal-to-noise ratio at the amplifier output is then just the photocurrent divided by the equivalent input noise current. In addition, these results are not affected by the feedback techniques the noise sources can be referred to the input before mentally closing the feedback loop; the loop is thus closed around the noiseless equivalent amplifier; by inspection, the equivalent input current noise source, which is outside the loop, is unchanged.

For digital receiver systems, the signal-to-noise ratio at the digital decision circuit input

is given by the photocurrent signal divided by the receiver-amplifier-equivalent input-noise expression, provided that the noise bandwidth is taken as the channel-filter bandwidth. As mentioned before the channel-filter bandwidth for a synchronous NRZ receiver system is typically 0.56 times the bit rate B . The photocurrent signal can be had from the optical signal power by remembering that a 1-eV photon has a wavelength of 1.240 μm ; as that wavelength, 1 watt of detected optical power gives 1 ampere of photocurrent. This then gives

$$\bar{I} = \eta \bar{P} \frac{\lambda}{1.240} \quad , \quad (12.3)$$

where \bar{I} the average photocurrent in amperes, η is the photodiode quantum efficiency, \bar{P} is the average optical power in watts, and λ is the optical wavelength in micrometers.

The problem now is to turn this signal-to-root-mean-square noise ratio at the decision circuit into a digital bit-error rate (BER). Consider a synchronous detection digital receiver system. The digital decision circuit samples the signal once at the center of each bit interval; if the signal, s , is below the decision threshold D , the bit is read as a zero; if s is greater than D it is read as a one. The probability of error, i.e., the BER, is the probability that the noise at the sampling instant will bring the zero signal above D or the one signal below D .

By the central limit theorem, the noise amplitude probability distribution is Gaussian if the noise amplitude is the sum of many small independent physical process. The probability distribution for the zero-level signal is then, following Smith and Personick[1]

$$P(s) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-[s-s(0)]^2/2\sigma_0^2} \quad , \quad (12.4)$$

where $s(0)$ is the noise-free, zero-level signal, σ_0 is the signal variance, or root-mean-square zero-level analog noise. The probability E_{01} of mistakenly identifying a zero bit as a one is then (Fig. 12.5)

$$E_{01} = P(s > D) = \frac{1}{\sqrt{2\pi}\sigma_0} \int_0^\infty ds e^{-[s-s(0)]^2/2\sigma_0^2} \quad .$$

The derivation of E_{10} , the probability of mistakenly identifying a one-bit as a zero, is similar. Changing variables, the general error probability is

$$P(E) = \frac{1}{\sqrt{2\pi}} \int_Q^\infty dx e^{-x^2/2} = \frac{1}{2} \text{erfc} \left(\frac{Q}{\sqrt{2}} \right) \quad , \quad (12.5)$$

where

$$Q = \frac{(D - s(j))}{\sigma_j} \quad (12.6)$$

and for $j = 0$, $P(E) = E_{01}$; for $j = 1$, $P(E) = E_{10}$.

For pin receivers, the zero-signal-level RMS noise, σ_0 , and the one-signal-level RMS noise, σ_1 , are essentially equal. Assuming a random bit stream (maximum information content), ones and zeros are equally frequent. The optimum decision level D is then midway between $s(0)$ and $s(1)$, and $E_{01} = E_{10}$. The bit-error rate (BER) is now

$$\text{BER} = \frac{1}{2} \text{erfc} \left(\frac{Q}{\sqrt{2}} \right) \quad , \quad (12.7)$$

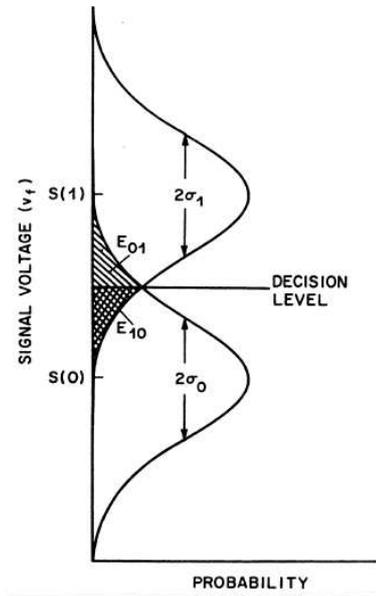


Figure 12.5: Error probabilities for two-level digital system. E_{01} is the probability of mistakenly identifying a zero bit as a one bit; E_{10} is the probability of mistakenly identifying a one bit as a zero bit.

where

$$Q = \frac{s(1) - s(0)}{2\langle s_n^2 \rangle^{1/2}} ,$$

where $\langle s_n^2 \rangle^{1/2} = \sigma$ is the RMS noise. Q is then just half the peak-to-peak signal to RMS noise ratio at the digital decider input. However, as mentioned, this SNR is just the photocurrent signal to equivalent spat RMS noise current ratio. Assuming that the zero-level photocurrent is zero, the average photo current fin half the peak (one-level) photocurrent I_{pk} and

$$Q = \frac{\bar{I}}{\langle i_n^2 \rangle^{1/2}} = \frac{I_{pk}}{2\langle i_n^2 \rangle^{1/2}} , \quad (12.8)$$

where $\langle i_n^2 \rangle^{1/2}$ is the RMS equivalent input noise current.

Equations (12.7) and (12.8) give the pin-receiver bit-error rate (BER) in terms of the photocurrent signal to RMS equivalent input-noise current ratio. The BER as a function of Q is shown in Fig. 12.6. A typical system requirement is for a BER of 10^{-9} ; this corresponds to $Q = 6$ or an average photocurrent of six times the RMS equivalent input-noise current. This corresponds to a peak-to-peak signal at the digital decider input that is 12 times the RMS noise. The optical power to get that photocurrent in then the digital optical receiver sensitivity.

A. Input Device Noise Theory for pin-FET Receivers

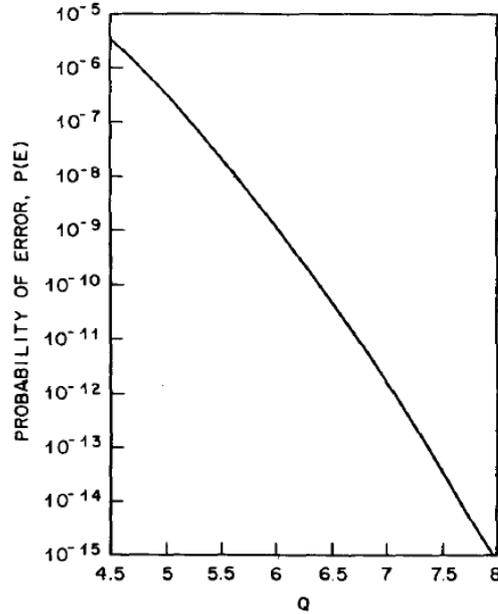


Figure 12.6: Probability of error versus the average signal-to-RMS-noise ratio, Q .

Figure 12.7 shows a generalized pin-FET receiver amplifier. This subsection presents a tutorial review for the noise due to the input devices shows in Fig. 12.7, i.e., the channel noise of the input FET, the Johnson noise of the bias/feedback or input resistor, and the shot noise due to the input FET gate and pin-photodiode leakage currents. This theory gives much of the information about FET- and pin-technology requirements and amplifier design constraints.

The input bias or feedback resistor R_i contributes a mean Johnson thermal input noise current squared per frequency bandwidth df of

$$d\langle i_n^2 \rangle_R = \frac{4kT}{R_i} df \quad . \quad (12.9)$$

The different frequency components of the noise are both independent and orthogonal. Therefore, the total mean-square input noise $\langle i_n^2 \rangle$ due to the input resistor is simply the mean-square spectral noise density of Eq. (12.9) integrated over the channel frequency bandwidth. If $|F(f)|$ is the magnitude of the normalized channel-filter frequency response, then

$$\langle i_n^2 \rangle_R = \int_0^\infty \frac{4kT}{R_i} |F(f)|^2 df \quad ; \quad (12.10)$$

the independent noise currents squared are multiplied by the filter response magnitude squared, then integrated.

By inspection, this resistor-noise integral is proportional to the bandwidth or to the bit rate B times a numerical factor that depends on the shape of the filter frequency response function $F(f)$. (The same filter function is scaled up or down in frequency for different bit

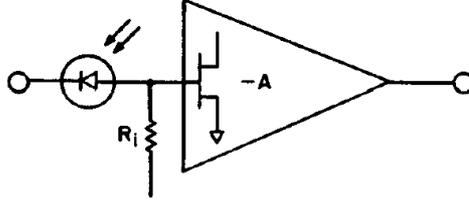


Figure 12.7: General pin-FET receiver amplifier.

rates.) Changing variables from the frequency f to $y = f/B$ (the frequency normalized to the bit rate) gives

$$\langle i_n^2 \rangle_R = \frac{4kT}{R_i} I_2 B \quad , \quad (12.11)$$

where

$$I_2 = \int_0^\infty dy |F^*(y)|^2 \quad (12.12)$$

where $F^*(y)$ is the filter frequency response shape; the numerical factor I_2 is called the second Personick integral.

The next question is how to normalize the channel frequency response $F(f)$. $F(f)$ must give the fraction of the input-noise-current frequency component that appears at the linear channel output to the digital decider. This is best done by first normalizing in the time domain so that a unit input photocurrent pulse maps to a unit output pulse to the decider. The corresponding filter frequency function normalization then applies to the noise as well.

The receiver input is a photocurrent; the input pulse shape $h_p(t)$ is normalized to correspond to a unit photocurrent over a bit interval $T = 1/B$:

$$\frac{1}{T} \int_{-\infty}^{\infty} h_p(t) dt = 1 \quad . \quad (12.13)$$

These receivers are used with synchronous decision circuits, which sample the waveform at the center of each bit interval when the pulse is a maximum (logic one bit) or a minimum (logic zero bit). Thus, a unit filtered pulse is one which is of unit amplitude at the sampling instant, t_0 ;

$$h_{\text{out}}(t_0) = 1 \quad . \quad (12.14)$$

Defining $H_p(f)$ and $H_{\text{out}}(f)$ as the fourier transforms of $h_p(t)$ and $h_{\text{out}}(t)$ then gives the normalized filter function:

$$F(f) = \frac{H_{\text{out}}(f)}{H_p(f)} \quad . \quad (12.15)$$

Note that, for a given filter design, changing the input pulse shape changes the magnitude but not the frequency dependence of $F(f)$; $F(f)$ is normalized such that a unit average photocurrent pulse of whatever pulse shape chosen gives a unit amplitude input at the sampling instant to the decision circuit, but the shape of $F(f)$ depends only on the channel filter design.

The remaining question is how to arrive at the frequency dependence of the channel filter function $F(f)$; in other words, how to design the channel filter. By inspection, a narrow channel-filter bandwidth means smaller Personick integrals and less receiver noise; however, a narrow filter bandwidth also means slow rise and fall times, which cause the filtered bit-pulse $h_{\text{out}}(t)$ to spread into the neighboring bit intervals. Thus, choosing the filter function is a tradeoff between noise considerations in the frequency domain and pulse-shape considerations in the time domain. Since the decision circuit samples the waveform at the center of each bit interval, the bit-pulse $h_{\text{out}}(t)$ should peak at the center of its own bit interval and should be zero at the centers of the neighboring bit-intervals to avoid inter-symbol interference (ISI). The filter typically is designed by iterating between the time and frequency domains, alternately minimizing the ISI and the noise bandwidth, respectively. The result usually is a filter bandwidth of 0.5-0.6 times the bit rate for a NRZ signal. For such a typical filter design, I_2 worked out to be 0.56.

The FET gate-leakage current and the pin-leakage current contribute an input shot noise current squared per unit frequency bandwidth of

$$d\langle i_n^2 \rangle_\ell = 2qI_\ell df \quad , \quad (12.16)$$

where the total leakage current I_ℓ is the sum of the pin-photodiode leakage current $I_{\ell\text{pin}}$ plus the FET leakage current $I_{\ell\text{FET}}$, and q is the electron charge. Integrating over frequency, as before, gives the total input-leakage-current noise in the channel filter bandwidth:

$$\langle i_n^2 \rangle_\ell = \int_0^\infty 2qI_\ell |F(f)|^2 df \quad (12.17)$$

or

$$\langle i_n^2 \rangle_\ell = 2qI_\ell I_2 B \quad , \quad (12.18)$$

where I_2 is the second Personick integral as before.

The input FET channel noise is essentially the Johnson noise of the unpinched-off portion of the channel next to the source. This channel conductance is simply the transconductance of the FET; the drain current noise per unit bandwidth is then

$$d\langle i_n^2 \rangle_{\text{drain}} = 4kTg_m\Gamma df \quad , \quad (12.19)$$

where Γ is a factor to account for high-field effects in short-channel transistors. For 1- μm gate length silicon MOSFETs, Γ is typically 1 to 1.2; for 1- μm GaAs FETs, Γ is typically 1.4 to 1.8.

This mean-square FET drain noise current must be turned into an equivalent input mean-square noise current. The drain noise current squared can be turned into an equivalent gate(input) noise voltage squared by dividing by the transconductance squared ($i_d^2 = g_m^2 e_g^2$):

$$d\langle e_n^2 \rangle_{\text{FET}} = \frac{4kT\Gamma}{g_m} df \quad . \quad (12.20)$$

The corresponding mean-square input noise current is simply the mean-square gate noise voltage divided by the input impedance squared (without feedback.) For small input bias/feedback resistor values, that mean-square equivalent input noise current would be just $\langle i_n^2 \rangle_{\text{FET}} = \langle e_n^2 \rangle_{\text{FET}} / R_i^2$. However, this case is of no practical interest because the

Johnson noise of that small-value resistor would swamp the FET noise and ruin the sensitivity. In high-sensitivity amplifiers, the input impedance (without feedback) is essentially the input FET gate capacitance, C_{FET} , in parallel with the junction capacitance of the pin photodiode, C_{pin} , plus any stray capacitance, C_s . The equivalent input noise current is then

$$d\langle i_n^2 \rangle_{\text{FET}} = \frac{4kT\Gamma(2\pi f C_T)^2}{g_m} df \quad , \quad (12.21)$$

where the total input capacitance, C_T is the sum of the FET input capacitance, C_{FET} , plus the photodiode capacitance C_{pin} , plus C_s .

The physical assumption behind Eq. (12.21) is that in the absence of feed-back, the photocurrent signal will be integrated by the input capacitance $C_T (V = 1/C_T \int i dt)$; that R_i is so large (for low noise) that it can be neglected over most of the bandwidth. For example, in a 45-Mb/s receiver with $R_i = 1 \text{ M}\Omega$, and $C_T = 1 \text{ pF}$, C_T dominates above 80 kHz. This means that the input voltage produced by the photocurrent is inversely proportional to the frequency; since the input noise voltage is fixed, the signal-to-noise ratio is also inversely proportional to the frequency. When the signal is equalized (differentiated), the low-frequency signals (and noise) are attenuated; the high frequencies are boosted. Thus, with the signal now proportional to the input photocurrent, the noise is proportional to frequency, and the mean-square noise is proportional to the frequency squared, as in Eq. (12.21). Again, using feedback to avoid signal integration does not change the signal-to-noise ratio versus frequency.

The total equivalent mean-square input noise current of the FET is obtained by integrating over the channel-filter frequency response as before:

$$\langle i_n^2 \rangle_{\text{FET}} = 4kT\Gamma \frac{(2\pi C_T)^2}{g_m} \int_0^\infty f^2 |F(f)|^2 df \quad , \quad (12.22)$$

or, changing variables from f to $y = f/B$ as before,

$$\langle i_n^2 \rangle_{\text{FET}} = 4kT\Gamma \frac{(2\pi C_T)^2}{g_m} I_3 B^3 \quad , \quad (12.23)$$

where

$$I_3 = \int_0^\infty |F(y)^*|^2 y^2 dy \quad (12.24)$$

is the third Personick integral.

The total mean-square equivalent input-noise current due to the input devices (pin photodiode plus FET pins bias resistor) is then

$$\langle i_n^2 \rangle_T = \langle i_n^2 \rangle_{R_i} + \langle i_n^2 \rangle_\ell + \langle i_n^2 \rangle_{\text{FET}} \quad (12.25)$$

or

$$\langle i_n^2 \rangle_T = \frac{4kT}{R_i} I_2 B + 2qI_\ell I_2 B + 4kT\Gamma \frac{(2\pi C_T)^2}{g_m} I_3 B^3 \quad , \quad (12.26)$$

where $I_\ell = I_{\ell\text{FET}} + I_{\ell\text{pin}}$ and $C_T = C_{\text{pin}} + C_{\text{FET}} + C_s$.

Equation (12.26) for the total equivalent input noise due to the input devices contains much of the information about pin-photodiode and FET technology requirements and about amplifier design constraints. In high-sensitivity designs, the input resistor Johnson

noise term $\langle i_n^2 \rangle_{R_i}$, is made almost negligible by making R_i large. The leakage current shot noise has been made negligible by reducing the pin photodiode and FET leakage currents. This leaves the input FET noise as the fundamental noise source; the receiver noise with low-leakage devices and state-of-the-art circuitry reduces to approximately.

$$\langle i_n^2 \rangle_T \cong 4kT\Gamma \frac{(2\pi C_T)^2}{g_m} I_3 B^3 \quad . \quad (12.27)$$

Thus, the circuit-equivalent input-noise power (or mean-square noise current) is approximately proportional to B^3 , C_T^2/g_m and the channel noise factor Γ . One can write an input circuit figure of merit that is independent of bit rate

$$M = \frac{g_m}{C_T^2 \Gamma} \quad . \quad (12.28)$$

The receiver optical sensitivity is inversely proportional to the root-mean-square noise current and therefore is approximately proportional to $\sqrt{M}B^{-3/2}$.

Note that if the mean-square noise current were proportional to B^2 rather than B^3 , the photocurrent charge per bit would be constant. In fact, the charge per bit goes up approximately as the square root of the bit rate.

B. Complete Receiver Circuit Noise Expressions with Device Figures of Merit

This subsection first derives figure-of-merit expressions for pin-photodiode, FET, and FET IC technologies, then extends the Smith and Personick noise expression to include the noise from the input FET load device and the rest of the receiver circuit. The resultant expression is used to calculate the theoretical receiver sensitivities.

The front-end figure of merit of Eq. (12.28) can be rewritten as the product of an FET technology figure-of-merit times a pin-photodiode figure-of-merit. In a typical high-performance FET technology, the source-to-drain spacing or channel length is fixed at the minimum reliable resolution of the lithography. The FET transconductance and input capacitance are both proportional to the channel width; the g_m/C_{FET} ratio is set by the FET technology. Rewriting g_m as $(g_m/C)C_{\text{FET}}$, where C_{FET} determines the FET size, and taking $C_T = C_{\text{FET}} + C_{\text{pin}} + C_s$ gives

$$M = \left(\frac{g_m}{\Gamma C_{\text{FET}}} \right) \left(\frac{C_{\text{FET}}}{(C_{\text{FET}} + C_{\text{pin}} + C_s)^2} \right) \quad . \quad (12.29)$$

Differentiating the second bracketed term with respect to C_{FET} says that the optimum-size FET in a given technology has a gate width such that

$$C_{\text{FET}} = C_{\text{pin}} + C_s = \frac{1}{2}C_T \quad . \quad (12.30)$$

Assuming such an optimum-sized FET, one can now write

$$M = M_{\text{FET}} M_{\text{pin}} \quad , \quad (12.31)$$

where the FET-technology figure of merit is

$$M_{\text{FET}} = \frac{g_m}{\Gamma C_{\text{FET}}} = \frac{2\pi f_T}{\Gamma} \quad , \quad (12.32)$$

where $f_T = g_m/2\pi C_{\text{FET}}$ is the unity gain frequency of the FET technology. M_{pin} , the pin-photodiode (and stray capacitance) figure of merit is

$$M_{\text{pin}} = \frac{1}{4(C_{\text{pin}} + C_s)} \quad (12.33)$$

The optical receiver sensitivity is, again, proportional to the square root of the product of these figures of merit.

The FET figure of merit indicates that high-sensitivity receivers should be made with microwave or VHSIC technologies because such FETs have the highest f_T s. The optical sensitivity of such receivers is approximately proportional to the square root of f_T ; thus, these high-frequency FETs are preferred, even at low bit rates (e.g., 10 Mb/s), but for noise, not frequency-response, reasons.

The photodiode and stray-capacitance figure of merit of Eq. (12.33) indicates that the pin-photodiode capacitance C_{pin} , and the stray capacitance C_s must be made as small as possible. Assuming an optimum-size amplifier input FET (gate width such that $C_{\text{FET}} = C_{\text{pin}} + C_s$), the optical receiver sensitivity is inversely proportional to the square root of $(C_{\text{pin}} + C_s)$. For small C_s , the optical sensitivity is inversely proportional to the square root of the pin-photodiode capacitance.

Thus, the pin-photodiode technology objectives are low capacitance, low leakage current, and high quantum efficiency. Low capacitance means either a low doping ($10^{14} - 10^{15}/\text{cc}$) for a wide depletion region, a small diameter (area), or both.

Thus, a high-sensitivity pin-FET receiver is a combined FET technology problem (high g_m/CT), pin-photodiode problem (low C_{pin} , low leakage current) circuit design problem (large R_i for low Johnson noise while preserving a wide bandwidth and dynamic range), and packaging problem (low stray capacitance C_s , small photodiode diameter if economic).

The pin-photodiode figure of merit can also be read as a total-input-capacitance figure of merit. For an optimized receiver, $M_{\text{pin}} = 1/(2C_T)$ by Eq. (12.30); thus, for a given FET figure of merit, the sensitivity of an optimized receiver is inversely proportional to the square root of C_T .

Actual sensitivity calculations should also include corrections for the noise due to the input FET drain load device and for the noise of following stages. The Smith and Personick noise calculations are readily expendable to include these effects.

The input FET's load device Q_1 is typically another FET. The IC load transistor Q_L adds an extra mean-square noise current at the drain of Q_1 of

$$d\langle i_n^2 \rangle_L = 4kT g_{\text{mL}} \Gamma df \quad . \quad (12.34)$$

where g_{mL} is the transconductance of Q_L . Equation (12.19) can then be rewritten as

$$d\langle i_n^2 \rangle_{\text{drain}} = 4kT(g_{\text{m1}} + g_{\text{mL}})\Gamma df \quad , \quad (12.35)$$

where g_{m1} is the transconductance of input FET Q_1 . Retracing the derivation of Eqs. (12.20)-(12.23) then gives a total mean-square equivalent input noise current due to Q_1 (input FET) and Q_L (load FET) of

$$\langle i_n^2 \rangle_{Q_1+Q_L} = 4kT\Gamma(2\pi C_T)^2 \left(\frac{g_{\text{m1}} + g_{\text{mL}}}{g_{\text{m1}}^2} \right) I_3 B^3 \quad . \quad (12.36)$$

The optimum ratio of Q_L size (gate width) to Q_1 size is a tradeoff. A large Q_L gives a higher drain current density in Q_1 and therefore a higher g_{m1} ; a small Q_L gives less drain current noise due to Q_L . Assuming an optimum Q_L to Q_1 ratio for the particular technology, one can revise Eq. (12.32) to give a figure of merit for the FET IC technology:

$$M_{\text{IC}} = \frac{g_{m1}^2}{(g_{m1} + g_{mL})\Gamma C_{\text{FET}}} \quad . \quad (12.37)$$

M_{IC} depends only on the IC technology, so long as the input FET Q_1 is scaled so that $C_{\text{FET}} = C_T/2$ and Q_L is scaled for minimum noise. The receiver figure of merit now is $M_{\text{IC}} \cdot M_{\text{pin}}$, where M_{pin} is given by Eq. (12.33) as before; the overall receiver sensitivity is approximately proportional to the square root of the receiver figure of merit. As will be shown below, the Q_L noise term typically is more important than the all following stage noise terms combined; thus, Equation (12.37) is a good IC technology figure of merit and gives good approximate sensitivities.

The noise of the following stages of the receiver can be represented as an equivalent stage input mean-square noise voltage of $d\langle e_s^2 \rangle$ per unit bandwidth df . If a_{s-1} is the total voltage gain of the $s - 1$ stages preceding stage s , the mean-square equivalent input noise voltage due to stage s is

$$d\langle e_s^2 \rangle_{\text{I}} = d\langle e_s^2 \rangle / a_{s-1}^2 \quad . \quad (12.38)$$

Assuming $\langle e_s^2 \rangle$ is constant in frequency, one can repeat the derivation of Eqs. (12.20)-(12.23), substituting $d\langle e_s^2 \rangle_{\text{I}}$ for $d\langle e_n^2 \rangle_{\text{FET}}$. This gives the mean-square equivalent input noise current due to stage s :

$$\langle i_n^2 \rangle_s = \frac{d\langle e_n^2 \rangle_s}{a_{s-1}^2} (2\pi C_T)^2 I_3 B^3 \quad . \quad (12.39)$$

The total mean-square equivalent input noise current of the receiver is then

$$\begin{aligned} \langle i_n^2 \rangle &= \frac{4kT}{R_1} I_2 B + 2qI_\ell I_2 B + 4kT\Gamma \frac{(2\pi C_T)^2}{g_{m1}} I_3 B^3 \\ &+ 4kT\Gamma (2\pi C_T)^2 \frac{g_{mL}}{g_{m1}^2} I_3 B^3 + (2\pi C_T)^2 I_3 B^3 \sum_{s=2}^N \frac{d\langle e_n^2 \rangle_s}{a_{s-1}^2} \quad , \quad (12.40) \end{aligned}$$

where the first two terms are the input-resistor Johnson noise and the leakage-current shot noise, respectively, the third term is the input FET noise, the fourth term is the noise of the input FET load (Q_L), and the last term is the sum over the noise contributions of the following stages. Generally, the following-stage noise current is less important than the load device contribution; even the second-stage mean-square noise is divided by the first-stage voltage gain squared. Thus, Eq. (12.37) is a good figure of merit for comparing FET IC technologies.

C. Sensitivities of Present-Technology pin-FET Receivers

This subsection calculates theoretical sensitivities for present-technology silicon and GaAs FET IC receivers and compares the two technologies. It also includes sensitivity results from the literature.

For the numerical sensitivity calculations, g_m/C is taken as 70 mS/pF for 1- μ m silicon MOSFETs and 90 mS/pF for 1- μ m GaAs MESFETs; Γ is taken as 1.2 for the silicon and 1.5 for the GaAs. Since the FET technology figure of merit is $g_m/(CT)$ by Eq. (12.32), the silicon and GaAs sensitivities are essentially equal in theory; GaAs FETs have higher transconductances, but silicon FETs have lower channel noise. In practice, GaAs designs are *presently* a few decibels more sensitive.

The Personick integrals I_2 and I_3 can be roughly estimated by remembering that the channel-filter bandwidth is typically 0.56 times the bit rate B for synchronous detection NRZ receivers. Thus, taking $|F(y)| = 1$ for $y < 0.56$ and $F(y) = 0$ for $y > 0.56$ in Eq. (12.12) for I_2 gives $I_2 = 0.56$; using this approximation in Eq. (12.24) for I_3 gives $I_3 = 0.059$. Values for I_2 and I_3 for different input and output (filtered) pulse shapes are found in [1] and [2].

Figure 12.8 shows theoretical digital optical-receiver sensitivities versus bit rate for 1- μ m gate length GaAs and silicon IC receivers using InGaAs pin photodiodes. The optical wavelength is 1.3 μ m. The calculations assume the new high-sensitivity, micro-FET feedback IC receiver designs, in which the input/feedback resistor noise is almost negligible; the sensitivities were calculated on paper designs using Eq. (12.40) for the total receiver noise. Figure 12.8 assumes a 1-pF total front-end capacitance, e.g., $C_{\text{pin}} = 0.40$ pF, $C_s = 0.10$ pF, $C_{\text{FET}} = 0.5$ pF; the silicon input FET then has a transconductance of 35 mS; the GaAs input FET has a transconductance of 45 mS. The photodiode leakage current is taken as 1 nA at 20°C and 15 nA at a maximum operating temperature of 85°C (the 15x increase assumes that the leakage is a G-R current via midgap states.) The bit-error rate is 10^{-9} .

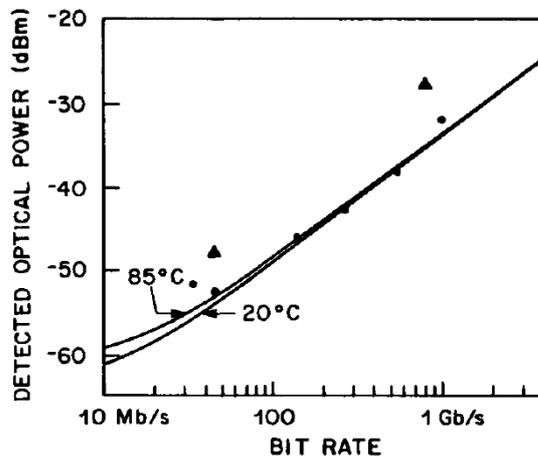


Figure 12.8: Theoretical sensitivities of present-technology pin-FET lightwave receivers. Dots show GaAs FET receiver measurements from the literature; triangles show silicon FET receiver measurements.

Since the theoretical sensitivities of the silicon and GaAs FET IC receivers are essentially the same, only one curve is plotted for both. Note also that the calculated

sensitivities go approximately as $B^{-3/2}$; the first two terms in Eq. (12.40) for the input noise squared are negligible except at low bit rates; the others go as B^3 except for the (small) following-stage noise summation, which increases slightly faster than B^3 because the gain per stage is less for wider bandwidth stages. Finally, note that the calculated sensitivities are competitive with those calculated for present APD receivers to 100-200 Mb/s.

Figure 12.8 also shows the best experimentally achieved pin-FET receiver sensitivities.

Unfortunately, the silicon MOSFET optical receivers are *presently* a few decibels less sensitive than either theory or RMS noise measurements would indicate. This may be because fine-line silicon FETs were developed as a digital IC technology in which noise was not important. The fine-line GaAs MESFET technology was developed as an analog technology for microwave receivers. Therefore, the GaAs noise problems were important and were solved.

12.2.4 Avalanche Photodiode Receiver Noise and Sensitivity Calculations

In an avalanche photodiode (APD), the primary photocurrent is multiplied by impact ionization in the p-n junction, which is operated reversebiased near breakdown. The multiplied photocurrent signal goes to the receiver circuit. Consider an APD in which the primary photocurrent is multiplied (on average) by a factor $\langle M \rangle$. If the multiplication process were noiseless and the APD leakage current negligible, the optical signal power required by the receiver would be decreased by a factor $\langle M \rangle$ and the optical receiver sensitivity increased by a factor $\langle M \rangle$.

In fact, the multiplication process is noisy because it is the result of random impact ionizations. The equivalent mean-square primary photocurrent noise $d\langle i_i^2 \rangle_D$ per bandwidth df due to the APD is the shot noise of the photocurrent I_{ph} plus leakage current I_ℓ times the McIntyre excess noise factor $F(\langle M \rangle)$ [3]

$$d\langle i_n^2 \rangle_D = 2q(I_{ph} + I_\ell)F(\langle M \rangle)df \quad . \quad (12.41)$$

For noiseless multiplication, $F(\langle M \rangle) = 1$, and the noise is just the photocurrent plus leakage-current shot noise; $F(\langle M \rangle)$ is the factor by which the real avalanche multiplication increases the noise over that of a noiseless multiplication.

If the avalanche is initiated by injection of photocarriers at one side of the avalanche region, $F(\langle M \rangle)$ is given by[3]

$$F(\langle M \rangle) = \langle M \rangle \left[1 - (1 - k) \left(\frac{\langle M \rangle - 1}{\langle M \rangle} \right)^2 \right] \quad , \quad (12.42)$$

where k is the ratio of the electron and hole ionization coefficients. In silicon, electrons have the higher ionization coefficient; the 0.8- μm wavelength silicon APDs use photoelectron initiated multiplication, and k is the ratio of the hole ionization coefficient to the electron ionization coefficient. In InGaAs/InP 1.3- to 1.6- μm wavelength APDs, the avalanche is photo-hole initiated and k is the ratio of the electron ionization coefficient to that of holes. In silicon APDs, k is typically 0.02[4]; in InGaAs/InP APDs, k is typically 0.4 at present.

The APD is not the “solid state equivalent of a photomultiplier” because both electrons and holes can impact ionize in an APD; in a photomultiplier, only electrons impact ionize

(on the dynodes); there are no holes in a vacuum. In an electron-initiated APD, the electrons of the primary avalanche travel downstream, creating electron-hole pairs by impact ionization; the resultant holes travel upstream and can create more electrons, thus initiating secondary electron avalanches, etc. This hole feedback makes avalanche carrier multiplication much noisier than photomultiplier electron multiplication. When the electron-hole feedback loop-gain becomes unity, the avalanche is self-sustaining and the diode breaks down. The higher the multiplication, the closer to breakdown and the noisier the multiplication.

For $k = 0$ (best case), only one carrier ionizes and the APD multiplication process is similar to that of a photomultiplier. In this limit, $F(\langle M \rangle)$ becomes

$$F(\langle M \rangle) = 2 - \frac{1}{\langle M \rangle} \quad . \quad (12.43)$$

For $k = 1$ (worst case), $F(\langle M \rangle)$ becomes

$$F(\langle M \rangle) = \langle M \rangle \quad ; \quad (12.44)$$

the avalanche multiplication is then more noisy due to the carrier feedback. In general, the lower the k , the less noisy the multiplication. Thus, the silicon APDs ($k \sim 0.02$) are much less noisy for a given multiplication than the InGaAs APDs ($k \sim 0.5$); unfortunately, silicon is transparent below $1.1 \mu m$ and cannot be used for 1.3 - to $1.6\text{-}\mu m$ APDs.

The total equivalent mean-square primary photocurrent noise $\langle i_n^2 \rangle_{\text{ph}}$ is the APD noise integrated over the receiver bandwidth plus the equivalent mean-square input noise current $\langle i_n^2 \rangle_{\text{T}}$ of the receiver amplifier, divided by $\langle M \rangle^2$:

$$\langle i_n^2 \rangle_{\text{ph}} = 2q(i_{\text{ph}} + I_{\ell})F(\langle M \rangle)I_1B + \frac{\langle i_n^2 \rangle_{\text{T}}}{\langle M \rangle^2} \quad , \quad (12.45)$$

where the integral over frequency is I_1B ; I_1 is the first Personick integral.

When a logic-zero bit is transmitted, the optical signal, hence the photocurrent i_{ph} , is ideally zero; the zero-level mean-square noise then is due only to the APD leakage current and to the amplifier noise. When a logic one bit is transmitted, the mean-square noise is increased by $2qI_{\text{ph}}F(\langle M \rangle)I_1B$, which is the noise due to the avalanche multiplication of the one-level photocurrent. Since an APD receiver's one-level noise is greater than its zero-level noise, the decision threshold D is typically set closer to the zero level than to the one level.

In practice, the optical transmitter is not completely turned off during zero bits, for reasons of transmitter response speed and (for laser transmitters) for reasons of optical frequency stability. Take the zero-bit optical signal level $P(0)$ as a fraction r of the one-bit optical signal level $P(1)$:

$$r = \frac{P(0)}{P(1)} = \frac{\langle i_{\text{ph}}(0) \rangle}{\langle i_{\text{ph}}(1) \rangle} \quad , \quad (12.46)$$

where $\langle i_{\text{ph}}(0) \rangle$ is the expected photocurrent at the sampling instant for a zero bit; $\langle i_{\text{ph}}(1) \rangle$ is the photocurrent for one bit. r is called the transmitter optical extinction ratio.

Ideally, r should be zero; in practice, r may be as high as 0.2. This means a smaller photocurrent signal component for a given average optical power. In addition,

the zero-level photocurrent is the functional equivalent of a leakage current and adds a corresponding noise term to both the zero and one signals.

Taking the avalanche-noise amplitude distribution as approximately Gaussian, the error probability or bit-error rate (BER) is given by Eq. (12.5): $\text{BER} = \frac{1}{2}\text{erfc}(Q/\sqrt{2})$ where, for a zero bit,

$$Q_0 = (i_{\text{th}} - \langle i_{\text{ph}}(0) \rangle) / \langle i_{n0}^2 \rangle_{\text{ph}}^{1/2} \quad , \quad (12.47)$$

and i_{th} is the decision threshold and $\langle i_{n0}^2 \rangle^{1/2}$ the root-mean-square noise for a zero bit. Similarly, for a one bit,

$$Q_1 = \frac{(\langle i_{\text{ph}}(1) \rangle - i_{\text{th}})}{\langle i_{n1}^2 \rangle_{\text{ph}}^{1/2}} \quad . \quad (12.48)$$

$Q_0 = Q_1 = 6$ gives a bit-error rate of 10^{-9} .

Taking $Q_0 = Q_1$ equal to the Q required for the desired, BER, and using the APD receiver noise Eq. (12.45) to give $\langle i_{n0}^2 \rangle^{1/2}$ and $\langle i_{n1}^2 \rangle^{1/2}$ in Eqs. (12.47) and (12.48), yields the photocurrent signal required for the desired BER. Assuming that zero bits and one bits are equally probable, and converting the photocurrent to an input optical power, gives an equation for the average optical power required for a given BER:

$$\begin{aligned} \eta \bar{P} = & \left(\frac{hc}{\lambda q} \right) \left(\frac{1+r}{1-r} \right) \left[(1+r) \frac{Q^2 q B I_1 F(\langle M \rangle)}{1-r} + \left(\left(\frac{Q^2 q B I_1 F(\langle M \rangle)}{1-r} \right)^2 \cdot 4r \right. \right. \\ & \left. \left. + Q^2 \left(\frac{\langle i_n^2 \rangle_T}{\langle M \rangle^2} + 2q I_{\ell m} F(\langle M \rangle) I_2 B \right) \right)^{1/2} \right] \quad , \quad (12.49) \end{aligned}$$

where Q is the photocurrent signal-to-average-noise ratio for the given BER, multiplied by a prefactor that is unity for zero (ideal) extinction ratio r :

$$Q = \frac{\langle i_{\text{ph}}(1) \rangle - \langle i_{\text{ph}}(0) \rangle}{\langle i_{n0}^2 \rangle_{\text{ph}}^{1/2} + \langle i_{n1}^2 \rangle_{\text{ph}}^{1/2}} = \left(\frac{1-r}{1+r} \right) \frac{\bar{I}_{\text{ph}}}{\frac{1}{2}(\langle i_{n0}^2 \rangle^{1/2} + \langle i_{n1}^2 \rangle^{1/2})} \quad , \quad (12.50)$$

$\langle i_n^2 \rangle_T$ is the equivalent mean-square amplifier noise, $I_{\ell m}$ is the leakage current of the APD that undergoes multiplication, and $F(\langle M \rangle)$ is the McIntyre excess noise factor of Eq. (12.42).

Note that a high multiplication $\langle M \rangle$ reduces the effect of the receiver amplifier noise but gives more avalanche multiplication noise and a higher $F(\langle M \rangle)$ in Eq. (12.45). A low $\langle M \rangle$ gives lower avalanche multiplication noise but increases the effect of amplifier noise. The optimum $\langle M \rangle$ is determined by this trade-off.

In theory, Eq. (12.49) can be differentiated to find the optimum gain; the result is an unmanageable expression of no particular physical interest. In practice, one is much better off using a minimum finder program to find the optimum gain $\langle M \rangle$ and the minimum $\eta \bar{P}$ numerically.

Figure 12.9 shows theoretical InGaAs/InP APD receiver sensitivities versus bit rate. The figure assumes that the APD is used with an optimized 1- μm gate technology GaAs FET receiver amplifier, which is the most sensitive amplifier type at present. It also assumes a total receiver input capacitance of 1 pF (this includes the APD capacitance), a primary leakage current at 20°C of 3 nA, and an APD k -ratio of 0.4. Figure 12.9 shows

the pin-FET receiver sensitivities from Fig. 12.8 for comparison, plus the 20°C and 85°C InGaAs/InP APD sensitivities. The 85°C leakage current is taken as 45 nA. This assumes that the APD leakage current is a generation-recombination current via midgap states in the InGaAs, which gives a 15x increase in leakage current from 20°C to 85°C.

Figure 12.9 also shows InGaAs/InP APD-FET receiver measurements from the literature.

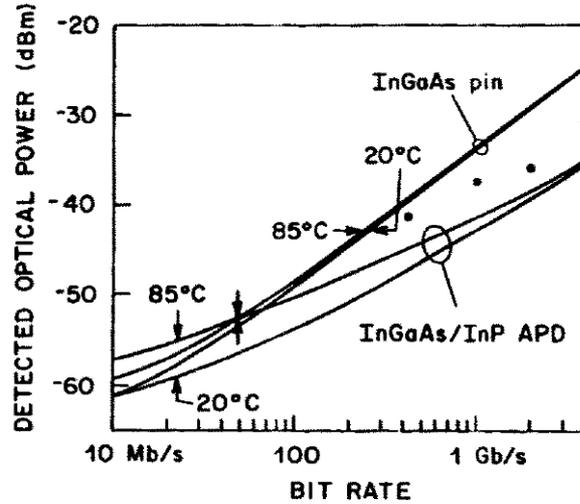


Figure 12.9: Theoretical InGaAs/InP APD receiver sensitivities. $\lambda = 1.3 \mu\text{m}$, $I_L = 3 \text{ nA}$ at 20°C, 45 nA at 85°C, $C_T = 1 \text{ pF}$, high-sensitivity receiver ICs. Dots show APD receiver measurements from the literature.

Note however, that these measurements were made at room temperature; however, in field use, the maximum operating temperature is typically 85°C. Unless the receiver and APD temperature is controlled, the 85°C sensitivity is what matters in practice.

Figure 12.9 shows that present InGaAs/InP 1.3- to 1.6- μm APDs in theory offer little sensitivity advantage below 100 Mb/s at a maximum operating temperature of 85°C. This is due to the noise caused by the avalanche multiplication of the primary leakage current; an APD without leakage current would offer a sensitivity improvement at all bit rates. In addition, reducing the leakage current is presently more important than improving the k -ratio for bit rates less than $\sim 500 \text{ Mb/s}$ at 85°C.

12.2.5 Optical Preamplicifier Receiver Noise

Consider an optical receiver system as shown in Fig. 12.10, in which an intensity modulated optical signal is incident on a traveling-wave (TW) optical linear preamplifier followed by an optical bandpass filter and photodiode. Electronic receiver circuitry consists of a baseband amplifier, an equalizer, and a decision circuit. The optical digital pulse signal is linearly amplified in the preamplifier, which generates a spurious spontaneous emission noise. The noise is partly reduced by the narrow-band optical bandpass filter.

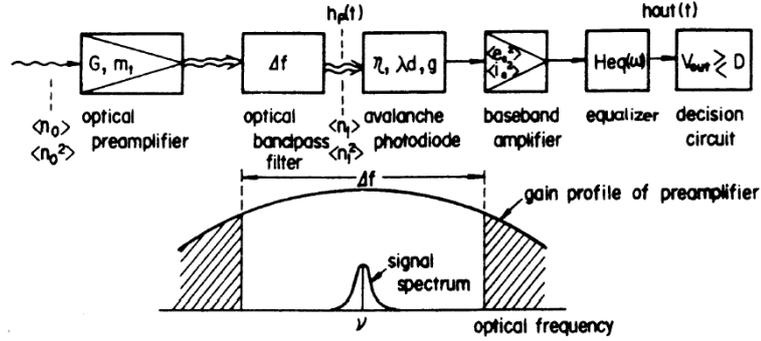


Figure 12.10: Optical receiver system model using a TW-type optical preamplifier in front of a photodetector. Here, n_0 and n_1 are numbers of preamplifier input and output photons, G is the preamplifier gain, m_t is the preamplifier transverse mode number, Δf is the optical filter bandwidth, η is the detector quantum efficiency, g is the detector internal gain, λd is the detector dark current count, $\sqrt{\langle i_a^2 \rangle}$ is the electronic amplifier input noise current, $\sqrt{\langle e_a^2 \rangle}$ is the input noise voltage, $H_{eq}(\omega)$ is the equalizer frequency response, D is the decision threshold voltage, $h_p(t)$ is the optical input pulse shape, and $h_{out}(t)$ is the equalized output pulse shape, respectively.

Let the mean and mean square of the photon number per second incident on the preamplifier be $\langle n_0 \rangle$ and $\langle n_0^2 \rangle$. The signal is assumed to be a single transverse mode. The variances in the incident photons for a completely coherent signal with Poisson distribution and a completely incoherent signal with Bose-Einstein distribution are of the form

$$\langle n_0^2 \rangle - \langle n_0 \rangle^2 = \begin{cases} \langle n_0 \rangle : & \text{coherent} \\ \langle n_0 \rangle^2 + \langle n_0 \rangle : & \text{incoherent} \end{cases} \quad (12.51)$$

The mean and variance in the photon number as the pre-amplifier output are described by the photon master equation for an unit noise bandwidth ($m_t \cdot \Delta f \cdot \tau = 1$) [5]

$$\frac{d\langle n \rangle}{dt} = (A - B - C)\langle n \rangle + A \quad (12.52)$$

$$\frac{d\langle n^2 \rangle}{dt} = 2(A - B - C)\langle n^2 \rangle + (3A + B + C)\langle n \rangle + A \quad (12.53)$$

where m_t is a transverse mode number, Δf is an optical bandwidth given, for example, by the optical filter, τ is a sampling interval over which the number of photons are counted, A and B are coefficients representing the stimulated emission and absorption, respectively, and C is a coefficient representing other loss mechanisms, such as free carrier absorption and waveguide scattering. The mean and variance of the photon number per second at the preamplifier output are

$$\langle n_1 \rangle = G\langle n_0 \rangle + (G - 1)\gamma m_t \Delta f \quad (12.54)$$

$$\begin{aligned}\langle n_1^2 \rangle &= G\langle n_0 \rangle + (G-1)\gamma m_t \Delta f + 2G(G-1)\gamma \langle n_0 \rangle \\ &\quad + (G-1)^2 \gamma^2 m_t \Delta f + G^2(\langle n_0^2 \rangle - \langle n_0 \rangle^2 - \langle n_0 \rangle)\end{aligned}\quad (12.55)$$

where G is a signal gain given by $\exp[(A-B-C)(\mathcal{L}/c_0)]$, \mathcal{L} is amplifier length, $\gamma = A/(A-B)$, and c_0 is the light velocity in the amplifier medium. Five terms on the right-hand side of Eq. (12.56) represent amplified signal shot noise, spontaneous emission shot noise, beat noise between signal and spontaneous emission, beat noise between spontaneous emission components, and signal excess noise, respectively. Factor $m_t \Delta f$ in the second terms of Eqs. (12.54) and (12.55) represents optical noise bandwidth. The beat noise between spontaneous emission components corresponds to an excess photon noise for lasers. The last term disappears when the input signal is completely coherent.

The optical power incident on the avalanche photodiode is assumed to be of the form for binary pulse signals:

$$\begin{aligned}p(t) &= h\nu[G\langle n_0 \rangle + (G-1)\gamma m_t \Delta f] \\ &= \sum_{k=-\infty}^{\infty} Gb_k h_p(t-kT) + h\nu(G-1)\gamma m_t \Delta f\end{aligned}\quad (12.56)$$

where $h_p(t-kT)$ is the amplified signal pulse shape normalized as $\int_{-\infty}^{\infty} h_p(t-kT)dt = 1$, T is a pulse spacing or the inverse of the data rate, $h\nu$ is the energy of a photon, and b_k is the energy of the incident optical pulse which takes one of the two values b_{\max} or b_{\min} . The average avalanche photodiode output current $\langle i_s(t) \rangle$ is given by

$$\langle i_s(t) \rangle = \left[\frac{e\eta}{h\nu} \sum_{k=-\infty}^{\infty} Gb_k h_p(t-kT) + e\eta(G-1)m_t \Delta f + e\lambda_d \right] \langle g \rangle \quad (12.57)$$

where

$$\begin{aligned}\langle g \rangle &= \text{average avalanche gain} \\ e &= \text{electron charge} \\ \lambda_d &= \text{detector dark current count} \\ \eta &= \text{quantum efficiency of the avalanche photodiode.}\end{aligned}$$

The average pulse voltage at the equalizer output is given by

$$\begin{aligned}\langle v_{\text{out}}(t) \rangle &= \frac{e\eta}{h\nu} \langle g \rangle R_L \left[\sum_{k=-\infty}^{\infty} Gb_k h_p(t-kT) \right] * h_{am}(t) * h_{eq}(t) \\ &= \frac{Ae\eta}{h\nu} \langle g \rangle R_L \left[\sum_{k=-\infty}^{\infty} \frac{Gb_k}{T} h_{\text{out}}(t-kT) \right]\end{aligned}\quad (12.58)$$

where A is an arbitrary constant, R_L is a detector load resistance, the symbol $*$ indicates convolution operation, $h_{am}(t)$ is the current impulse response of the baseband amplifier input circuit, $h_{eq}(t)$ is the current impulse response of the equalizer, and $h_{\text{out}}(t)$ is the equalized output pulse shape, respectively. The equalizer response $h_{eq}(t)$ is designed so

that the equalized output pulse signal has no intersymbol interference at sampling time $t = \{kT\}$:

$$h_{\text{out}}(0) = 1 \quad (12.59)$$

$$h_{\text{out}}(kT) = 0 \quad (k \neq 0) \quad . \quad (12.60)$$

Having defined the equalized output signal voltage in Eq. (12.58) and the input photon statistics in Eqs. (12.54) and (12.55), it is now possible to calculate the worst case variance of $\nu_{\text{out}}(t)$, the noise voltage at the equalizer output. The worst case variance at sampling time $t = 0$ is given by[2]

$$NW(b_0) = \max_{\{b_k\}, k \neq 0} [\langle \nu_{\text{out}}(0)^2 \rangle - \langle \nu_{\text{out}}(0) \rangle^2] \quad (12.61)$$

where the maximization is over all possible sets $\{b_k\}$ for $k \neq 0$, and b_0 takes either b_{max} or b_{min} . The worst case variance occurs when all the b_k , except b_0 , are b_{max} . Combining Eqs. (12.54), (12.55), (12.58) and (12.61), and after some algebraic calculation, we obtain

$$NW(b_0) = \langle N_s^2 \rangle + \langle N_{sp}^2 \rangle + \langle N_{s-sp}^2 \rangle + \langle N_{sp-sp}^2 \rangle + \langle N_d^2 \rangle + \langle N_a^2 \rangle \quad (12.62)$$

where

$$\langle N_s^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g^2 \rangle}{T^2} \frac{G\eta}{h\nu} \{I_1 b_0 + (I_0 - I_1) b_{\text{max}}\} : \quad (12.63)$$

signal shot noise

$$\langle N_{sp}^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g^2 \rangle}{T} I_2 \eta (G - 1) \gamma m_t \Delta f : \quad (12.64)$$

spontaneous emission shot noise

$$\langle N_{s-sp}^2 \rangle = 2A^2 e^2 R_L^2 \frac{\langle g \rangle^2}{T^2} \frac{G(G-1)\gamma\eta^2}{h\nu} \cdot \{I_1 b_0 + (I_0 - I_1) b_{\text{max}}\} : \quad (12.65)$$

beat noise between signal and spontaneous emission

$$\langle N_{sp-sp}^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g \rangle^2}{T} I_2 \eta^2 (G - 1)^2 \gamma^2 m_t \Delta f : \quad (12.66)$$

beat noise between spontaneous emission components

$$\langle N_d^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g^2 \rangle}{T} I_2 \lambda_d : \quad (12.67)$$

shot noise caused by detector dark current

$$\langle N_a^2 \rangle = \frac{I_2}{T} [\langle i_a^2 \rangle R_L^2 + \langle e_a^2 \rangle + 2K\theta R_L] + \frac{2\pi C_d}{T^3} R_L^2 I_3 \langle e_a^2 \rangle : \quad (12.68)$$

thermal noise generated in baseband electronic amplifier circuit.

Here, $\langle g^2 \rangle$ is the mean square of avalanche gain, which is conventionally expressed by $\langle g \rangle^{2+x}$, $\sqrt{\langle i_a^2 \rangle}$ is the baseband amplifier input noise current, $\sqrt{\langle e_a^2 \rangle}$ is the input noise voltage, $K\theta$ is Boltzmann's constant times absolute temperature, and C_d is the total parallel

capacitance in the detector circuit, respectively. The input signal excess noise expressed by the last term on the right-hand side of Eq. (12.55) is neglected here. The zeroth to third Personick integrals are given by[1],[2]

$$I_0 = \sum_{k=-\infty}^{\infty} H_p \left(\frac{2\pi k}{T} \right) \left\{ \frac{H_{\text{out}} \left(\frac{2\pi k}{T} \right)}{H_p \left(\frac{2\pi k}{T} \right)} * \frac{H_{\text{out}} \left(\frac{2\pi k}{T} \right)}{H_p \left(\frac{2\pi k}{T} \right)} \right\} \quad (12.69)$$

$$I_1 = \int_{-\infty}^{\infty} H_p \left(\frac{2\pi f}{T} \right) \left\{ \frac{H_{\text{out}} \left(\frac{2\pi f}{T} \right)}{H_p \left(\frac{2\pi f}{T} \right)} * \frac{H_{\text{out}} \left(\frac{2\pi f}{T} \right)}{H_p \left(\frac{2\pi f}{T} \right)} \right\} df \quad (12.70)$$

$$I_2 = \int_{-\infty}^{\infty} \left| \frac{H_{\text{out}} \left(\frac{2\pi f}{T} \right)}{H_p \left(\frac{2\pi f}{T} \right)} \right|^2 df \quad (12.71)$$

$$I_3 = \int_{-\infty}^{\infty} \left| \frac{H_{\text{out}} \left(\frac{2\pi f}{T} \right)}{H_p \left(\frac{2\pi f}{T} \right)} \right|^2 f^2 df \quad (12.72)$$

Error rate is the probability that the noise voltage superimposed on signal voltage crosses a prescribed threshold voltage D at sampling time $t = \{kT\}$. By approximating the statistics of $v_{\text{out}}(t)$ as Gaussian, the error rate is given by

$$P_e = \frac{1}{2} \left\{ \frac{1}{\sqrt{2\pi NW(b_{\text{min}})}} \int_D^{\infty} \cdot \exp \left[-\frac{\left(v - \frac{Ae\eta R_L \langle g \rangle G b_{\text{min}}}{h\nu T} \right)^2}{2NW(b_{\text{min}})} \right] dv \right. \\ \left. + \frac{1}{\sqrt{2\pi NW(b_{\text{max}})}} \int_{-\infty}^D \cdot \exp \left[-\frac{\left(v - \frac{Ae\eta R_L \langle g \rangle G b_{\text{max}}}{h\nu T} \right)^2}{2NW(b_{\text{max}})} \right] dv \right\} . \quad (12.73)$$

The threshold voltage D is optimized so as to minimize the error rate. The incident average power on the preamplifier is given by

$$P_{\text{in}} = \frac{1}{2T} (b_{\text{max}} + b_{\text{min}}) . \quad (12.74)$$

Some numerical results on the TW-type preamplifier performance are presented in Fig. 12.11. Figure 12.11 shows the incident optical average power P_{min} at the preamplifier input to achieve a 10^{-9} error rate, versus preamplifier gain G . It is assumed that avalanche gain $\langle g \rangle$ is the unity photodiode operation, the detector dark current count λ_d is $6.3 \times 10^{-9} \text{ s}^{-1}$, the optical wavelength λ is $1.5 \mu\text{m}$, data rate B_0 is 1 Gbit/s, electronic amplifier input noise current $\sqrt{\langle i_a^2 \rangle}$ is 10 pA, and noise voltage $\sqrt{\langle e_a^2 \rangle}$ is 0.5 nV. Input pulse shape $h_p(t)$ is assumed to be a nonreturn-to-zero (NRZ) rectangular pulse:

$$h_p(t) = \begin{cases} \frac{1}{T} & \left(-\frac{T}{2} < t < \frac{T}{2} \right) \\ 0 & \text{(otherwise)} \end{cases} \quad (12.75)$$

$$H_p \left(\frac{2\pi f}{T} \right) = \frac{\sin \pi f}{\pi f} . \quad (12.76)$$

The equalized output pulse shape is assumed to be a full cosine roll-off pulse:

$$h_p(t) = \sin\left(\frac{\pi t}{T}\right) \cos\left(\frac{\pi t}{T}\right) / \left[\frac{\pi t}{T} \left(1 - \left(\frac{2t}{T}\right)^2\right)\right] \quad (12.77)$$

$$H_{\text{out}}(t) = \begin{cases} \frac{1}{2} \left[1 - \sin\left(\pi f - \frac{\pi}{2}\right)\right] & (0 < |f| < 1) \\ 0 & (\text{otherwise}) \end{cases} \quad (12.78)$$

The minimum detectable power obtained by an avalanche photodiode direct detection scheme is depicted for comparison in Fig. 12.11. The excess noise coefficient x of the avalanche detector depends on the diode material and structure.

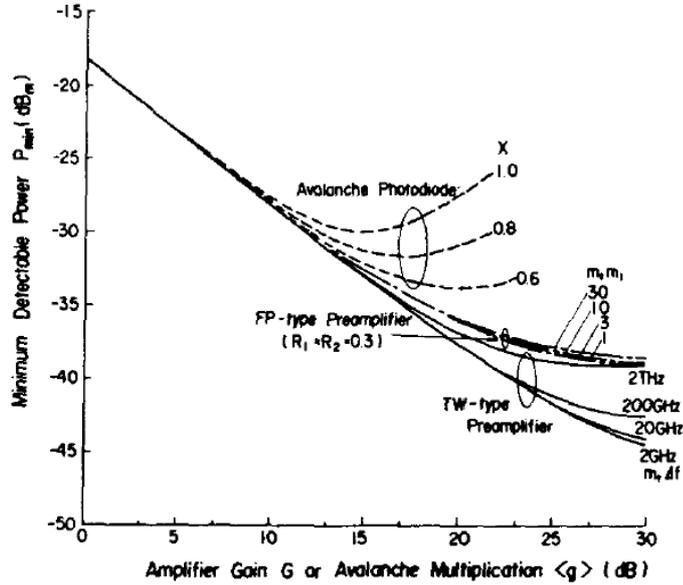


Figure 12.11: Minimum detectable powers P_{\min} to achieve a 10^{-9} error rate in the optical preamplifier and photodiode combination scheme versus preamplifier gain G . The minimum detectable power in the avalanche photodiode direct detection scheme versus the average avalanche gain $\langle g \rangle$ is also plotted. Here, optical wavelength $\lambda = 1.5\mu\text{m}$, data rate $B_0 = (1/T) = 1 \text{ Gbit/s}$, $\eta = 0.6$, $\lambda d = 6.3 \times 10^9 \text{ s}^{-1}$, detector load resistance $R_L = 50\Omega$, $\sqrt{\langle i_a^2 \rangle} = 10 \text{ pA}$, and $\sqrt{\langle e_a^2 \rangle} = 0.5 \text{ nV}$.

The dominant noise source in an optimized preamplifier realized by $m_t \Delta f \geq 2 \times 10^2 B_0$ is the beat noise between signal and spontaneous emission given by Eq. (12.65), which is inevitable in laser amplifiers. The signal-to-noise ratio is obtained from Eqs. (12.58) and (12.65):

$$\frac{\langle v_{\text{out}}(0) \rangle^2}{\langle N_{s-sp}^2 \rangle} = \frac{b_0}{2h\nu} : \quad G \gg 1 \quad (12.79)$$

The criterion $m_t \Delta f \leq 2 \times 10^2 B_0$ for optimized operation indicates that another dominant noise source in a high gain amplifier is the beat noise between spontaneous emission components given by Eq. (12.66), and it has to be made smaller than the beat noise between signal and spontaneous emission. The above condition is calculated from Eqs. (12.65) and (12.66) as

$$\frac{m_t \Delta f}{2B_0} < \frac{b_0}{h\nu} \quad . \quad (12.80)$$

The right side of Eq. (12.76) shows the number of input photons per pulse, which is more than 200, to achieve a 10^{-9} error rate.

12.3 Optical Repeater Amplifiers in PCM-IM Systems

Consider an optical linear repeater system shown in Fig. 12.12, in which the attenuated intensity modulated signal due to fiber loss L is recovered by gain G of the TW-type amplifier.

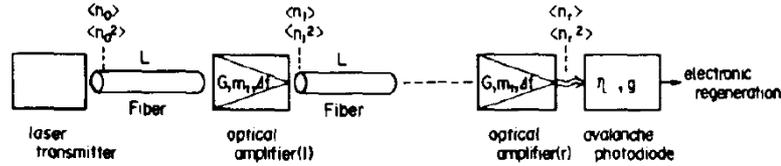


Figure 12.12: Optical repeater system using TW-type optical linear amplifiers. Here L is fiber loss and n_r is the number of r th repeater output photons.

The mean and variance in the number of output photons from the r th amplifier are calculated by using Eqs. (12.54) and (12.55) iteratively. Assuming that all amplifiers have the same gain G and optical noise bandwidth $m_t \Delta f$, and that fiber loss L is constant between any two repeaters, we have

$$\langle n_r \rangle = (GL)^r \langle n_0 \rangle + (G - 1) \gamma m_t \Delta f \cdot \frac{1 - (GL)^r}{1 - GL} \quad (12.81)$$

$$\begin{aligned} \langle n_r^2 \rangle - \langle n_r \rangle^2 &= (GL)^r \langle n_0^2 \rangle + (G - 1) \gamma m_t \Delta f \cdot \frac{1 - (GL)^r}{1 - GL} \\ &+ 2(GL)^r \langle n_0 \rangle (G - 1) \gamma \frac{1 - (GL)^r}{1 - GL} + (G - 1)^2 \gamma^2 m_t \Delta f \left[\frac{1 - (GL)^r}{1 - GL} \right]^2 \\ &+ (GL)^r (\langle n_0^2 \rangle - \langle n_0 \rangle^2 - \langle n_0 \rangle) \quad . \end{aligned} \quad (12.82)$$

Here $\langle n_0 \rangle$ and $\langle n_0^2 \rangle$ are the mean and mean square values for the laser transmitter output photons.

From Eqs. (12.77) and (12.78) it is possible to calculate signal $\langle v_{out} \rangle$ and worst case variance $NW(b_0)$ for the equalized output voltage defined by Eqs. (12.58) and (12.62). Let

the optical signal pulse shape falling on an avalanche photodiode be $h_p(t)$, the equalized pulse shape $h_{\text{out}}(t)$, and the fiber input energy of a pulse b_k . Then, we have

$$\langle v_{\text{out}}(t) \rangle = \frac{Ae\eta}{h\nu} \langle g \rangle R_L \left\{ \sum_{k=-\infty}^{\infty} (GL)^r \frac{b_k}{T} h_{\text{out}}(t - kT) \right\} \quad (12.83)$$

$$\langle N_s^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g^2 \rangle}{T^2} \frac{(GL)^r \eta}{h\nu} \{I_1 b_0 + (I_0 - I_1) b_{\text{max}}\} \quad (12.84)$$

$$\langle N_{\text{sp}}^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g^2 \rangle}{T} I_2 \eta (G - 1) \gamma m_t \Delta f \cdot \frac{1 - (GL)^r}{1 - GL} \quad (12.85)$$

$$\langle N_{\text{s-sp}}^2 \rangle = 2A^2 e^2 R_L^2 \cdot \frac{\langle g \rangle^2}{T^2} \cdot \frac{\eta^2 (GL)^r (G - 1) \gamma [1 - (GL)^r]}{h\nu(1 - GL)} \cdot \{I_1 b_0 + (I_0 - I_1) b_{\text{max}}\} \quad (12.86)$$

$$\langle N_{\text{sp-sp}}^2 \rangle = A^2 e^2 R_L^2 \frac{\langle g \rangle^2}{T} I_2 \eta^2 (G - 1)^2 \gamma^2 m_t \Delta f \cdot \left[\frac{1 - (GL)^r}{1 - GL} \right]^2 \quad (12.87)$$

$$\langle N_d^2 \rangle = \text{same as Eq. (12.67)}$$

$$\langle N_a^2 \rangle = \text{same as Eq. (12.68)}$$

The error rate is calculated by replacing G with $(GL)^r$ in Eq. (12.73).

The main feature of the repeater system performance is described by the signal-to-noise ratio defined as follows:

$$S/N = \frac{[\langle v_{\text{out}}(b_{\text{max}}) \rangle - \langle v_{\text{out}}(b_{\text{min}}) \rangle]^2}{\left[\frac{1}{2} (\sqrt{NW}(b_{\text{max}}) + \sqrt{NW}(b_{\text{min}})) \right]^2} \quad (12.88)$$

When the value of GL is equal to unity, the output signal level is constant for all repeaters, while the noise component is accumulated with the number of repeaters. The signal-to-noise ratio at the output of the r th repeater is shown in Fig. 12.13. It is assumed that the fiber input power is -10 dBm and amplifier gain is 25 dB. The worst case variance due to the signal-spontaneous emission beat noise $\langle N_{\text{s-sp}}^2 \rangle$ is in proportion to the number of repeaters r , while the variance due to the beat noise between spontaneous emission components $\langle N_{\text{sp-sp}}^2 \rangle$ is proportional to r^2 . Therefore, the signal-to-noise ratio decreases with r^{-1} for a narrow bandwidth ideal amplifier, while it decreases with r^{-2} for a broad bandwidth amplifier, as shown in Fig. 12.13.

Next, let us consider an optical amplifier repeater communication system by a simplified model. The S/N ratios at the transmitter output, the first fiber output, and the first amplifier output are given, respectively, by

$$(S/N)_0 = \frac{\langle n \rangle^2}{\langle \Delta n^2 \rangle} = \langle n \rangle \quad (\text{Poisson statistics}) \quad , \quad (12.89)$$

$$(S/N)_1 = \frac{L^2 \langle n \rangle^2}{L^2 \langle \Delta n^2 \rangle + L(1 - L) \langle n \rangle} = L \langle n \rangle \quad (\text{random deletion noise}) \quad , \quad (12.90)$$

$$(S/N)_2 = \frac{G^2 L^2 \langle n \rangle^2}{G^2 L \langle n \rangle + G(G - 1)(L \langle n \rangle + 1)} \simeq \frac{L}{2} \langle n \rangle \quad (3 \text{ dB noise figure}) \quad . \quad (12.91)$$

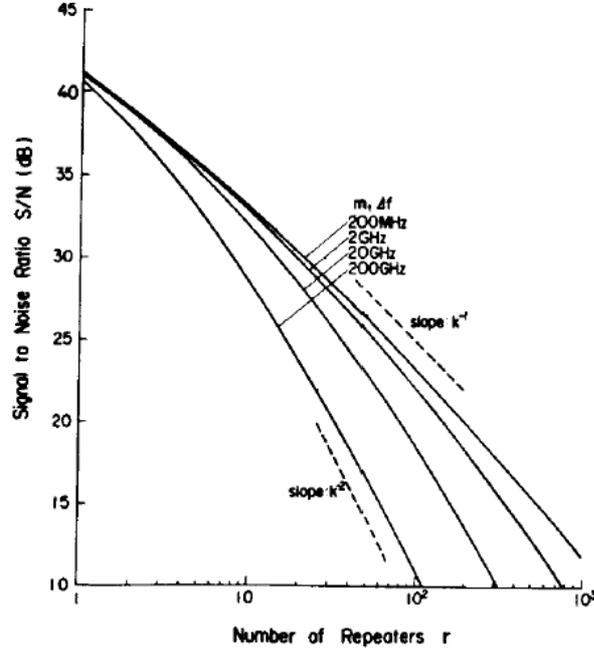


Figure 12.13: Signal-to-noise ratio S/N versus number of linear amplifier repeaters. Here, $\lambda = 1.5\mu\text{m}$, $B_0 = 100$ Mbits/s, linear amplifier input optical power is -35 dBm, $G = 25$ dB, $\eta = 0.6$, $\lambda_d = 6.3 \times 10^9$ s⁻¹, $\sqrt{\langle i_a^2 \rangle} = 10$ pA, and $\sqrt{\langle e_a^2 \rangle} = 0.5$ nV.

Equation (12.90) accounts for the random deletion noise due to fiber loss (Vergues variance theorem), which was derived in Chapter 2. The second equality in Eqs. (12.89) and (12.90) stems from the assumption that the signal does not have excess noise (i.e., $\langle \Delta n^2 \rangle = \langle n \rangle$). The S/N ratio is degraded by the fiber loss L and further degraded by 3 dB by the amplifier noise. The S/N ratio at the k -th amplifier output is calculated similarly:

$$(S/N)_{2k} = \frac{L}{2k} \langle n \rangle \quad . \quad (12.92)$$

In order to achieve the bit error rate of $P_e = 10^{-9}$, the S/N ratio must be larger than 20 dB. If one assumes $\langle n \rangle = 10^8$ and $L = \frac{1}{G} = 10^{-3}$ (30 dB loss/gain), the number of amplifier repeaters which can be put in the system is $k \simeq 500$. If the total input/output coupling loss of the amplifier and the isolator is 5 dB per section, the net gain is 25 dB, which corresponds to a total fiber loss of 12,500 dB or a total fiber length of 62,500 Km if the fiber loss is assumed to be $0.2 \left(\frac{\text{dB}}{\text{Km}} \right)$. The transmitter and optical amplifier output photon number of 10^8 per pulse corresponds to the optical power of 10 mW (10 dBm) if the bit rate is 1 Gbit/s ($T = 10^{-9}$ s) and the wavelength is $1.5 \mu\text{m}$. If the same output power of 10 mW is employed in a 10-Gbit/s system ($T = 10^{-10}$ s), the average photon number output per pulse is $\langle n \rangle = 10^7$. In this case, the maximum number of amplifier repeaters is $k \simeq 50$ and the total fiber length is 6250 Km.

If the optical amplifier repeater is not used and only the optical preamplifier is used, the total fiber loss and fiber length necessary to achieve $P_e = 10^{-9}$ are only 60 dB and 300 Km for a 1-Gbit/s system and 50 dB and 250 Km for a 10-Gbit/s system. Without the optical preamplifier, the total fiber loss and fiber length are, respectively, 40 dB and 200 Km for a 1 Gbit/s system and 30 dB and 150 Km for a 10 Gbit/s system. Therefore, one can see that the linear optical amplifier repeater can enormously expand the electronic terminal repeater spacing and thus is very useful for intercontinental under-sea communication systems.

12.4 Fundamental Limits of Communication Systems

12.4.1 Channel Capacity

Quantum and thermal noise place fundamental limits on communication systems. Shannon's channel capacity is one way to elucidate such fundamental limits. The channel capacity formulates how many bits of information can be transmitted over a given channel bandwidth B and S/N ratio:

$$C = B \log_2(1 + S/N) \quad (\text{bits/s}) \quad . \quad (12.93)$$

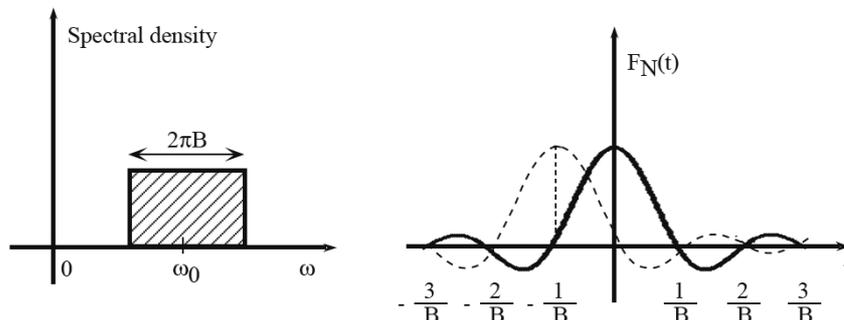


Figure 12.14: A channel bandwidth and displaced Nyquist functions.

The basis of the Shannon formula is the counting of degrees of freedom (DOF) which arrive at the receiver through the channel with a bandwidth B . This calculation was performed earlier by Nyquist. Consider the filtered spectral density with a bandwidth $\Delta\omega = 2\pi B$ centered at an angular frequency ω_0 , as shown in Fig. 12.14. The Fourier transform of this rectangular spectrum results in a Nyquist function, as shown in Fig. 12.14. A sequence of Nyquist functions,

$$F_N\left(t - \frac{k}{B}\right) = \frac{\sin\left[\pi B\left(t - \frac{k}{B}\right)\right]}{\pi B\left(t - \frac{k}{B}\right)} \quad (k = 0, \pm 1, \pm 2, \dots) \quad , \quad (12.94)$$

are orthogonal to each other and fully reproduce all bandwidth-limited functions. Each Nyquist function carries independent information and thus the arrival rate of the DOF in a communication channel with a bandwidth B is equal to B .

In the intensity modulation scheme, one can assign a different number of photons n for each Nyquist function (*DOF*) to transmit information. The (information theoretic) entropy of the signal is described by the probability $p(n)$ of having n photons:

$$H = - \sum_n p(n) \ln p(n) \quad . \quad (12.95)$$

One can maximize Eq. (12.105) under the constraints

$$\sum_n p(n) = 1 \quad (12.96)$$

and

$$\sum_n np(n) = \langle n \rangle \quad , \quad (12.97)$$

Equation (12.97) indicates that the average photon number is fixed in the entropy maximization procedure and the result is given by

$$H_{\max} = \langle n \rangle \ln \left(1 + \frac{1}{\langle n \rangle} \right) + \ln(1 + \langle n \rangle) \quad , \quad (12.98)$$

where $p(n)$ satisfies the so-called geometrical (thermal) distribution:

$$p(n) = \frac{\langle n \rangle^n}{(1 + \langle n \rangle)^{n+1}} \quad . \quad (12.99)$$

The channel capacity is simply the product of the arrival rate B of the *DOF* and the maximum entropy H_{\max} per *DOF*:

$$C = B \left[\langle n \rangle \ln \left(1 + \frac{1}{\langle n \rangle} \right) + \ln(1 + \langle n \rangle) \right] \quad . \quad (12.100)$$

The channel capacity in Eq. (12.100) is expressed in terms of natural digits (nats) per second and, in order to translate it into ordinary bits per second, one can divide Eq. (12.100) by $\ln 2$. The first term on the RHS is the product of the arrival rate of photon $B\langle n \rangle$ (photons/s) and the information carried by the number of *DOFs* per photon, $\frac{1}{\langle n \rangle}$, i.e., $H = \log \left(1 + \frac{1}{\langle n \rangle} \right)$; this term is called “photon entropy” or “particle entropy”. The second term on the RHS is the product of the arrival rate of *DOF*, B (*DOF*/s), and the information carried by the number of photons per *DOF* $\langle n \rangle$, i.e., $H = \log(1 + \langle n \rangle)$; this term is called “wave entropy.” As shown in Fig. 12.15, the channel capacity C is dominated by the particle entropy when $\langle n \rangle \ll 1$ and is dominated by the wave entropy when $\langle n \rangle \gg 1$.

The implicit assumption for the channel capacity Eq. (12.110) is that each pulse has a definite number of photons without fluctuation and the statistics of the photon number over many pulses obey the geometric distribution Eq. (12.109). It is also assumed that the photon counter is free from any noise and thus the photon number of each pulse is determined without error. The channel capacity provides the ultimate upper bound for such an ideal communication system.

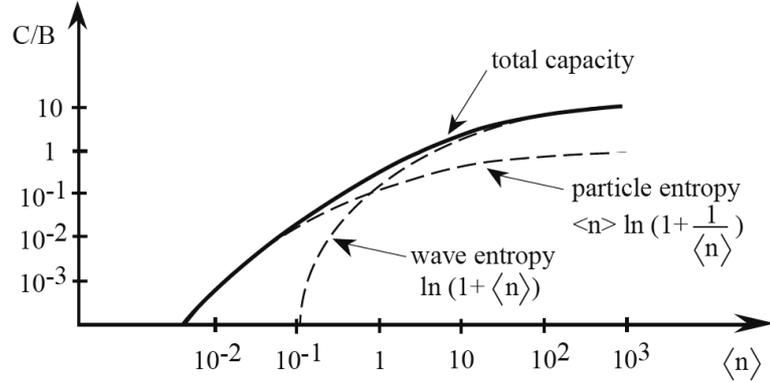


Figure 12.15: A quantum channel capacity normalized by a channel bandwidth vs. average photon number.

12.4.2 Quantum Limit of Communication Systems

The minimum energy cost to transmit one bit of information is calculated by

$$\Delta E \equiv \frac{\hbar\omega_0 B \langle n \rangle}{\left(\frac{C}{\ln 2}\right)} = \frac{\hbar\omega_0 \ln 2}{\frac{1}{\langle n \rangle} \ln(1 + \langle n \rangle) + \ln\left(1 + \frac{1}{\langle n \rangle}\right)} . \quad (12.101)$$

The second term in the denominator goes to infinity when $\langle n \rangle$ goes to zero, which means that the minimum energy cost to transmit one bit of information is reduced to zero in the limit of $\langle n \rangle \rightarrow 0$. In other words, in principle, “one photon can transmit infinite bits of information.” The information is carried by the particle entropy of a photon in such a case.



Figure 12.16: A pulse-position-modulation (PPM) signal.

Consider a pulse position modulation (PPM), as illustrated in Fig. 12.16. Each word with a time duration T is divided into M slots with a time interval $\tau = T/M$. A single photon arbitrarily occupies one slot out of M slots in each word; thus the information per word (i.e., per photon) is given by $\log_2 M$ (bits). When M becomes much larger than one, the average photon number per DOF, which corresponds to each slot, $\langle n \rangle = \frac{1}{M}$, becomes

much smaller than unity and the full particle entropy is approached by the PPM signal:

$$C_{\text{PPM}} = \frac{1}{T} \ln M \simeq B \langle n \rangle \ln \left(1 + \frac{1}{\langle n \rangle} \right) . \quad (12.102)$$

Obviously, $C_{\text{PPM}}/B \langle n \rangle$ (nats/photon) goes to infinity in the limit of $M \rightarrow \infty$ ($\langle n \rangle \rightarrow 0$).

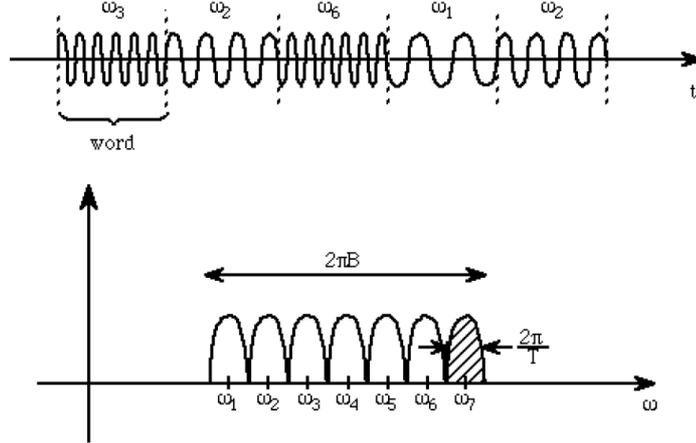


Figure 12.17: A multiple-frequency-shift-keying (MFSK) signal.

Another example of a modulation scheme for achieving more than one bit of information per photon is multiple-frequency shift keying (MFSK), as shown in Fig. 12.17. Each word with a time duration T is occupied by a photon of a different color. If the total channel bandwidth is $\Delta\omega \simeq 2\pi B$, there are $M = \frac{2\pi B}{(\frac{2\pi}{T})} = BT$ independent colors that can be accommodated in this channel. A single-color photon occupies one word out of M different colors, so that the information per word (i.e., per photon) is again given by $\log_2 M$ (bits). The channel capacity is the same as the PPM case:

$$C_{\text{MFSK}} = \frac{1}{T} \ln M \simeq B \langle n \rangle \ln \left(1 + \frac{1}{\langle n \rangle} \right) . \quad (12.103)$$

The above examples illustrate that there is no fundamental quantum limit as far as the minimum energy cost per bit is concerned and a finite photon energy $\hbar\omega_0$ does not impose a minimum energy cost per bit of information.

Next, consider the minimum time-energy product necessary to transmit one bit of information. The time duration required to transmit one bit of information is given by the inverse of the bit rate:

$$\Delta T \equiv \frac{\ln 2}{C} = \frac{\ln 2}{B \left[\langle n \rangle \ln \left(1 + \frac{1}{\langle n \rangle} \right) + \ln(1 + \langle n \rangle) \right]} . \quad (12.104)$$

Since the maximum available bandwidth B is on the order of the carrier frequency $f_0 = \omega_0/2\pi$, the time-energy product is written as

$$\Delta E \Delta T \simeq \frac{2\pi(\ln 2)^2 \hbar}{\langle n \rangle \left[\ln \left(1 + \frac{1}{\langle n \rangle} \right) + \frac{1}{\langle n \rangle} \ln(1 + \langle n \rangle) \right]^2} . \quad (12.105)$$

As shown in Fig. 12.18, the time-energy product takes the minimum value $\Delta E \Delta T \sim \frac{\hbar}{2}$ when $\langle n \rangle$ is equal to one. This relation is often referred to as the Bohr's time-energy uncertainty product; that is, when one tries to minimize the product of the energy cost and time duration per one bit of information, the quantum limit emerges and places a fundamental limit on communication systems equal to the Planck constant.

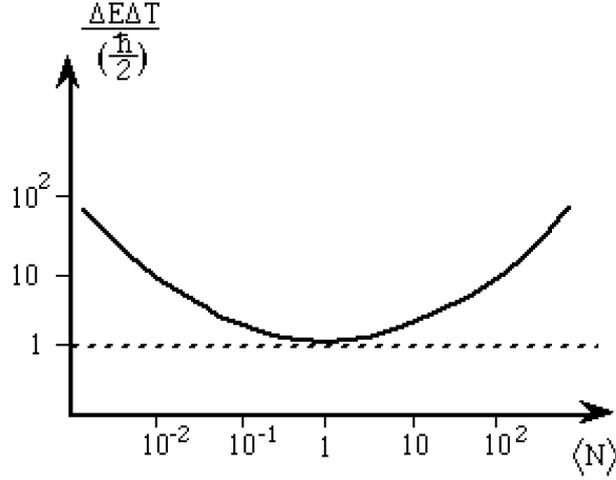


Figure 12.18: The time-energy product versus the average photon number.

12.4.3 Thermal Limit of Communication Systems

The above argument holds when the temperature of the communication system is absolute zero. When the temperature is finite, there are finite (thermal) noise photons n_{th} in each *DOF* as well as signal photons n_s . Suppose the total photon number $n = n_s + n_{\text{th}}$ obeys the thermal distribution Eq. (12.99). In this case, the maximum entropy is also equal to Eq. (12.98), but this maximum entropy cannot be fully extracted as useful information due to the presence of (thermal) noise photons. If the photon counter reports a photon number n_0 , the signal photon number is not necessarily equal to n_0 . The probability density for n_s has the following distribution:

$$P_{\text{meas}}(n_s) = \begin{cases} 0 & : n_s > n_0 \\ \frac{\langle n_{\text{th}} \rangle^{n_0 - n_s}}{(1 + \langle n_{\text{th}} \rangle)^{n_0 - n_s + 1}} & : n_s < n_0 \end{cases} . \quad (12.106)$$

This means that, even after the detection of the received photon number n_0 , there is still uncertainty with respect to the signal photon number n_s . This residual (noise) entropy is given by

$$\begin{aligned} H_{\text{noise}} &= \sum_{n_s} P_{\text{meas}}(n_s) \ln P_{\text{meas}}(n_s) \\ &= \langle n_{\text{th}} \rangle \ln \left(1 + \frac{1}{\langle n_{\text{th}} \rangle} \right) + \ln(1 + \langle n_{\text{th}} \rangle) . \end{aligned} \quad (12.107)$$

The useful information that can be extracted under thermal background photons is thus given by

$$\begin{aligned}
I \equiv H_{\max} - H_{\text{noise}} &= (\langle n_s \rangle + \langle n_{\text{th}} \rangle) \ln \left(1 + \frac{1}{\langle n_s \rangle + \langle n_{\text{th}} \rangle} \right) + \ln(1 + \langle n_s \rangle + \langle n_{\text{th}} \rangle) \\
&\quad - \langle n_{\text{th}} \rangle \ln \left(1 + \frac{1}{\langle n_{\text{th}} \rangle} \right) - \ln(1 + \langle n_{\text{th}} \rangle) \quad . \quad (12.108)
\end{aligned}$$

One can now calculate the minimum energy cost necessary to transmit one bit of information:

$$\Delta E = \frac{\hbar\omega_0 B \langle n_s \rangle}{\left(\frac{C}{\ln 2}\right)} \xrightarrow{\langle n_s \rangle \ll 1} k_B T \ln 2 \quad . \quad (12.109)$$

This is a thermal limit on the minimum energy cost per bit; for instance, at $T = 300^\circ \text{ K}$ and $\lambda = 1.5 \mu\text{m}$, a single photon can transmit

$$\frac{\hbar\omega_0}{k_B T \ln 2} = 46 \text{ bits} \quad , \quad (12.110)$$

or one bit of information costs ~ 0.022 photon.

Bibliography

- [1] Smith and Personick, (1980).
- [2] S. D. Personick, *Bell Syst. Tech. J.* **52**, 843 (1973).
- [3] McIntyre, (1966).
- [4] Melchior, (1978).
- [5] K. Shimoda, H. Takahashi, and C. H. Townes, “Fluctuations in amplification of quanta with application to maser amplifiers,” *J. Phys. Soc. Japan*, **12**, 686 (1957).
- [6] S. D. Personick, *Bell Syst. Techn. J.* **52**, 507 (1973).
- [7] Y. Yamamoto, *IEEE J. Quantum Electron.* **QE-16**, 1073 (1980).
- [8] H. Nyquist, *AIEE Trans.* **47**, 617 (1928).
- [9] C. E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948).
- [10] L. Brillouin, *Science and Information Theory* (Academic, New York, 1956).
- [11] Y. Yamamoto and H. A. Haus, *Rev. Mod. Phys.* **58**, 1001 (1986).