PERFORMANCE OF COHERENT OPTICAL RECEIVERS

by

John R. Barry

Memorandum No. UCB/ERL M88/16

1 March 1988

PERFORMANCE OF COHERENT OPTICAL RECEIVERS

by

John R. Barry

Memorandum No. UCB/ERL M88/16

1 March 1988

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

PERFORMANCE OF COHERENT OPTICAL RECEIVERS

by

John R. Barry

Memorandum No. UCB/ERL M88/16

1 March 1988

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Performance of Coherent Optical Receivers

John R. Barry

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, California 94720

ABSTRACT

Coherent optical fiber receivers have numerous advantages over direct detection receivers, most notably increased sensitivity and increased selectivity, at the cost of increased receiver complexity. It is shown that while the performance of practical intensity-modulation/direct-detection systems is well below the level predicted by the quantum limit, coherent systems show promise of attaining this limit. The sensitivity of coherent optical receivers under shot-noise-limited conditions is derived for PSK, ASK, FSK, and DPSK modulation formats. When applicable, homodyne, heterodyne, synchronous, and asynchronous systems are considered. In addition, laser phase noise is characterized, and its effects on the performance of FSK, ASK, and DPSK systems is analyzed.

This research was supported in part by the National Science Foundation, Grant MIP-86-57523.

Table of Contents

Performance of Coherent Optical Receivers

John R. Barry

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, California 94720

1. Introduction

Since its introduction nearly two decades ago, optical fiber communications has become the predominant choice for high-throughput point-to-point digital transmission. The advantages of optical fiber as a medium are many: large bandwidth, low attenuation, immunity to interference, and high security are the most obvious. The earliest fiber optic systems, employed in the 1970's, used multimode fibers at short wavelengths (850 nm). Improvements in photodetector technology allowed subsequent systems in the 1980's to operate at 1300 nm, the wavelength at which the attenuation of the fiber is minimum. In the mid-1980's, improved fiber design, as well as better coupling efficiencies, allowed operation at wavelengths of 1500 nm through single-mode fibers. These single-mode fibers avoid inter-modal dispersion, a primary disadvantage of multimode fibers. And through proper fiber design, the chromatic dispersion for single-mode fibers can be nearly eliminated for wavelengths in the 1300–1500 nm range.

The modulation scheme used by most commercial optical transmission systems, including those described above, is called *intensity-modulation/direct-detection* (IM/DD). This means that the intensity of the transmitting laser is modulated on and off, or on-off keying (OOK). The term *direct detection* stems from the receiver configuration where the received signal is applied directly to a photodetector. Note that only the intensity or power of the light is relevant; phase and frequency are for the most part ignored. The throughput of the optical fiber links presently installed throughout the country represents a significant improvement over that of their predecessors, coaxial cable and microwave radio links. Nevertheless, only a small fraction of fiber's potential is met through the use of these IM/DD systems. To make full use of the capabilities of fiber, *coherent* techniques must be used.

Coherent receivers employ the heterodyne technique used extensively in microwave communications. A common analogy compares IM/DD receivers to RF crystal radios, whereas optical coherent receivers are more like the modern superheterodyne radio receivers. In coherent detection, the received optical signal is added to a local optical signal, and the combined lightwave is directed towards a photodetector. For a heterodyne receiver, the frequency of the local oscillator (LO) is slightly different from the carrier frequency of the signal. The resulting current produced by the photodetector is then centered at an intermediate frequency (IF) equal to the difference in LO and carrier frequency, usually in the GHz range. This allows processing of the signal at microwave frequencies, where well-established radio techniques can be employed.

To fully understand the advantages of coherent techniques over direct detection, one must first study the limitations of IM/DD systems. Section 2 derives a fundamental limit on the sensitivity of an optical receiver called the *quantum limit.* Sensitivity is defined as the number of photons per bit necessary to maintain a bit-error ratio of 10^{-9}. The sensitivity of a practical IM/DD receiver is then derived, and it if found to be much less sensitive than the quantum limit.

Section 3 develops the theory behind coherent reception. First an example is given, where the theoretical sensitivity of homodyne phase-shift keying is found to be better than the quantum limit. The example neglects many sources of noise. However, under certain conditions, the most important being high local oscillator power, the receiver is said to be "shot-noise limited." In this case, it is shown that the quantum limit can be approached.

In section 3.2, the shot-noise-limited bit-error ratios are derived for various optical receivers. First synchronous receivers are analyzed, which offer the highest sensitivity, but require phase-locked loops (PLL), and are thus complicated. Asynchronous receivers, which use envelope detectors, are then discussed. Their sensitivity is shown to be only slightly worse (~ 0.5 dB) than their synchronous counterparts. Also a type of weakly-synchronous processing of differential phase-shift keying is discussed.

Although there are many similarities between conventional microwave communication systems and fiber systems, one can not always apply well known principles from one regime to the other. Section 4 discusses some of the unique traits of optical communications systems. A discussion on fibers, modulators, photodetectors, and polarization is presented. The automatic 3 dB loss in sensitivity of heterodyne systems with respect to homodyne is explained. And laser phase noise, a critical impairment in coherent systems, is discussed.

Finally, section 5 explores the impact of laser phase noise on receiver performance. Expressions for the maximum allowable laser linewidths are derived for FSK, DPSK, and ASK.

2. Direct Detection

This section develops some of the basic principles of optical communications. First, the quantum limit, a bound on the sensitivity of an optical receiver, is derived. Then, the sensitivity of a realistic IM/DD system is examined. Important expressions for the photodetector current and shot noise are derived, which will be used extensively in subsequent analysis.

A comment on notation: optical transmissions systems generally operate with an alphabet size of two. The improvements gained through M-ary signaling are to this date not worth their trouble. For this reason, ASK in the context of optical systems refers to OOK. Similarly, PSK means binary antipodal signaling. Also, this paper assumes that a *ONE* and a *ZERO* occur with equal probability.

2.1. The Photon Counter and the Quantum Limit

Light is a form of electromagnetic radiation, with frequencies in the 400 to 700 THz range. At these high frequencies, the quantum mechanical nature of the electromagnetic wave becomes prevalent, and the lightwave can act as if it were a stream of particles.

Each quanta of energy (called a photon) has energy $h\nu$ (Joules), where h = Planck's constant (Joules-sec) and ν is the frequency of the lightwave (Hz). If a lightwave were directed towards a surface, such as that of a photodetector, the arrival times of the photons would be randomly distributed according to a Poisson process.

Just as in microwave frequencies, a lightwave can be represented by either its electric or magnetic field. We know from the Poynting theorem that the power of a wave is proportional to the product of the amplitudes of the electric and magnetic fields. For the purpose of analysis, a lightwave is commonly represented by

$$x(t) = \sqrt{P_s}\cos\omega_0 t \tag{1}$$

where P_s is the average power (Joules/sec) of the wave, and ω_0 is the angular frequency. Since an average of P_s Joules arrive each second, and each photon has $h\nu$ Joules, the average number of photons per second is

$$\lambda_p \triangleq \frac{P_s}{h\nu} \text{ (Photons/sec).} \tag{2}$$

Define $n(t)$ as the photon arrival process, such that for $t > 0$, $n(t)$ = the number of photons that have arrived in the interval from 0 to t. Then equation (2) implies that $n(t)$, which as stated earlier is a Poisson counting process, has a mean arrival rate of λ_p.

An ideal OOK optical transmission system across a noiseless channel would transmit a pulse of light for a *ONE*, and no light for a *ZERO*. The receiver would then count n, the number of photons it receives in the bit interval (T seconds), and decide *ONE* if one or more photons were detected, and *ZERO* otherwise. This ideal system is depicted in figure 1 for an isolated pulse. The coefficient a_k is either 0 or 1, depending on the k^{th} data bit. This corresponds to amplitude-shift keying (ASK).

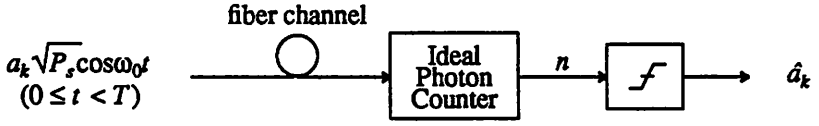

Figure 1. A photon counter: an idealized optical receiver counts the number of photons received in a bit interval, and compares the result with a zero threshold.

If $a_k = 0$, then there is zero probability of receiving any photons. If $a_k = 1$, then the photons arrive according to a Poisson process with mean rate λ_p. That is, for a *ONE*, the probability of obtaining n photons in T seconds is given by the Poisson distribution:

$$\Pr[n \text{ photons} \mid \textit{ONE}] = \frac{(\lambda_p T)^n e^{-\lambda_p T}}{n!}. \tag{3}$$

Thus assuming *ONE*s and *ZERO*s are equally likely, the probability of a bit error, or bit-error ratio (BER), is

$$P_e = \frac{1}{2}\Pr[0 \text{ photons} \mid \textit{ONE}] + \frac{1}{2}\Pr[>0 \text{ photons} \mid \textit{ZERO}]$$

$$= \frac{1}{2}e^{-\lambda_p T} + \frac{1}{2}(0)$$
$$= \frac{1}{2}e^{-P_s T/h\nu}. \tag{4}$$

To simplify notation, use equation (2) to define the exponent in equation (4) as

$$\bar{M} \triangleq \frac{P_s}{h\nu}T = \text{average number of photons per } \textit{ONE} \text{ bit.} \tag{5}$$

This leads to an important lower bound on the probability of a bit-error called the *quantum limit* [1] [2]:

$$P_e = \frac{1}{2}e^{-\bar{M}}. \tag{6}$$

Equation (6) represents a fundamental limit on the performance of an optical communications system: no ASK receiver can do better. For example, to achieve a BER of 10^{-9}, an uncoded ASK modulation scheme requires a minimum of $\bar{M} = 20$ photons per *ONE* bit, or an average of 10 photons per bit. The first value of 20 is more significant, however, because in optical fiber communications, the real limitation is peak power, not average power. Throughout the rest of this paper, the quantity of interest will be the *peak* power per bit.

2.2. Photodetector Current and Shot Noise.

In practice, photons cannot be counted directly, which renders the photon counter of figure 1 unrealistic. Instead, a photodetector is needed, which inevitably introduces additional noise. A photodetector converts incident photons to electron-hole pairs with efficiency η. The statistical properties of the resulting current will now be examined. The derivations that follow rely heavily on [3] and [4].

Define $n_e(t)$ as the electron generation process, such that for a given t, $n_e(t)$ is the number of electrons generated in the interval from 0 to t. Then $n_e(t)$ is also a Poisson process, with a mean rate of generation $\lambda_e \triangleq \eta\lambda_p$. Let $\{t_k\}$ be the Poisson generation times of the electrons, each of which will produce a small pulse of current, $h(t)$. The total photodetector current will be the sum of these individual pulses. The current can be modeled as the output of a filter driven by $z(t)$, a train of impulses at the times $\{t_k\}$. This impulse train can in turn be modeled as the time derivative of the electron counting process $n_e(t)$. A diagram of this representation is shown in figure 2(b). Since $h(t)$ is the current due to one electron, its total integral must be q, the charge of an electron:

$$\int_{-\infty}^{\infty} h(t)\,dt = q. \tag{7}$$

The mean and autocorrelation of the electron generation process $n_e(t)$ are [3]

$$\mathrm{E}[n_e(t)] = \lambda_e t, \tag{8}$$

and

$$R_n(t_1,t_2) = \lambda_e{}^2 t_1 t_2 + \lambda_e \min(t_1,t_2). \tag{9}$$

From these statistics of $n_e(t)$, we can derive the statistics of $z(t)$ (following [3]):

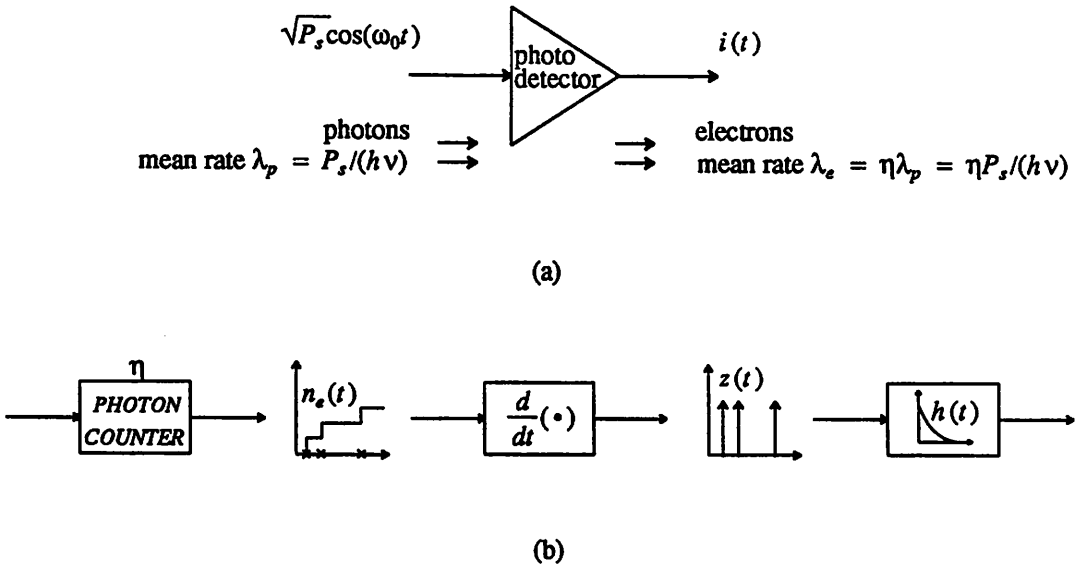

Figure 2. A photodetector converts photons to electrons with efficiency η: (a) block diagram representation, (b) equivalent statistical model.

$$\begin{aligned} \mathrm{E}[z(t)] &= \mathrm{E}\left[\frac{\partial}{\partial t} n_e(t)\right] \\ &= \frac{\partial}{\partial t}\mathrm{E}[n_e(t)] \\ &= \frac{\partial}{\partial t}(\lambda_e t) \\ &= \lambda_e. \end{aligned} \tag{10}$$

The autocorrelation function for $z(t)$ is

$$\begin{aligned} R_z(t_1,t_2) &= \frac{\partial^2}{\partial t_1 \partial t_2} R_{n_e}(t_1,t_2) \\ &= \frac{\partial^2}{\partial t_1 \partial t_2}\left[\lambda_e{}^2 t_1 t_2 + \lambda_e \min(t_1,t_2)\right] \\ &= \frac{\partial}{\partial t_1}\left[\lambda_e{}^2 t_1 + \lambda_e u(t_1 - t_2)\right] \end{aligned} \tag{11}$$

where $u(t)$ is the unit step function. This yields

$$R_z(\tau) = \lambda_e{}^2 + \lambda_e \delta(\tau). \tag{12}$$

We are now in a position to determine the second order statistics of $i(t)$. Since $i(t)$ is the convolution of $z(t)$ and $h(t)$, we have

$$\mathrm{E}[i(t)] = \mathrm{E}\left[\int_{-\infty}^{\infty} h(\tau)z(t-\tau)d\tau\right]$$

$$= \int_{-\infty}^{\infty} h(\tau)\mathrm{E}\left[z(t-\tau)\right]d\tau$$

$$= \lambda_e \int_{-\infty}^{\infty} h(\tau)d\tau$$

$$= q\lambda_e, \tag{13}$$

where the last equality follows from (7). Rewriting $i(t)$ in terms of its mean and deviation from this mean:

$$i(t) = q\lambda_e + i_{sh}(t), \tag{14}$$

where $i_{sh}(t)$ represents the current *shot noise :*

$$i_{sh}(t) \triangleq i(t) - q\lambda_e. \tag{15}$$

The goal of this discussion is to characterize the statistics of this shot noise current. Its autocorrelation function is

$$R_{i_{sh}}(\tau) = \mathrm{E}[i(t)i(t+\tau)] - q^2{\lambda_e}^2$$

$$= \mathrm{E}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z(u)z(v)h(t-u)h(t+\tau-v)\,dudv\right] - q^2{\lambda_e}^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R_z(u-v)h(t-u)h(t+\tau-v)\,dudv - q^2{\lambda_e}^2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[{\lambda_e}^2 + \lambda_e\delta(u-v)\right]h(t-u)h(t+\tau-v)\,dudv - q^2{\lambda_e}^2$$

$$= {\lambda_e}^2\left[\int_{-\infty}^{\infty} h(t)dt\right]^2 + \lambda_e\int_{-\infty}^{\infty} h(t)h(t+\tau)\,dt - q^2{\lambda_e}^2$$

$$R_{i_{sh}}(\tau) = \lambda_e\int_{-\infty}^{\infty} h(t)h(t+\tau)\,dt \tag{16}$$

Thus the spectral density of the shot noise process $i_{sh}(t)$ is the Fourier transform of (16):

$$S_{sh}(f) = \lambda_e|H(f)|^2. \tag{17}$$

The spectrum of $h(t)$ will roll-off at frequencies above $1/T_d$, where T_d is the

photodetector response time [5]. Alternatively, T_d can be thought of as the time-constant of $h(t)$. When T_d is small, so that the bandwidth of interest is less than $1/T_d$, then $H(f)$ can be approximated by $H(0)$. Thus from (17) we see that the shot noise is approximately white, with density

$$\begin{aligned} S_{sh}(f) &\approx \lambda_e |H(0)|^2 \\ &= \lambda_e \left[\int_{-\infty}^{\infty} h(t)\, dt \right]^2 \\ &= q^2 \lambda_e. \end{aligned} \tag{18}$$

The photodetector current due to a lightwave with power P_s can now be represented by

$$\begin{aligned} i(t) &= q\lambda_e + i_{sh}(t) \\ &= \left[\frac{\eta q}{h\nu} \right] P_s + i_{sh}(t) \\ &= RP_s + i_{sh}(t), \end{aligned} \tag{19}$$

where $R \triangleq (\eta q / h\nu)$ is the *responsivity* of the photodetector, measured in Amperes per Watt. The shot noise $i_{sh}(t)$ is a zero-mean white noise process, with density

$$\begin{aligned} S_{sh}(f) &= q^2 \lambda_e \\ &= q \left[\frac{\eta q}{h\nu} \right] P_s \\ &= qRP_s. \end{aligned} \tag{20}$$

It will be demonstrated in section 2.4 that IM/DD systems cannot approach the quantum limit due to extraneous noises, such as dark current and thermal noise. Therefore, to achieve acceptable performance, the received optical power P_s must be large. Consider the case when P_s is large enough so that the mean time between electron formations $\lambda_e^{-1} = h\nu/(\eta P_s)$ is much less than the time-constant T_d. For a given t, the current $i(t)$ is then the sum of a large number of tails of $h(t - t_k)$, where $\{t_k\}$ are independent. Therefore, $i(t)$ is the sum of a large number of independent random variables, and so it will have a distribution that is nearly Gaussian [4]. Thus for large P_s (so that $\lambda_e^{-1} \ll T_d$), the photodetector current will be approximated by

$$i(t) = RP_s + i_{sh}(t), \tag{21}$$

where $i_{sh}(t)$ is a zero-mean Gaussian white noise process with two-sided spectral density $S_{sh}(f) = qRP_s$. This concludes the derivation of the statistical properties of the

photodetector current and shot noise.

2.3. A Current Averager.

The ideal IM/DD receiver which originally used a photon counter is redrawn in figure 3 with a photodetector and current averager.

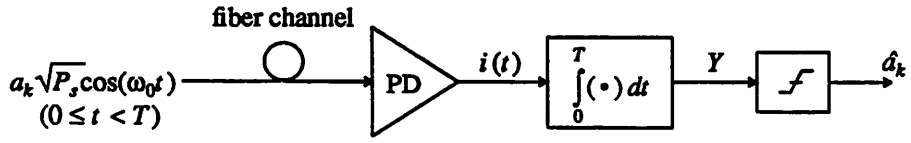

Figure 3. An ideal current averaging IM/DD receiver.

Assume that P_s is large. Using equation (21), the current produced by the photodetector in figure 3 is

$$i(t) = a_k R P_s + i_{sh}(t), \tag{22}$$

$i_{sh}(t)$ is approximated as a zero-mean Gaussian white noise process with density

$$S_{sh}(f) = a_k q R P_s. \tag{23}$$

The output of the integrator is

$$Y = a_k R P_s T + N, \tag{24}$$

where

$$N = \int_0^T i_{sh}(t)dt \tag{25}$$

is a zero-mean Gaussian random variable with variance

$$\begin{aligned} \mathrm{E}[N^2] &= \mathrm{E}\int_0^T\int i_{sh}(u) i_{sh}(v)\,du dv \\ &= a_k q R P_s \int_0^T\int \delta(u-v)\,du dv \\ &= a_k q R P_s T. \end{aligned} \tag{26}$$

The probability density function for Y conditioned on *ONE* and *ZERO* is shown in figure 4. Note that since we are considering shot noise only, Y is exactly zero when a *ZERO* is transmitted. Choosing a zero threshold, the receiver decides *ONE* when $Y > 0$, and *ZERO* otherwise. This corresponds to a BER of

$$\begin{aligned} P_e &= \frac{1}{2}\Pr[Y > 0 \mid \mathit{ZERO}] + \frac{1}{2}\Pr[Y \le 0 \mid \mathit{ONE}] \\ &= \frac{1}{2}(0) + \frac{1}{2}Q\left[\frac{RP_sT}{\sqrt{qRP_sT}}\right] \end{aligned}$$

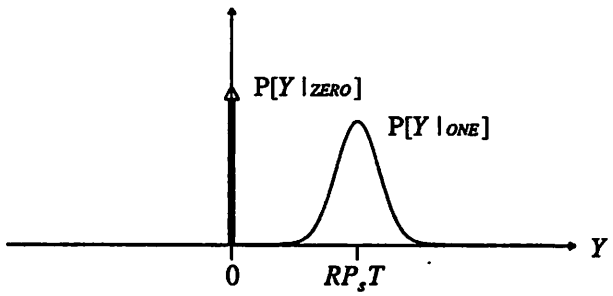

Figure 4. The likelihood functions for the idealized current averaging receiver of figure 3. When a *ZERO* is transmitted, Y is 0 with probability one, so that P[Y |*ZERO*] is a delta function.

$$= \frac{1}{2} Q\left[\sqrt{\frac{R}{q} P_s T}\right], \tag{27}$$

where the special Q-function is defined by

$$Q(\rho) \triangleq \int_{\rho}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$= \frac{1}{2} \operatorname{erfc}\left[\frac{\rho}{\sqrt{2}}\right]. \tag{28}$$

But the responsivity $R = \frac{\eta q}{h\nu}$, therefore

$$\frac{R}{q} P_s T = \eta \frac{P_s}{h\nu} T = \eta \bar{M}. \tag{29}$$

Define

$$M \triangleq \eta \frac{P_s}{h\nu} T = \eta \bar{M} = \text{number of photoelectrons per } \textit{ONE} \text{ bit.} \tag{30}$$

For coherent receivers, PIN diodes are generally used, which have efficiency η near unity. Therefore, the distinction between M and $\bar{M}$ is a subtle one.

The probability of error from (27) then becomes

$$P_e = \frac{1}{2} Q(\sqrt{M}) \approx \frac{1}{2} e^{-M/2}, \tag{31}$$

the second approximation being valid for $M \gg 1$. Since the exponent in equation (31) is half that in the quantum limit, we conclude that this system is approximately 3 dB *less sensitive* than the ideal photon counter. In other words, twice as much power per bit is needed to maintain the same BER. This difference is due entirely to the approximation of the electron flow as a white Gaussian process.

2.4. A Practical IM/DD Receiver

The above analysis of the photon counter and current averager considered only shot noise. As one might expect, however, in practice other noise sources must be accounted for. Real photodetectors produce a *dark current* regardless of the presence or absence of incident photons, due to the spontaneous formation of electron-hole pairs [6]. Also, background light can hit the detector's surface [7]. And since the current produced by a photodetector can be relatively small, the circuit or thermal noise of the next stage (usually a FET amplifier) can be significant [6]. A more realistic IM/DD receiver will now be considered, using the block diagram in figure 5.

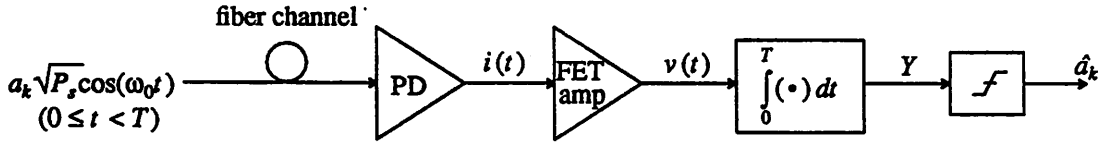

Figure 5. In practical IM/DD receivers, the noises due to dark current and thermal noise in the amplifier front-end must be dealt with.

When the above mentioned noises are accounted for, the voltage at the output of the FET amplifier is

$$v(t) = a_k ZRP_s + n_{sh}(t) + n_{dk}(t) + n_{th}(t) + n_{bk}(t). \tag{32}$$

Here Z is the equivalent impedance of the FET amplifier, and $n_{sh}(t)$, $n_{dk}(t)$, $n_{th}(t)$, and $n_{bk}(t)$ represent the shot noise, dark current noise, thermal noise, and background noise respectively. They can be modeled as zero-mean white Gaussian random processes over the frequency range of interest [7], with two-sided spectral densities

$$\begin{aligned} S_{sh}(f) &= a_k qZ^2 RP_s, \\ S_{dk}(f) &= N_{dk} = qZ^2 I_{dk}, \\ S_{th}(f) &= N_{th}, \\ S_{bk}(f) &= N_{bk} = qZ^2 RP_{bk}. \end{aligned} \tag{33}$$

Notice that there is more noise when a *ONE* is transmitted than for a *ZERO*. Computing Y, the output of the integrator:

$$Y = \int_0^T [a_k ZRP_s + n_{sh}(t) + n_{bk}(t) + n_{th}(t)]dt \tag{34}$$

Define the conditional mean and variance of Y as

$$m_0 \triangleq \mathrm{E}[Y \mid \mathit{ZERO}] = 0, \tag{35}$$

$$m_1 \triangleq \mathrm{E}[Y \mid \mathit{ONE}] = ZRP_s T, \tag{36}$$

$$\sigma_0^2 \triangleq \mathrm{var}[Y \mid \mathit{ZERO}] = N_{dk}T + N_{th}T + N_{bk}T, \tag{37}$$

$$\sigma_1^2 \triangleq \mathrm{var}[Y \mid \mathit{ONE}] = N_{dk}T + N_{th}T + N_{bk}T + qZ^2 RP_s T. \tag{38}$$

Figure 6 shows the probability distributions for Y conditioned on *ONE* and *ZERO*.

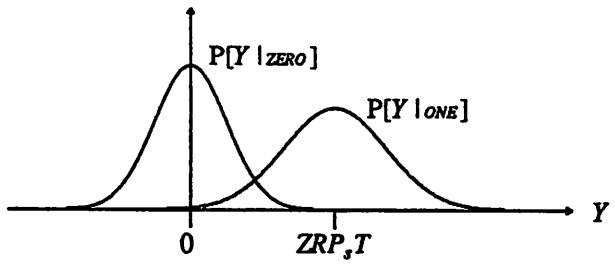

Figure 6. The variance of the noise in IM/DD receivers depends on the transmitted bit.

Since *ONE*s and *ZERO*s are assumed equally likely, the optimum threshold Y_{th} that minimizes the BER is the value of Y at which the two likelihood functions intersect (hence the name *maximum likelihood,* or *ML).* Since the variance of the two distributions are not equal, this threshold is not immediately obvious. One can approximate it by finding the $\hat{Y}_{th}$ that causes the probability of a false alarm to equal that of a miss [6] [7]:

$$\Pr[Y > \hat{Y}_{th} \mid ZERO] = \Pr[Y < \hat{Y}_{th} \mid ONE], \tag{39}$$

or

$$Q\left[\frac{\hat{Y}_{th} - m_0}{\sigma_0}\right] = Q\left[\frac{m_1 - \hat{Y}_{th}}{\sigma_1}\right],$$

$$\frac{\hat{Y}_{th} - m_0}{\sigma_0} = \frac{m_1 - \hat{Y}_{th}}{\sigma_1},$$

$$\hat{Y}_{th} = \frac{m_0\sigma_1 + m_1\sigma_0}{\sigma_0 + \sigma_1}. \tag{40}$$

To justify this approximation, proceed more formally: the exact threshold Y_{th} satisfies

$$\frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(Y_{th} - m_0)^2/2\sigma_0^2} = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-(Y_{th} - m_1)^2/2\sigma_1^2}. \tag{41}$$

For the example at hand, this leads to

$$Y_{th} = m_1 \left[\frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2}\right] \left[1 - \frac{\sigma_1}{\sigma_0}\sqrt{1 + \frac{(\sigma_1/\sigma_0) ln\,(\sigma_1/\sigma_0)^2}{M}}\right], \tag{42}$$

where $M = m_1/(qZ)$ is the mean number of photons per *ONE* bit. Then for $M \gg (\sigma_1/\sigma_0)\ln(\sigma_1/\sigma_0)^2$, this reduces to

$$Y_{th} \approx m_1 \left[\frac{\sigma_0^2}{\sigma_1^2 - \sigma_0^2}\right] \left[1 - \frac{\sigma_1}{\sigma_0}\right], \tag{43}$$

or

$$Y_{th} \approx \frac{m_1\sigma_0}{\sigma_1+\sigma_0}, \tag{44}$$

which, because $m_0 = 0$, supports the threshold obtained from equation (40).

With this approximation, the BER becomes

$$P_e = \Pr[Y > \hat{Y}_{th} \mid ZERO], \tag{45}$$

and substituting equation (40) yields

$$P_e = Q\left[\frac{m_1 - m_0}{\sigma_1 + \sigma_0}\right]. \tag{46}$$

Plugging in the conditional means and variances from equations (35)-(38), we find the BER for a realistic IM/DD receiver to be

$$P_e = Q\left[\frac{ZRP_sT}{\sqrt{N_{dk}T + N_{th}T + N_{bk}T} + \sqrt{N_{dk}T + N_{th}T + N_{bk}T + qZ^2RP_sT}}\right]. \tag{47}$$

This BER is plotted versus peak received power P_s in figure 7, using the following typical parameters [7]:

I_{dk} = 1 nAmp (dark current)

$N_{th}/Z^2 = 10^{-16}$ Amps^2 (thermal current variance)

$P_{bk} = 10^{-19}$ Watts (background light power)

$1/T$ = 100 Mbps (bit rate)

λ = 1500 nm (wavelength)

For purposes of comparison, the BER for the ideal photon counter is included in the graph. It is evident that the performance of the receiver is very much inferior, as approximately 26 dB more power is needed to maintain a given BER. In terms of photons per bit, this receiver needs nearly 8000 photons per *ONE* bit for a BER of 10^{-9}. And although there are some methods for improving the sensitivity of IM/DD receivers, (e.g., through the use of an avalanche photodiode), practical IM/DD receivers cannot achieve the level of performance predicted by the quantum limit.

3. Coherent Detection

It was shown in the last section that the dark current and preamplifier thermal noise caused a 26 dB degradation in sensitivity with respect to the quantum limit. To combat the thermal noise, IM/DD receivers typically use an avalanche photodiode (APD), which has an internal gain. The APD current is then large enough so that the preamplifier thermal noise is negligible. However, the gain of an APD fluctuates randomly, which introduces another source of noise. Furthermore, the efficiency of an APD is typically half that of a PIN diode. Overall, the net improvement brought on by an APD for an IM/DD system is 5-10 dB; still 15-20 dB worse than the quantum limit [7].

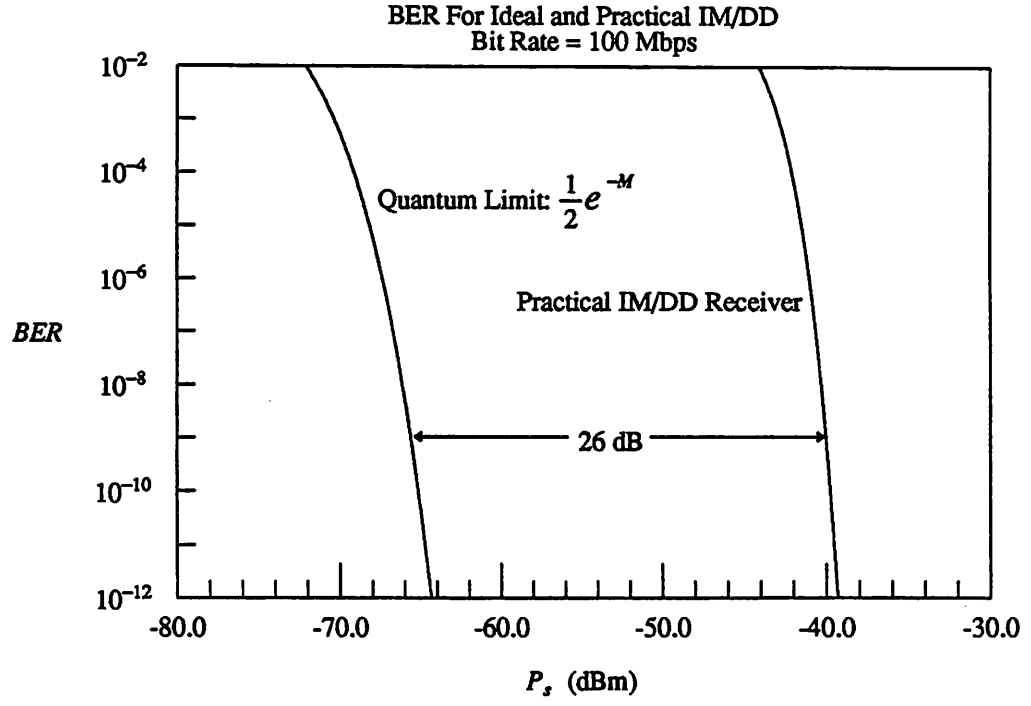

Figure 7. The BER curves for the quantum limit and a practical IM/DD receiver, at a bit rate of 100 Mbps and wavelength λ = 1500nm. The parameters for the practical receiver are mean gain $<M>$ = 1, η = 1 (PIN diode), thermal current variance = 10^{-16} A^2, dark current = 1nA, background light power 10^{-19} W.

If the optical power could somehow be amplified before the photodetector, then a PIN diode could be used, since its current would then be large enough to overcome the effects of thermal noise. This avoids the noisy gain and low efficiency of the APD. But how does one achieve this high signal power? An optical amplifier is one possibility. It turns out that these devices are theoretically possible, but difficult to realize. But by mixing the received lightwave with a local lightwave, the power of the resulting sum is proportional to the received power times the LO power. If P_{LO} is large, this has the effect of amplification, and the advantages described above can be gained.

In optical communications, *coherent* has come to describe any receiver that adds a LO lightwave to the incoming wave, even if subsequent processing ignores its phase, as do envelope detectors. To differentiate between a correlation type receiver and an envelope detector, the terms *synchronous* and *asynchronous* are used.

3.1. PSK: An Example

To illustrate the principles of a coherent receiver, consider a single pulse of a PSK modulated lightwave $a_k\sqrt{P_s}\cos\omega_0 t$, where $a_k = \pm 1$. Add to this signal a local lightwave with power P_{LO} and frequency ω_{LO}. The resulting sum is then directed towards a photodetector, as shown in figure 8. With proper attention to the phase and

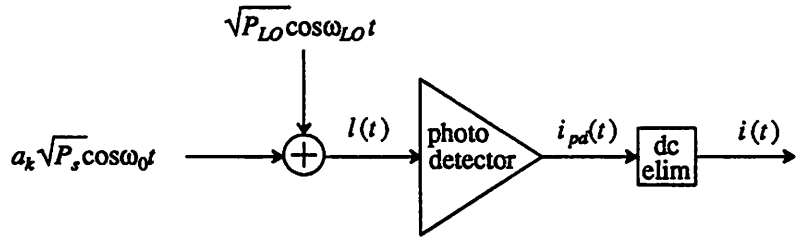

Figure 8. The front end of a coherent receiver *adds* an LO lightwave to the incoming signal. The resulting photodetector current is passed through a *dc eliminator,* which removes dc components, and can be realized by a simple RC high-pass filter.

polarization of the lightwaves, the sum $l(t)$ is

$$\begin{aligned} l(t) &= a_k\sqrt{P_s}\cos\omega_0 t + \sqrt{P_{LO}}\cos\omega_{LO}t \\ &= \sqrt{a_k P_s + P_{LO} + 2a_k\sqrt{P_s P_{LO}}\cos(\omega_0 - \omega_{LO})t}\ \cos[\omega_0 t + \beta(t)]. \end{aligned} \quad (48)$$

The second equality follows from an expansion of $l(t)$ in terms of its *envelope* and phase about ω_0. The power of $l(t)$ is

$$P_{sum}(t) = a_k P_s + P_{LO} + 2a_k\sqrt{P_s P_{LO}}\cos\omega_{IF}t, \quad (49)$$

where $\omega_{IF} = |\omega_0 - \omega_{LO}|$ is the intermediate frequency.

Using the result from last section (see (32)), the current produced by the photodetector can be modeled by

$$\begin{aligned} i_{pd}(t) &= RP_{sum}(t) + i_{sh}(t) + i_{other}(t) \\ &= a_k RP_s + RP_{LO} + 2a_k R\sqrt{P_s P_{LO}}\cos\omega_{IF}t + i_{sh}(t) + i_{other}(t), \end{aligned} \quad (50)$$

where $i_{sh}(t)$ is a zero-mean Gaussian random process with spectral density

$$\begin{aligned} S_{sh}(f) &= qRP_{sum} \\ &= qR(a_k P_s + P_{LO} + 2a_k\sqrt{P_s P_{LO}}\cos\omega_{IF}t). \end{aligned} \quad (51)$$

When $P_{LO} \gg P_s$, which will be shown to be desirable and is easy to achieve, the density of the shot noise is constant, independent of the transmitted bit:

$$S_{sh}(f) \approx qRP_{LO}. \quad (52)$$

The term $i_{other}(t)$ in (50) models all of the other noises (dark current, thermal, etc.) as a zero-mean white Gaussian process with density $S_{other}(f) = N_{other}$. After dc elimination, the current is

$$i(t) = a_k 2R\sqrt{P_s P_{LO}}\cos\omega_{IF}t + i_{sh}(t) + i_{other}(t). \quad (53)$$

The SNR for this current relative to the signal space representation with basis function $\psi(t) = \sqrt{2/T}\cos\omega_{IF}t$ can be computed:

$$\text{SNR} = \frac{\left(\int_0^T a_k 2R\sqrt{P_s P_{LO}}\cos\omega_{IF}t\ \psi(t)\,dt\right)^2}{\mathrm{E}\left[\int_0^T (i_{sh}(t) + i_{other}(t))\,\psi(t)\,dt\right]^2},$$

$$= \frac{2TR^2 P_s P_{LO}}{qRP_{LO} + N_{other}}. \tag{54}$$

The signal term is proportional to P_{LO}, as is the shot noise term. But the other noises represented by N_{other} are generally independent of P_{LO} (see [8] for an exception). Thus one can reduce the effect of these other noises by increasing P_{LO}. In the limit, as P_{LO} tends to ∞, the resulting SNR is

$$\text{SNR} = \frac{2TR^2 P_s P_{LO}}{qRP_{LO}}$$

$$= 2M. \tag{55}$$

For this shot-noise-limited case, then, we can ignore the other noises, and equation (53) becomes

$$i(t) \approx a_k 2R\sqrt{P_s P_{LO}}\cos\omega_{IF}t + i_{sh}(t), \tag{56}$$

where $i_{sh}(t)$ is a zero-mean white Gaussian process with density $S_{sh}(f) = qRP_{LO}$. The problem has thus been reduced to the classical detection problem of a signal in additive white Gaussian noise. The optimum receiver is well known to be a correlator followed by a threshold test, as illustrated in figure 9.

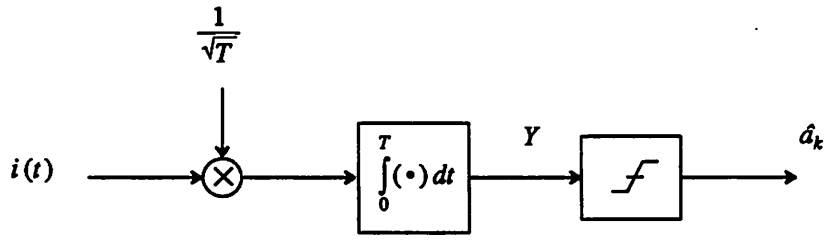

Figure 9. The ML receiver for signals in additive white Gaussian noise is a correlator, often implemented with a matched filter.

In a signal space representation, the received signal $i(t)$ is modeled as a vector in a finite-energy infinite-dimensional linear space. For known signals in additive white

Gaussian noise, the space can be restricted to the subspace spanned by the *known* signals, before they are corrupted by noise. For PSK, the subspace is one dimensional. For the example at hand, let $\omega_{LO} = \omega_0$, so that $\omega_{IF} = 0$; i.e., homodyne PSK. In this case, the known signal portion of the current is just a constant:

$$i(t) = a_k 2R\sqrt{P_s P_{LO}} + i_{sh}(t). \tag{57}$$

Let the one-dimensional subspace be represented by the normalized basis vector ψ, where

$$\psi \leftrightarrow \psi(t) = \sqrt{\frac{1}{T}}$$

The above notation means that the signal $\psi(t)$ is represented in signal space by the vector ψ. For signals in additive white Gaussian noise, the optimal receiver need only compute the *sufficient statistic,* which is the component of the incoming signal in the basis vector ψ direction. This is because the components of the noise in any other direction are uncorrelated with those in the *sufficient subspace.* (For a more complete description of signal space representations, see, e.g., [9], [10].) The component of $i(t)$ in the $\psi(t)$ direction is found by integrating their product:

$$Y = \sqrt{1/T}\int_0^T i(t)dt = a_k 2R\sqrt{P_s P_{LO} T} + N, \tag{58}$$

where N is a zero-mean Gaussian random variable with variance qRP_{LO}. The signal space constellation is shown in figure 10.

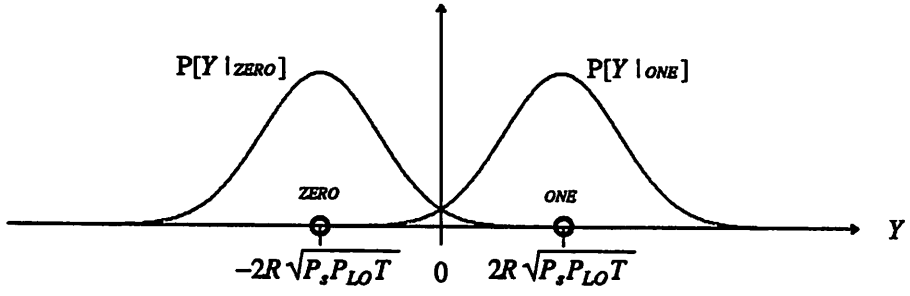

Figure 10. The signal constellation for 2-level PSK, often referred to as binary antipodal signaling.

The BER is then

$$P_e = Q\left[\frac{4R\sqrt{P_s P_{LO} T}}{2\sqrt{qRP_{LO}}}\right] \tag{59}$$

$$= Q(\sqrt{4M}). \tag{60}$$

Comparing with equation (6), we see that under shot-noise-limited conditions, the performance of homodyne PSK is slightly (~ 3.5 dB) better than the quantum limit! Note that 3 dB of this advantage is due to the fact that a PSK signal is transmitting at peak power for both *ones* s and *zero* s, something an IM/DD system cannot do. Equation (60) implies that 9 photons per bit are needed for a homodyne PSK receiver to achieve a BER of 10^{-9}. This result illustrates the increase in receiver sensitivity promised by coherent detection. And although the above analysis neglected some important issues (such as laser phase noise), it will be shown that the degradation of this result in a more complete analysis will not be severe.

3.2. Computation of BER for Shot Noise Limited Operation

What follows is a compilation of BER computations for various homodyne, heterodyne, synchronous and asynchronous modulation formats. The spirit of these calculations is to develop a framework in which the relative merits of the various formats can be compared. They will serve only as design guidelines, since these values will represent the lower limit on the BER in the ideal case of shot-noise-limited operation. The receiver structures and associated analysis in this section are for the most part standard; see, e.g., [11] [12] [10]

To make the analysis tractable, many assumptions were made:

(i) No laser phase noise.

(ii) Shot noise $\gg$ thermal, dark current noise, etc.

(iii) Perfect modulation of laser.

(iv) Square pulse shape ($0 \le t < T$), no intersymbol interference.

(v) Perfect phase tracking by PLL (for synchronous receivers).

(vi) Combined lightwaves are identically polarized.

If these assumptions are not made, the analysis would become excessively complicated. The effects of laser phase noise will be considered in section 5. If other noise sources besides the shot noise are significant, the result will be a uniform power penalty, so that the relative performance levels derived will stay intact [13]. The difficulty in cleanly modulating lasers results in inadvertent intensity and frequency noise, which in general must be accounted for. Since optical fibers have a large bandwidth and optical PAM is very narrow-band, dispersion, and thus pulse shaping, equalization, and intersymbol interference, are neglected. The problem of realizing optical phase-locked loops (PLL) is also an important one, but it too will be neglected here. Finally, a means of ensuring identical polarization when combining lightwaves is needed. This can be done through polarization-maintaining fiber, or more practically through some form of automatic polarization adjusters [14].

The front end of the receivers considered throughout this section will all be the same: the received lightwave is mixed with a LO lightwave, the sum directed towards a PIN photodetector, and the resulting current passed through a dc eliminator (see figure 8.)

Heterodyne demodulation results in an IF signal, which can be processed in three distinct ways. *Synchronous* receivers use a correlator or matched filter, and thus require a PLL. *Asynchronous* receivers use an energy or envelope detector. Another alternative, denoted as *weakly-synchronous* processing, uses some form of an autocorrelator. Since

homodyning results in a baseband signal, the transmitted signal can be recovered directly, so homodyne receivers are usually included in the class of synchronous receivers.

3.2.1. Synchronous Processing

PSK

In the last section, the shot-noise-limited BER for homodyne PSK was found to be $Q(\sqrt{4M})$. The result for heterodyne PSK can be found using similar methods. From equation (56), the current produced when a PSK signal is heterodyned using the technique in figure 8 is

$$i(t) = a_k 2R\sqrt{P_s P_{LO}}\cos\omega_{IF}t + i_{sh}(t), \tag{61}$$

where $i_{sh}(t)$ is a zero-mean Gaussian white noise process with density $S_{sh}(f) = qRP_{LO}$. The ML receiver is a correlator, as shown in figure 9, with basis function

$$\psi(t) = \sqrt{2/T}\cos\omega_{IF}t. \tag{62}$$

The input to the slicer can be computed:

$$Y = \int_0^T i(t)\psi(t)dt = a_k R\sqrt{2TP_s P_{LO}} + N, \tag{63}$$

where N is a zero-mean Gaussian random variable with variance qRP_{LO}. The signal constellation is shown below in figure 11.

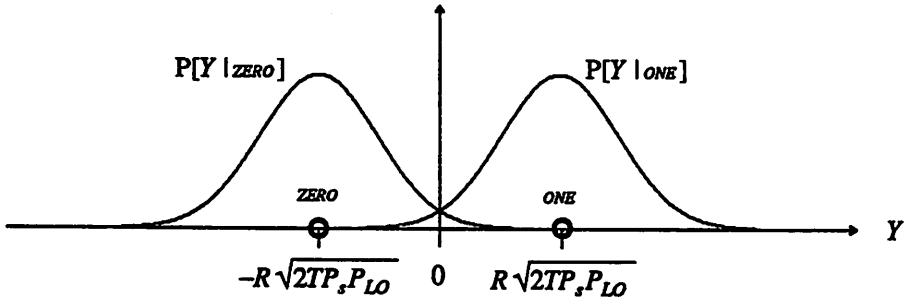

Figure 11. The signal space constellation for heterodyne PSK. It is the same as for homodyne PSK, except the basis vector $\psi(t)$ has changed.

The BER is then

$$P_e = Q\left[\frac{R\sqrt{2TP_sP_{LO}}}{\sqrt{qRP_{LO}}}\right] = Q(\sqrt{2M}). \tag{64}$$

The performance of heterodyne PSK is seen to be 3dB worse than homodyne PSK, and asymptotically equal to the quantum limit, as illustrated in figure 12. The corresponding sensitivity for heterodyne PSK is 18 photons per bit.

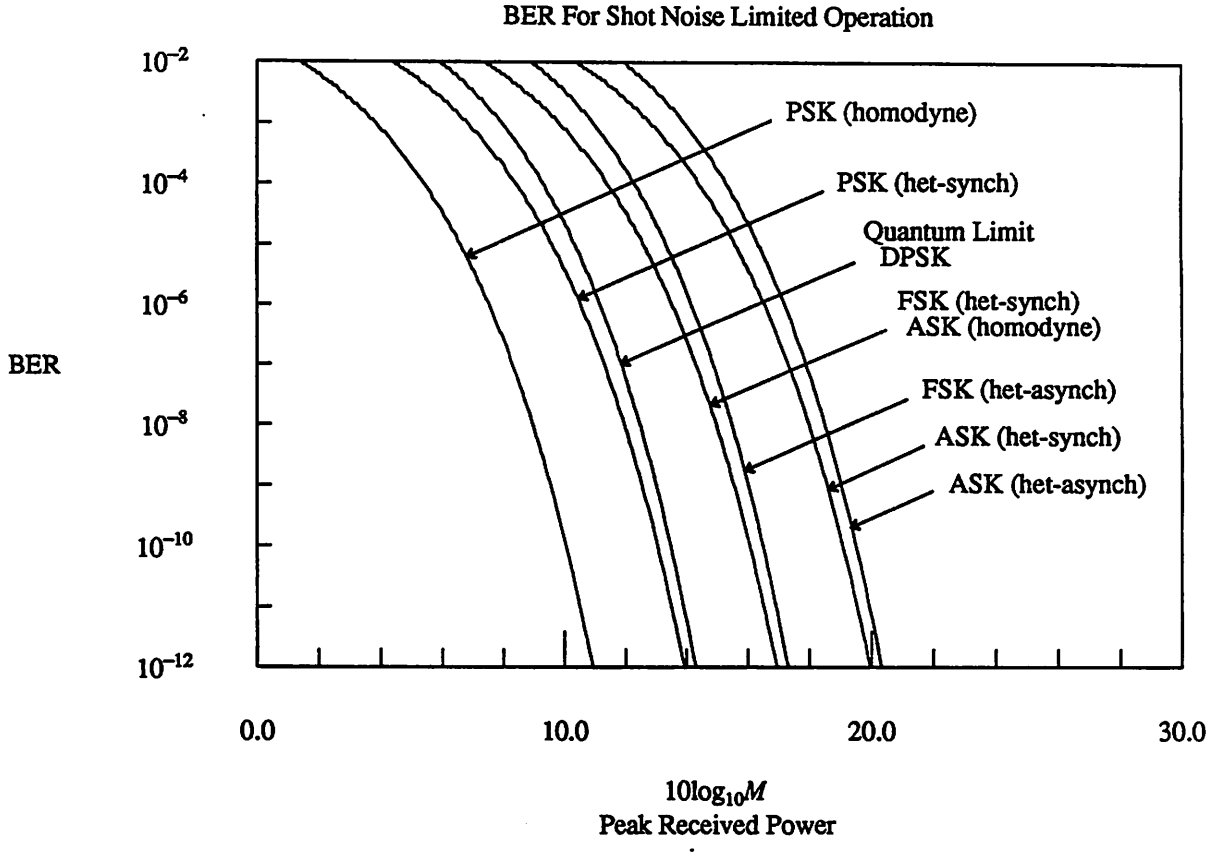

Figure 12. BER curves for shot-noise-limited operation, where BER is the probability of a symbol error, and M is the number of photons per *ONE* bit. M is proportional to the received peak power; $M = (\eta P_s / h\nu)T$.

FSK

For frequency-shift keying (FSK), the transmitter sends a light pulse with one of two possible frequencies $f_0 \pm f_d$, where f_d is the deviation frequency. The coherent receiver adds a LO lightwave as before, see figure 13.

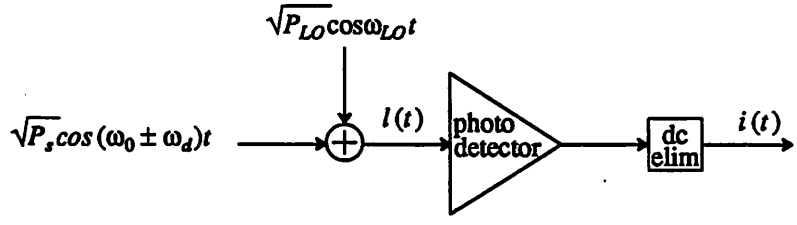

Figure 13. The FSK coherent receiver is similar to that for PSK.

In this case, the combined lightwave $l(t)$ is

$$l(t) = \sqrt{P_s}\cos(\omega_0 \pm \omega_d)t + \sqrt{P_{LO}}\cos\omega_{LO}t$$
$$= \sqrt{P_s + P_{LO} + 2\sqrt{P_s P_{LO}}\cos(\omega_0 - \omega_{LO} \pm \omega_d)t}\ \cos[\omega_0 t + \beta(t)], \quad (65)$$

so that the resulting current, proportional to the power of this lightwave, is

$$i(t) = 2R\sqrt{P_s P_{LO}}\cos(\omega_{IF} \pm \omega_d)t + i_{sh}(t), \quad (66)$$

where $i_{sh}(t)$ is again a zero-mean Gaussian white noise process with density $S_{sh}(f) = qRP_{LO}$. For large ω_{IF}, this corresponds to orthogonal signaling. The resulting signal space has two dimensions, as shown in figure 14.

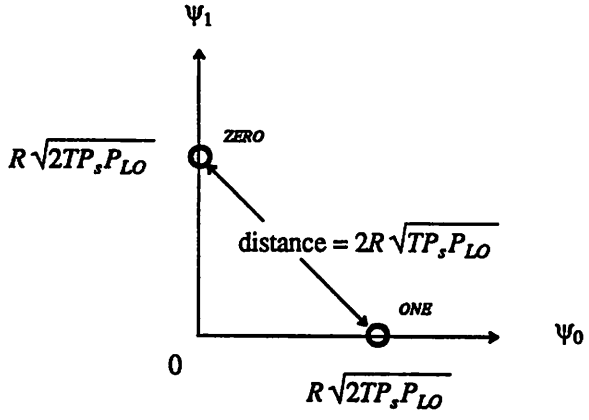

Figure 14. FSK is an orthogonal signaling modulation scheme.

Here,

$$\psi_0 \leftrightarrow \sqrt{\frac{2}{T}}\cos(\omega_{IF} + \omega_d)t, \quad (67)$$

and

$$\psi_1 \leftrightarrow \sqrt{\frac{2}{T}}\cos(\omega_{IF} - \omega_d)t. \quad (68)$$

In this space, a *ONE* is represented by $R\sqrt{2TP_sP_{LO}}\psi_1$, and a *ZERO* by $R\sqrt{2TP_sP_{LO}}\psi_0$. The ML receiver finds the projection of the received signal $i(t)$ onto the (ψ_0,ψ_1) plane, and makes the decision that minimizes the Euclidean distance between this projection and the decision. A block diagram of this receiver is shown in figure 15. Since ψ_0 and ψ_1 are orthonormal, the ψ_0 and ψ_1 components of the shot noise $i_{sh}(t)$ are independent and identically distributed (i.i.d.). The BER can then be computed as the probability that the noise component is greater than half the Euclidean distance between the *ONE* and *ZERO* in signal space [10]:

$$P_e = Q\left[\frac{\text{distance}/2}{\sigma}\right]$$
$$= Q\left[\frac{2R\sqrt{TP_sP_{LO}}/2}{\sqrt{qRP_{LO}}}\right]$$

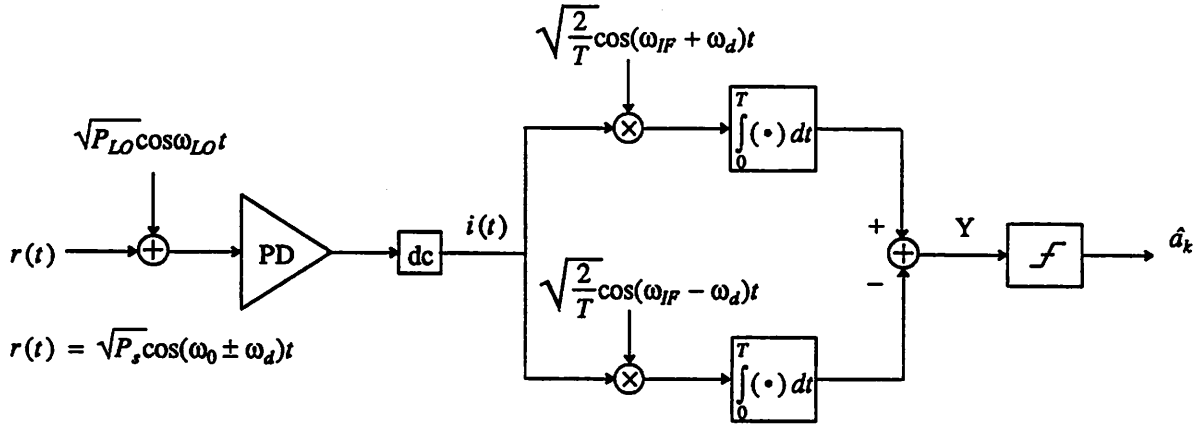

Figure 15. The ML FSK heterodyne receiver finds the correlation of the current signal $i(t)$ with the two expected pulses for a *ONE* and *ZERO*.

$$= Q(\sqrt{M}) \tag{69}$$

This result is plotted in figure 12, where it can be seen that heterodyne synchronous FSK reception is 3 dB worse than the quantum limit. It is also 3 dB worse than heterodyne PSK. This is as expected, since FSK is an orthogonal signaling scheme, while PSK is antipodal. In terms of sensitivity, heterodyne synchronous FSK requires 36 photons per bit for a BER of 10^{-9}. FSK is a popular modulation scheme, however, due to the advantages of direct modulation: by modulating the current of a semiconductor laser, one can achieve FSK directly. PSK and ASK require external modulators, which can introduce significant insertion loss.

Note that a homodyne FSK receiver would require two local lasers, one for each possible frequency. Not only does this make them impractical, but there would be an additional 3 dB penalty in sensitivity due to the inevitable loss of power when the received lightwave is separated into the two receiver branches. Therefore, homodyne FSK will not be considered here.

ASK

The BER for homodyne PSK and heterodyne-synchronous PSK have already been found to be $Q(\sqrt{4M})$, $Q(\sqrt{2M})$, respectively. The same analysis is applicable to ASK, with two adjustments:

- The distance in signal space for ASK is half that for PSK.
- The average signal power for ASK is half that for PSK.

Each of these alone cause a 3 dB degradation in sensitivity, resulting in a total of 6 dB. Therefore, without repeating the analysis, we conclude that for homodyne ASK,

$$P_e = Q(\sqrt{M}), \tag{70}$$

and for heterodyne-synchronous ASK,

$$P_e = Q(\sqrt{\frac{1}{2}M}). \tag{71}$$

These BERs correspond to a 3 and 6 dB penalty with respect to the quantum limit, requiring 36 and 72 photons per bit, respectively. Although these ASK modulation schemes perform worse than the PSK schemes, their use stems from the relative ease in performing external ASK modulation of a laser. An interesting application of homodyne ASK will be considered in section 5.3.

3.2.2. Asynchronous Processing

Although synchronous receivers perform well, they require sophisticated components, such as a PLL. An alternative way of processing the IF current uses an energy or envelope detector, which results in simpler receiver design. Their performance, however, will be seen to be inferior to their synchronous counterparts by ~0.5 dB. An envelope detector can be thought of as an absolute magnitude rectifier followed by a low-pass filter. It will be assumed in this section that an envelope detector exactly replicates the envelope of its input. Since an envelope detector ignores the phase of its input, PSK cannot be processed asynchronously, so only ASK and FSK will be considered in this section.

ASK

In the last section, it was shown that the synchronous ASK receiver found the component of $i(t)$ in the ψ direction using a correlator. In practice, this is implemented using a matched filter, as illustrated in figure 16(b). These synchronous receivers use a PLL to track the phase of the received signal. The motivation behind asynchronous receivers is the desire to detect the signal without regard to the phase, and thus simplifying the receiver design. If the matched filter implementation of figure 16(b) were used without phase tracking (modeled as a uniform random variable θ over $[0,2\pi]$), sampling $p(t)$ at $t=T$ will in general not yield Y. In fact, $p(T)$ could be zero, or worse negative! But even if $p(t)$ were not maximum at $t=T$, its *envelope* will be close to its maximum. Therefore, an asynchronous receiver samples the envelope of $p(t)$, as illustrated in figure 16(c). For the ideal envelope detector, the result r will be independent of θ.

Consider the asynchronous receiver for heterodyne ASK, shown in figure 17. Here, the IF current is

$$i(t) = a_k 2R\sqrt{P_s P_{LO}}\cos(\omega_{IF}t + \theta) + i_{sh}(t) \tag{72}$$

where θ represents the fact that the absolute phase of the pulse is not tracked, so it's an unknown random variable. The matched filter - envelope detector - sampler combination shown in the figure estimates r, the envelope of $i(t)$. Analytically, this is equivalent to projecting

$$\mathbf{i} \leftrightarrow i(t) \tag{73}$$

onto the (ψ_c,ψ_s) plane and finding the distance from the origin, where

$$\psi_c \leftrightarrow \sqrt{\frac{2}{T}}\cos\omega_{IF}t, \tag{74}$$

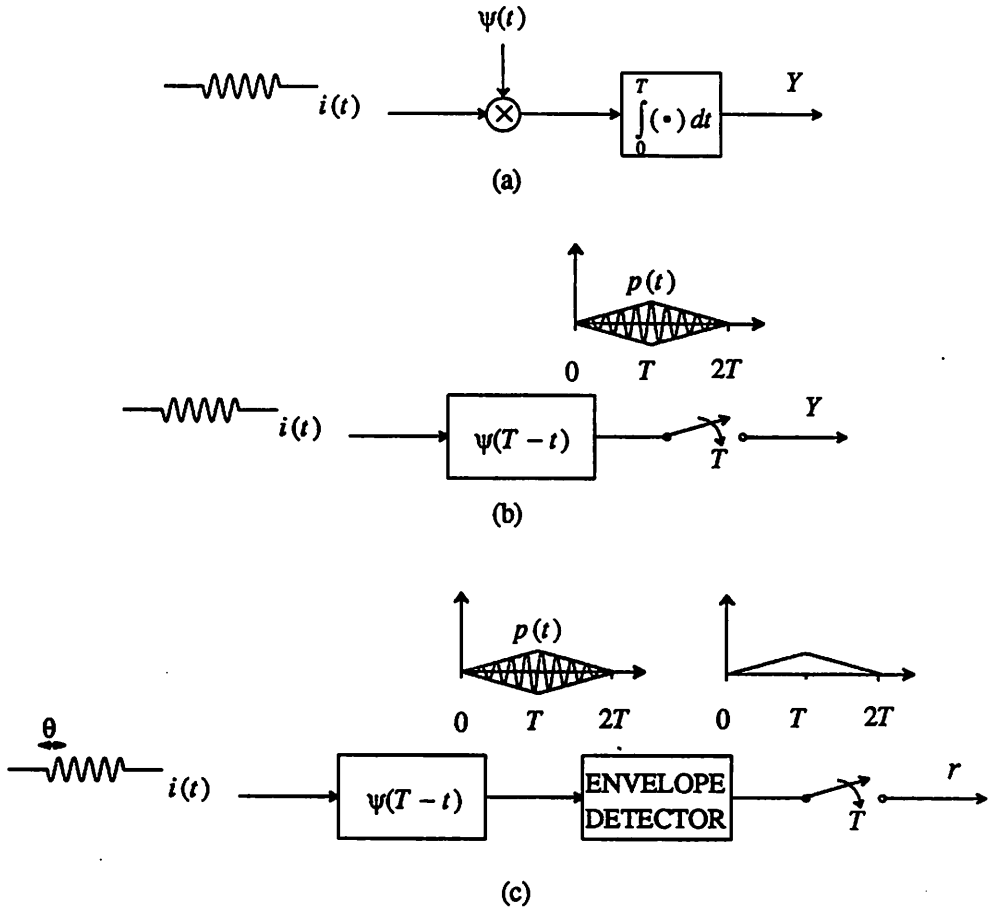

Figure 16. Evolution of asynchronous receivers: (a) the ideal correlator; (b) a matched filter implementation; (c) an asynchronous envelope detector implementation. Adapted from [15].

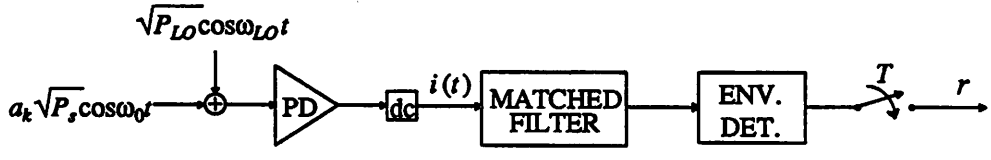

Figure 17. The heterodyne asynchronous ASK receiver uses a filter matched to the expected pulse at IF, followed by an envelope detector.

and

$$\psi_s \leftrightarrow \sqrt{\frac{2}{T}}\sin\omega_{IF}t, \tag{75}$$

Figure 18 shows the signal space representation. It is evident from figure 18 that the enelope r will be independent of θ. So without loss of generality, assume $\theta = 0$.

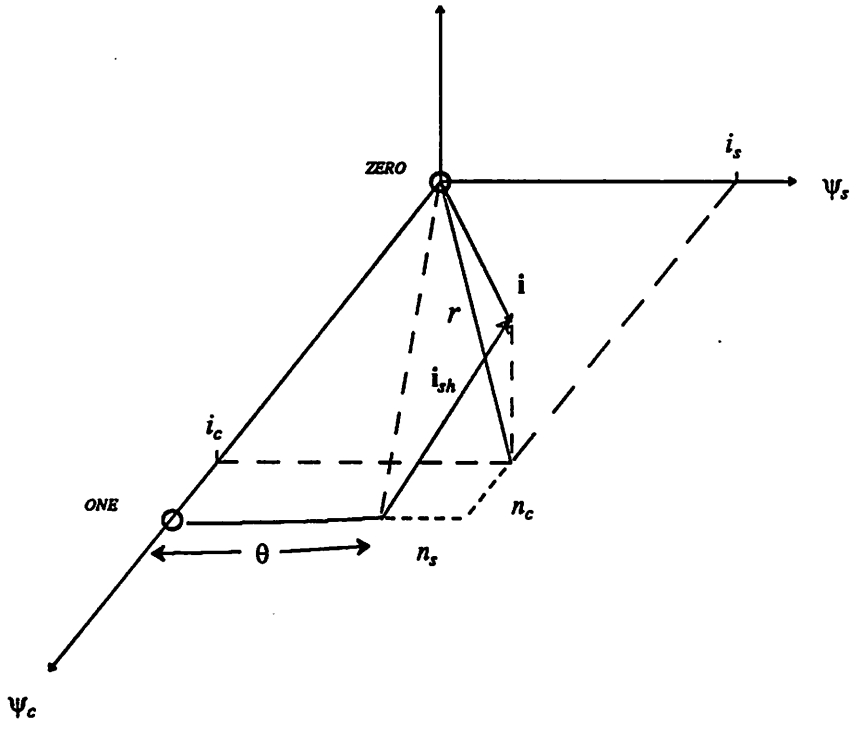

Figure 18. The signal space for ASK: θ represents the unknown phase, and $\mathbf{i}_{sh}$ represents the shot noise. The envelope of $i(t)$ is r, the distance from the origin to the projection of **i** onto the (ψ_c,ψ_s) plane.

Computing the projections of **i** onto the plane axes:

$$i_c = \int_0^T i(t)\psi_c(t)\,dt = a_k R\sqrt{2TP_sP_{LO}} + n_c, \tag{76}$$

and

$$i_s = \int_0^T i(t)\psi_s(t)\,dt = n_s, \tag{77}$$

where n_c and n_s are i.i.d. zero-mean Gaussian random variables with variance qRP_{LO}. To simplify notation, define

$$I = R\sqrt{2TP_sP_{LO}} \tag{78}$$

and

$$\sigma^2 = qRP_{LO} \tag{79}$$

as the normalized current amplitude and noise variance, respectively. Then the envelope r is

$$\begin{aligned} r &= \sqrt{i_c^2 + i_s^2} \\ &= \sqrt{(a_k I + n_c)^2 + n_s^2}. \end{aligned} \tag{80}$$

The probability distribution function for r can be obtained as follows (see, e.g., [15] [16]).

$$\begin{aligned} p_r(\rho)d\rho &= \Pr[\rho < r < \rho + d\rho] \\ &= \Pr[\rho < \sqrt{(a_k I + n_c)^2 + n_s^2} < \rho + d\rho] \\ &= \int\int_{\Delta N} p_{n_c n_s}(n_c - a_k I,\ n_s)\, dn_c dn_s, \end{aligned} \tag{81}$$

where ΔN is a ring of radius ρ and thickness $d\rho$ in the (n_c, n_s) plane, as shown in figure 19.

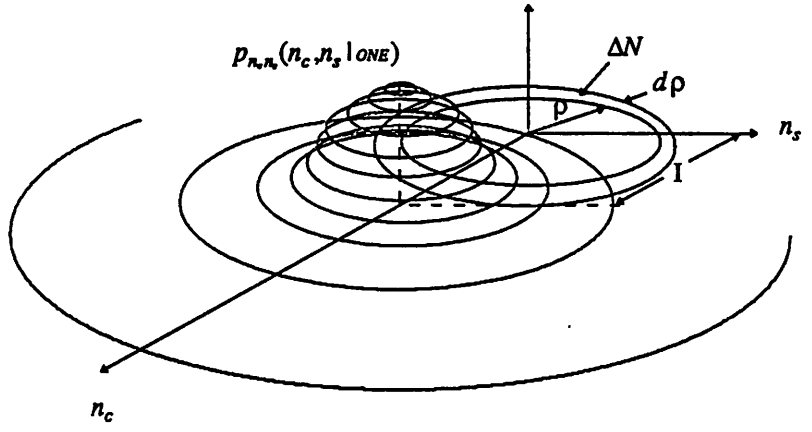

Figure 19. The probability distribution function for r can be found by integrating a two-dimensional jointly Gaussian distribution centered at $(I,0)$ over ΔN.

Since n_c and n_s are independent zero-mean Gaussian random variables with variance σ^2, the joint distribution is the product of the marginal distributions, so that (81) becomes

$$p_r(\rho)d\rho = \int\int_{\Delta N} \frac{1}{2\pi\sigma^2} e^{-(n_c - a_k I)^2/2\sigma^2} e^{-n_s^2/2\sigma^2}\, dn_c dn_s. \tag{82}$$

Substitute

$$\begin{aligned} n_c &= \rho\cos\theta, \\ n_s &= \rho\sin\theta, \\ dn_c dn_s &= \rho d\rho d\theta \end{aligned} \tag{83}$$

to get

$$\begin{aligned} p_r(\rho)d\rho &= \int_{\rho}^{\rho + d\rho} \int_{0}^{2\pi} \frac{1}{2\pi\sigma^2} e^{-(\rho\cos\theta - a_k I)^2/2\sigma^2 - \rho^2\sin^2\theta/2\sigma^2} \rho\, d\theta d\rho \\ &= \rho d\rho \int_0^{2\pi} \frac{1}{2\pi\sigma^2} e^{-(\rho^2 - 2\rho a_k I\cos\theta + a_k I^2)/2\sigma^2} d\theta \\ &= d\rho \frac{\rho}{\sigma^2} e^{-(\rho^2 + a_k I^2)/2\sigma^2} \left[\frac{1}{2\pi} \int_0^{2\pi} e^{a_k \rho I\cos\theta/\sigma^2} d\theta \right]. \end{aligned} \tag{84}$$

Recognizing the bracketed term as the zeroth order modified Bessel function of the first kind, we find that

$$p_r(\rho) = \frac{\rho}{\sigma^2} e^{-(\rho^2 + a_k I^2)/2\sigma^2} I_0\left[\frac{a_k \rho I}{\sigma^2}\right], \qquad (\rho > 0). \tag{85}$$

This result, which holds for the envelope of a narrow-band signal in additive white Gaussian noise, is well known to be the Rician density. When a_k is zero, this reduces to the Rayleigh distribution. The conditional distributions for the envelope are

$$p_r(\rho \mid ZERO) = \frac{\rho}{\sigma^2} e^{-\rho^2/2\sigma^2}, \qquad (\rho > 0), \tag{86}$$

$$p_r(\rho \mid ONE) = \frac{\rho}{\sigma^2} e^{-(\rho^2 + I^2)/2\sigma^2} I_0\left[\frac{\rho I}{\sigma^2}\right], \qquad (\rho > 0)$$

$$\approx \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\rho - I)^2/2\sigma^2}, \tag{87}$$

where the second approximation is valid for large

$$\begin{aligned} \text{SNR} &= \frac{I^2}{\sigma^2} \\ &= \frac{R^2 2TP_s P_{LO}}{qRP_{LO}} \\ &= 2M. \end{aligned} \tag{88}$$

The conditional densities for heterodyne ASK are sketched in figure 20.

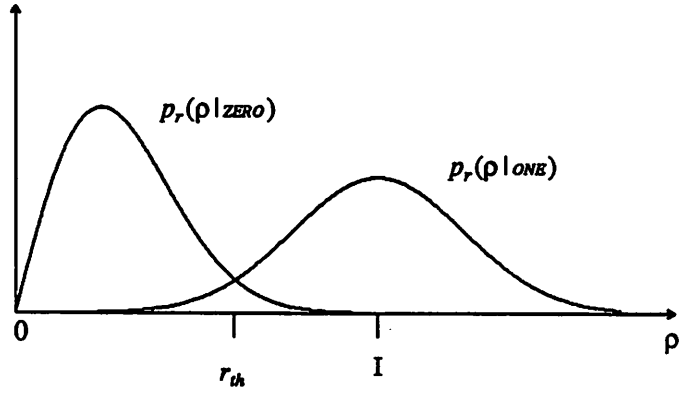

Figure 20. The Rayleigh distribution (*ZERO*) and Rician distribution (*ONE*).

The optimum threshold r_{th} is where $p_r(\rho \mid ZERO) = p_r(\rho \mid ONE)$, which again for high SNR is $r_{th} = I/2$. The BER for heterodyne asynchronous ASK with $M >> 1$ is then

$$P_e = \frac{1}{2} \Pr[r > \frac{I}{2} \mid ZERO] + \frac{1}{2} \Pr[r < \frac{I}{2} \mid ONE]$$

$$= \frac{1}{2}\int_{I/2}^{\infty} \frac{\rho}{\sigma^2} e^{-\rho^2/2\sigma^2} d\rho + \frac{1}{2} Q\left[\frac{I/2}{\sigma}\right]$$

$$= \frac{1}{2} e^{-I^2/8\sigma^2} + \frac{1}{2} Q\left[\frac{I/2}{\sigma}\right]$$

$$\approx \frac{1}{2} e^{-I^2/8\sigma^2}, \tag{89}$$

where the last approximation follows from

$$Q\left[\frac{I/2}{\sigma}\right] < e^{-I^2/4\sigma^2} \ll e^{-I^2/8\sigma^2}. \tag{90}$$

Notice that a false alarm is more probable than a miss. Substituting the original definitions for I and σ from equations (78) and (79), we get

$$P_e = \frac{1}{2} e^{-M/4}. \tag{91}$$

Thus heterodyne asynchronous ASK is 6 dB worse that the quantum limit, as shown in figure (12). The sensitivity is 80 photons per *ONE* bit.

FSK

An FSK modulated lightwave can be demodulated asynchronously using two branches, one for each possible frequency. Consider the receiver shown in figure 21.

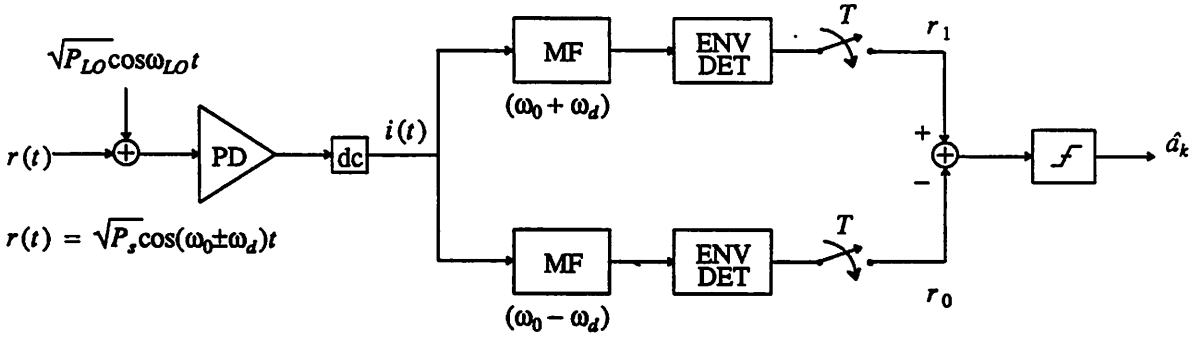

Figure 21. A dual filter heterodyne asynchronous FSK receiver.

The IF current is

$$i(t) = 2R\sqrt{P_s P_{LO}}\cos(\omega_0 \pm \omega_d)t + i_{sh}(t)$$

$$= I\sqrt{\frac{2}{T}}\cos(\omega_0 \pm \omega_d)t + i_{sh}(t), \tag{92}$$

where again,

$$S_{sh}(f) = \sigma^2 = qRP_{LO} \tag{93}$$

is the the shot noise density, and

$$I = R\sqrt{2TP_sP_{LO}} \tag{94}$$

is the normalized current amplitude. The two branch filters are matched to the expected RF pulse at frequencies $\omega_0 \pm \omega_d$. For large frequency deviation ($\omega_0 \gg \omega_d \gg 1/T$), there will be no cross-talk between branches, so the system is symmetric (i.e., the probability of a false alarm equals that of a miss). Assuming a *ONE* is sent, then

$$P_e = \Pr[r_1 - r_0 < 0 \mid ONE], \tag{95}$$

where r_1 and r_0 are the envelopes detected by the two branches. Each branch is seen to be a replica of the ASK heterodyne asynchronous receiver considered earlier (see figure 16). The probability distribution of r_1 and r_0 can thus be found from equation (86) and (87) (see, e.g., [15] [16]):

$$p_{r_1}(\rho_1 \mid ONE) = \frac{\rho_1}{\sigma^2} e^{-(\rho_1^2 + I^2)/2\sigma^2} I_0\left[\frac{\rho_1 I}{\sigma^2}\right] \qquad (\rho > 0), \tag{96}$$

$$p_{r_0}(\rho_0 \mid ONE) = \frac{\rho_0}{\sigma^2} e^{-\rho_0^2/2\sigma^2} \qquad (\rho > 0). \tag{97}$$

The noise in each branch will be independent due to the separate frequency ranges spanned by the two matched filters. Therefore

$$p_{r_1 r_0}(\rho_1, \rho_0 \mid ONE) = p_{r_1}(\rho_1 \mid ONE)\, p_{r_0}(\rho_0 \mid ONE). \tag{98}$$

The BER can now be computed by integrating this joint distribution over the range dictated by equation (95):

$$\begin{aligned} P_e &= \int_0^\infty \int_{\rho_1}^\infty p_{r_1}(\rho_1 \mid ONE)\, p_{r_0}(\rho_0 \mid ONE)\, d\rho_0 d\rho_1 \\ &= \frac{1}{\sigma^2} e^{-I^2/2\sigma^2} \int_0^\infty \rho_1 I_0\left[\frac{\rho_1 I}{\sigma^2}\right] e^{-\rho_1^2/\sigma^2} \left[\int_{\rho_1}^\infty \frac{\rho_0}{\sigma^2} e^{-\rho_0^2/2\sigma^2} d\rho_0\right] d\rho_1 \\ &= \frac{1}{\sigma^2} e^{-I^2/2\sigma^2} \int_0^\infty \rho_1 I_0\left[\frac{\rho_1 I}{\sigma^2}\right] e^{-2\rho_1^2/2\sigma^2} d\rho_1. \end{aligned} \tag{99}$$

Substituting $x = \sqrt{2}\rho_1$, $\alpha = I/\sqrt{2}$, yields

$$= \frac{1}{2} e^{-I^2/4\sigma^2} \int_0^\infty \frac{x}{\sigma^2} I_0\left[\frac{\alpha x}{\sigma^2}\right] e^{-(\alpha^2 + x^2)/2\sigma^2} dx. \tag{100}$$

Recognizing the integrand as the Rician density, the total integral must be one. Thus

$$P_e = \frac{1}{2} e^{-I^2/4\sigma^2}. \tag{101}$$

Substituting I and σ^2 from (93) and(94) yields the BER for heterodyne asynchronous FSK:

$$P_e = \frac{1}{2} e^{-M/2}. \tag{102}$$

This corresponds to a 3 dB penalty with respect to the quantum limit, as illustrated in figure (12). The sensitivity is 40 photons per bit.

3.2.3. Weakly Synchronous Processing: DPSK

The synchronous receivers of section 3.2.1 are known to be optimum for additive white Gaussian channels. They essentially found the correlation of the received lightwave with a local replica of the transmitted pulse. Since this requires accurate phase synchronization which complicates the receiver structure, the asynchronous receivers described in section 3.2.2 are commonly used, with an accompanying penalty of ~0.5 dB. These asynchronous receivers used envelope detectors, which obliterate all knowledge of phase, and thus they are inappropriate for PSK systems. Fortunately, however, there is a way to demodulate a certain type of PSK called *differential phase-shift keying (DPSK)* without replicating the transmitted pulse locally, and without the need for tracking phase. To send a *ZERO* in DPSK, the transmitter changes the phase of the pulse by 180° with respect to the previous pulse. To send a *ONE*, the phase of the transmitted pulse is kept the same as the previous one. That is, the received signal is

$$\sqrt{P_s}\cos(\omega_0 t + \phi_k + \theta) \tag{103}$$

where $\phi_k \in \{0,\pi\}$ represents the modulated phase of the k^{th} bit:

$$\begin{aligned} \textit{ONE} &\leftrightarrow \phi_k - \phi_{k-1} = 0, \\ \textit{ZERO} &\leftrightarrow \phi_k - \phi_{k-1} = \pm\pi. \end{aligned} \tag{104}$$

Here θ represents the unknown absolute phase, which is not tracked. If this unknown phase varies slowly, enough so that it is virtually constant over a time interval of two bit periods, then the pulses in two adjacent slots will differ by either 0 or π. A DPSK receiver uses this knowledge and finds the correlation of the incoming pulse with the previous pulse. Except for the effects of shot noise, they will be the same for a *ONE*, resulting in a positive correlation, and opposite for a *ZERO*, resulting in a negative correlation. The block diagram of such a receiver is shown in figure 22.

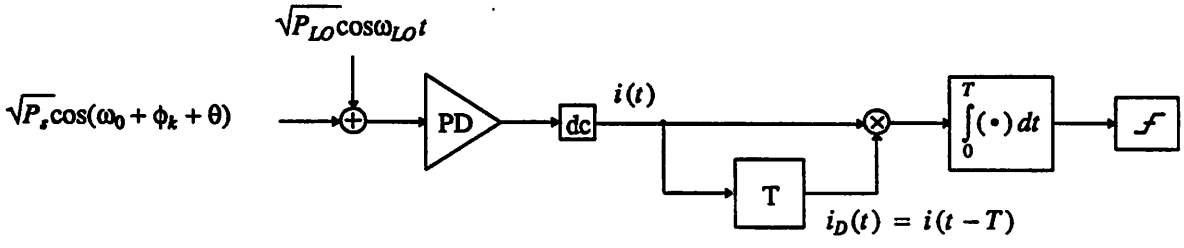

Figure 22. The DPSK receiver uses a one bit-delay demodulator. When the phase of two consecutive pulses are the same, the result is positive, and when the two IF pulses have opposite phase, the result is negative.

The IF current is

$$i(t) = 2R\sqrt{P_s P_{LO}}\cos(\omega_{IF} t + \theta) + i_{sh}(t), \tag{105}$$

where

$$S_{sh}(f) = \sigma^2 \triangleq qRP_{LO}. \tag{106}$$

To analyze the DPSK receiver, it is easiest to use signal space concepts. From equation (104), one can surmise the best receiver will estimate $\beta \triangleq$ the difference in phase between $i(t)$ and $i(t-T)$, and

$$\text{decide } \textit{ONE} \textit{ if } |\beta|_{\text{mod}2\pi} < \pi/2,$$

and

$$\text{decide } \textit{ZERO} \textit{ if } |\beta|_{\text{mod}2\pi} > \pi/2. \tag{107}$$

From equation (105), we see that without the additive shot noise, the set of all possible expected signals falls on the (ψ_c, ψ_s) plane, where

$$\psi_c \leftrightarrow \sqrt{\frac{2}{T}}\cos\omega_{IF}t,$$

and

$$\psi_s \leftrightarrow \sqrt{\frac{2}{T}}\sin\omega_{IF}t. \tag{108}$$

And since the shot noise is white, its (ψ_c, ψ_s) components are uncorrelated with any of its other components. Therefore, no information is lost if the currents $i(t)$ and $i(t-T)$ are restricted to the *sufficient subspace* spanned by ψ_c and ψ_s. In this signal subspace, β is the angle between the projections of $\mathbf{i}$ and $\mathbf{i}_D$ onto the (ψ_c, ψ_s) plane, where

$$\mathbf{i} \leftrightarrow i(t),$$

and

$$\mathbf{i}_D \leftrightarrow i(t-T). \tag{109}$$

Also define

$$\begin{aligned} \mathbf{s} &\leftrightarrow I\cos(\omega_{IF}t + \theta), \\ \mathbf{n} &\leftrightarrow i_{sh}(t), \\ \mathbf{n}_D &\leftrightarrow i_{sh}(t-T), \end{aligned} \tag{110}$$

where $\mathbf{s}$ represents the expected signal, and again

$$I = R\sqrt{2TP_sP_{LO}}. \tag{111}$$

Notice that since $\mathbf{n}$ and $\mathbf{n}_D$ are sample functions of a white noise process over two disjoint time intervals, they are independent. Furthermore, since

$$\langle\psi_c,\psi_s\rangle \leftrightarrow \int_0^T \psi_c(t)\psi_s(t)\,dt = 0, \tag{112}$$

then

$$\begin{aligned} n_c &\triangleq \langle\mathbf{n},\psi_c\rangle, \\ n_s &\triangleq \langle\mathbf{n},\psi_s\rangle, \\ n_{D,c} &\triangleq \langle\mathbf{n}_D,\psi_c\rangle, \end{aligned}$$

$$n_{D,s} \triangleq \langle \mathbf{n}_D, \psi_s \rangle \tag{113}$$

are all i.i.d. zero-mean Gaussian random variables with variance $\sigma^2 = qRP_{LO}$. With these definitions, the signal space diagram is shown in figure 23, where only the projections of the vectors onto the (ψ_c, ψ_s) plane are illustrated.

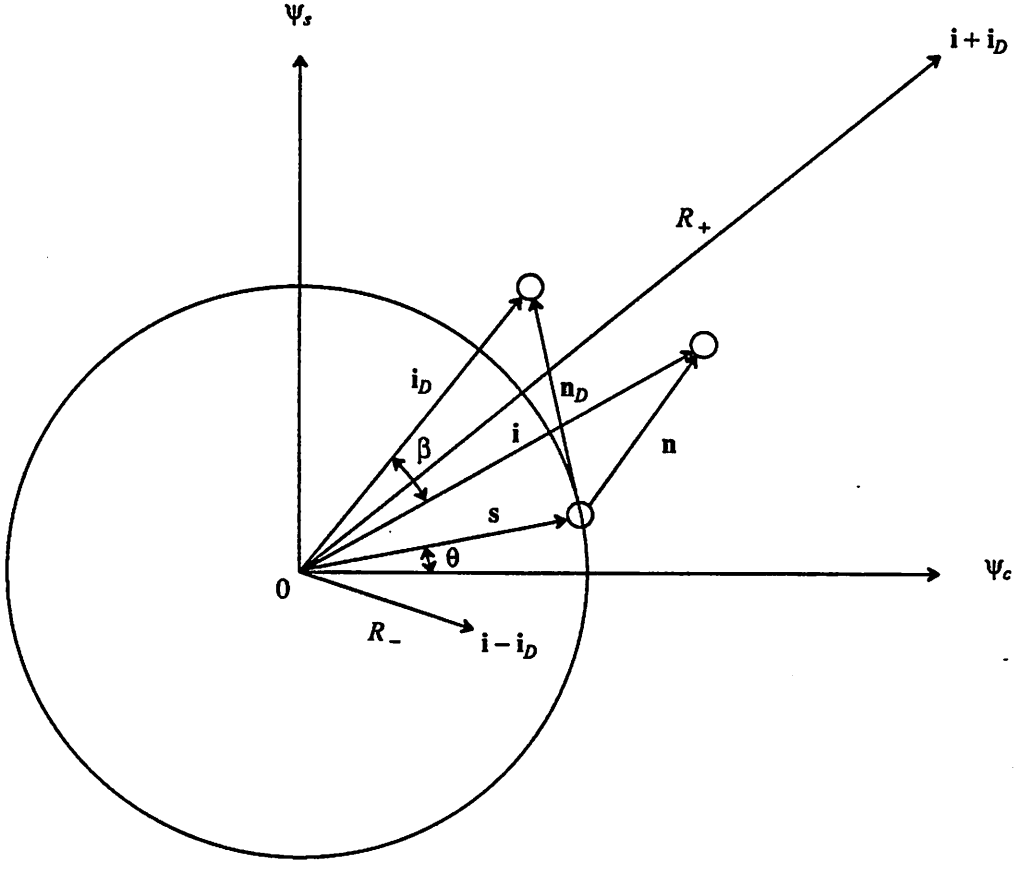

Figure 23. Signal space representation for DPSK. The absolute phase θ is irrelevant, only the relative phase β between **i** and $\mathbf{i}_D$ is important.

Since DPSK is a symmetric modulation scheme [12], the BER can be found by considering only the probability of a miss:

$$\begin{aligned} P_e &= \Pr[\text{miss}] \\ &= \Pr[|\beta|_{\text{mod}2\pi} > \pi/2 \mid ONE] \\ &= \Pr[\cos\beta < 0 \mid ONE] \\ &= \Pr\left[\frac{\langle \mathbf{i}, \mathbf{i}_d \rangle}{\|\mathbf{i}\| \cdot \|\mathbf{i}_D\|} < 0 \mid ONE\right] \end{aligned} \tag{114}$$

where the last equality follows from $\langle \mathbf{i}, \mathbf{i}_d \rangle = \|\mathbf{i}\| \cdot \|\mathbf{i}_D\| \cos\beta$. Next use the identity

$$4\langle \mathbf{i}, \mathbf{i}_D \rangle = \|\mathbf{i} + \mathbf{i}_D\|^2 - \|\mathbf{i} - \mathbf{i}_D\|^2 \tag{115}$$

to get:

$$P_e = \Pr[\|\mathbf{i}+\mathbf{i}_D\|^2 - \|\mathbf{i}-\mathbf{i}_D\|^2 < 0 \mid ONE] = \Pr[R_+ < R_- \mid ONE], \quad (116)$$

where

$$R_+ \triangleq \|\mathbf{i}+\mathbf{i}_D\| = \|2\mathbf{s}+\mathbf{n}+\mathbf{n}_D\|, \quad (117)$$

and

$$R_- \triangleq \|\mathbf{i}-\mathbf{i}_D\| = \|\mathbf{n}-\mathbf{n}_D\|, \quad (118)$$

That is, R_+ and R_- are the lengths of two-dimensional vectors with mean lengths $2I$ and 0, respectively, corrupted by additive white Gaussian noise. From the analysis of heterodyne asynchronous ASK reception in section 3.2.2, we know that R_+ and R_- have a Rician distribution. Now, however, the mean and variance are $2I$ and $2\sigma^2$, instead of I and σ^2. From equations (86) and (87) we get the conditional distributions for R_+ and R_-:

$$p_{R_+}(\rho \mid ONE) = \frac{\rho}{(2\sigma^2)} e^{-\rho^2/2(2\sigma^2)}, \qquad (\rho > 0), \quad (119)$$

$$p_{R_-}(\rho \mid ONE) = \frac{\rho}{(2\sigma^2)} e^{-(\rho+(2I))^2/2(2\sigma^2)} I_0\left[\frac{\rho(2I)}{(2\sigma^2)}\right], \qquad (\rho > 0). \quad (120)$$

The evaluation of (116) for Rician distributions was already carried out in section 3.2.2, for heterodyne asynchronous FSK. The result can be directly applied here, by replacing I by $2I$, and σ^2 by $2\sigma^2$ in equation (101). Thus for DPSK reception, the BER under shot-noise-limited conditions is

$$P_e = \frac{1}{2} e^{-(2I)^2/4(2\sigma^2)} = \frac{1}{2} e^{-I^2/2\sigma^2} = \frac{1}{2} e^{-M}, \quad (121)$$

where I and σ^2 were substituted from equations (111) and (113). This BER represents a ~0.5 dB degradation with respect to heterodyne-synchronous PSK, equaling the performance of the quantum limit, at the gain of simpler receiver structure. This means that a DPSK receiver needs only 20 photons per bit for a BER of 10^{-9}. As one might expect, however, both PSK and DPSK are sensitive to phase noise, which turns out to be a significant impairment for semiconductor lasers (see section 4.8).

4. Unique Characteristics of Optical Communication Systems

In many ways, communicating with the visible portion of the electromagnetic spectrum is the same as using RF frequencies. Enough differences exist, however, so that one cannot always directly apply proven principles from one realm to the other. For example, the bandwidth of RF systems is generally limited, whereas for optical systems it is

plentiful. And a virtually noise-free RF oscillator is relatively easy to build, whereas the best of today's semiconductor lasers have significant phase noise. So although lightwave communications overcomes many of the limitations in RF communications, it introduces a new set of impairments. This section describes some of these unique traits.

4.1. Narrow-band

An optical pulse is very narrow-band, since the frequency of its carrier is large. As an example, consider an isolated pulse of a 100 Mbps system, with wavelength $\lambda = 1300nm$. Then the ratio of the bandwidth of this pulse to its carrier is $\sim 10^{-6}$, whereas the same square pulse with a carrier of only 1 GHz would have a ratio of $\sim 10^{-2}$. Also, the bandwidth of optical fiber is immense. Furthermore, today's high quality fiber has a very small dispersion coefficient. Together, these facts make the dispersion of an optical pulse relatively small, so it is commonly ignored in the analysis of optical receivers. This precludes the need for equalization to eliminate inter-symbol interference. Consider next some of the components of an optical communications system.

4.2. Optical Fiber

Multimode fibers have a large core with respect to the wavelength of the propagating wave. When a pulse of light is applied to one end of a fiber, the light can be thought of as traveling down the fiber via total reflection. When the core radius is large enough, many different modes reflecting at different angles can propagate. Since the resulting path traveled by the lower modes (small angle of reflection) is shorter than that of the higher order modes (large angle of reflection), the lower order modes will arrive at the receiver before the higher modes, resulting in a spreading of the pulse. This type of distortion is called *inter-modal dispersion,* and can be overcome by making the core radius small enough so that only a single mode can propagate, namely a single-mode fiber. For this reason, single-mode fibers are expected to prevail in high-throughput communication applications, especially coherent systems. Because of the small core radius, however, it is difficult to launch light into a single-mode fiber.

Another form of degradation is *chromatic dispersion,* which arises from the fact that lightwaves at different frequencies travel through the dielectric waveguide at different speeds. The wavelength of zero chromatic dispersion for standard fiber is near 1300 *nm*, but it can be shifted anywhere from 1300–1500 *nm* through proper fiber design.

The attenuation of high-quality production fibers is $\lesssim 0.2$ dB/km, which is quite favorable compared to the ~ 10 dB/km loss for metallic coaxial cables. And since this low attenuation exists in the 1450–1650 nm wavelength range, optical fiber offers a potential bandwidth of over 30,000 GHz. Other attributes of fiber as a medium are its immunity to interference, absence of energy radiation (which provides high security), and reduced overall effective cost. Together, these characteristics make optical fiber a nearly ideal communications medium.

4.3. Lasers

A crucial element of coherent communication systems is the laser. Ideally, it should be inexpensive, rugged, spectrally pure (small $\Delta\nu$), and tunable over the entire low attenuation window. Gas lasers can have small linewidths, but they are bulky, expensive, and sensitive to vibrations. The future of coherent systems is thought to rely heavily on

advancements in semiconductor laser technology. Presently, semiconductor lasers need external cavities to get $\Delta\nu < 1$ MHz. But these external optical devices lack robustness, so it is hoped the properties of semiconductor lasers can be improved through other means (such as distributed feedback, or DFB). The best of today's DFB lasers have a linewidth $\Delta\nu$ of no less than 5 MHz.

4.4. Modulators

There are basically two ways of modulating a laser:

Direct modulation:

Where FSK can be obtained by controlling the injection current of the semiconductor laser.

External modulation:

Where the lightwave is modulated via a phase modulator, usually a lithium niobate device, to achieve ASK, PSK, or FSK.

Attempting to perform ASK or PSK via direct modulation results in spurious *chirp* noise, and so is not commonly attempted. Externally modulated FSK seems to be possible, although there is an extra requirement of some form of line coding. To investigate this idea further, consider the phase modulator modeled in figure 24.

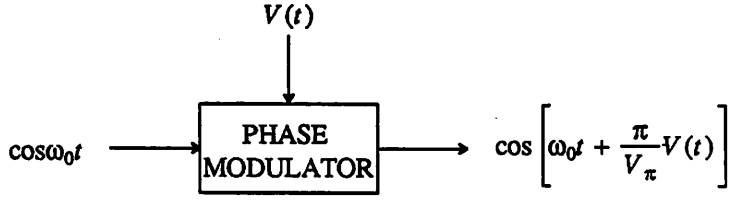

Figure 24. The phase of the output of an ideal phase modulator is proportional to the modulating voltage input $V(t)$.

V_π is the voltage necessary for a phase change of π radians. To perform FSK, $V(t)$ would need to be linearly increasing or decreasing. Neglecting coding considerations, let's say we want to transmit a 1010 pattern. This will provide a quick check of the viability of this device in an FSK system. Then $V(t)$ would need to be triangular waveform, as shown in figure 25:

$$V(t) = \pm\frac{V_{pp}}{T}t + \phi_k \tag{122}$$

where ϕ_k is a constant during each bit interval T. It maintains $V(t)$ as a piecewise continuous function. Then the output of the phase modulator is

$$\begin{aligned}\cos[\omega_0 t + \frac{\pi}{V_\pi}V(t)] &= \cos\left[\omega_0 t \pm \frac{\pi V_{pp}}{V_\pi T}t + \frac{\pi}{V_\pi}\phi_k\right] \\ &= \cos\left[2\pi\left[f_0 \pm \frac{V_{pp}}{2V_\pi T}\right]t + \frac{\pi}{V_\pi}\phi_k\right].\end{aligned} \tag{123}$$

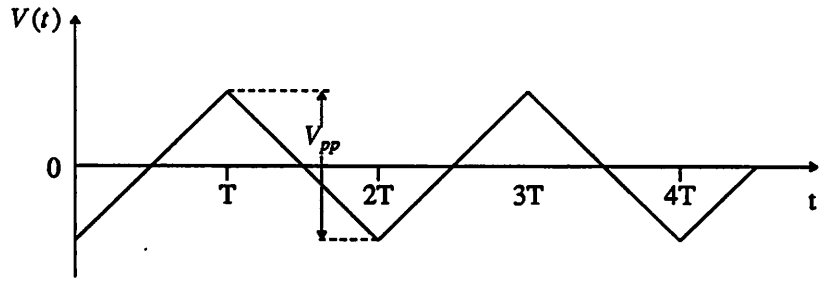

Figure 25. Modulating voltage $V(t)$ necessary for a 1010 FSK pattern.

Notice that the since $V(t)$ is continuous, the output of the phase modulator has continuous phase. From the previous equation, we see that the effective frequency deviation is

$$f_d = \frac{V_{pp}}{2V_\pi T}. \tag{124}$$

A typical value for V_π is 10 volts [17]. Since $\frac{1}{T} = R$ = bit rate, we can rewrite

$$f_d = \frac{V_{pp}}{20} R, \tag{125}$$

which indicates a reasonable range for V_{pp}. For example, for MSK ($f_d = R/4$), $V_{pp} = 5$ volts. But what about the phase modulator, can it track a triangle wave at a frequency of $\frac{1}{2T} = \frac{1}{2}R$? Typical lithium niobate phase modulators have a bandwidth of 8-13 GHz [17]. Thus, as long as $\frac{1}{2}R \ll$ 8-13 GHz, the triangle wave will be sufficiently tracked.

The preceding calculations seem to assert that FSK using an external phase modulator is practical. However, it has neglected an important aspect of phase modulators: their modulating input voltage range. To ensure the linear phase relationship depicted in figure 24 is valid, $V(t)$ must be maintained within specific limits, typically ± 50 volts. One solution would be to make a discrete jump by some multiple of $2V_\pi$. But this would not be tracked adequately by the lithium niobate modulator, resulting in a distorted signal. An alternative is to code the bits before modulation, ensuring that $V(t)$ stays within $\pm V_{max}$. A simple (and inefficient) code is a 1B2B code, illustrated in figure 26. Transmitting $v_1(t)$ for a *ONE* will appear as a *ONE–ZERO* sequence to the FSK receiver. Similarly, applying $v_0(t)$ to the phase modulator would appear as a *ZERO–ONE* sequence. Although this is a practical system, the effective bit rate has been cut in half. However, further study may prove the advantages of external modulation outweigh the coding limitations. A simple FSK system using an external FSK modulator is shown in figure 27.

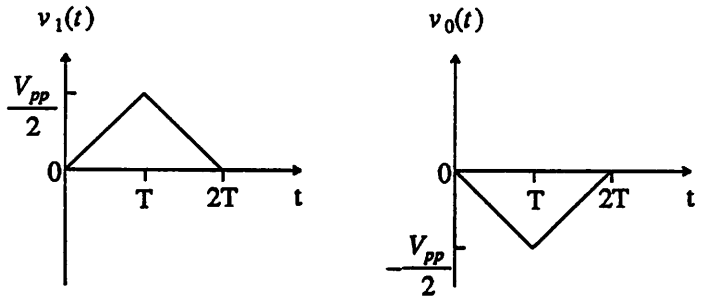

Figure 26. Pulse shapes for $V(t)$: (a) *ONE–ZERO* sequence, (b) *ZERO–ONE* sequence.

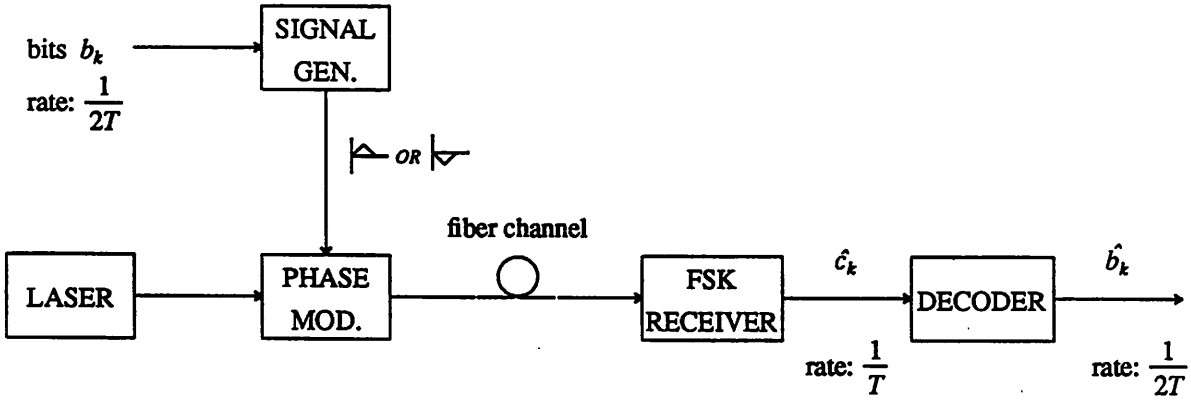

Figure 27. An external modulator in an FSK system.

4.5. Photodetectors

There are two classes of photodetectors:

- PIN diodes: Quantum efficiency $\eta \approx 1$, mean gain $<M> = 1$, less sensitive than avalanche photodiodes.
- Avalanche Photodiodes (APD): Quantum efficiency $\eta \sim 0.5$, mean gain $<M> \sim 50$, more sensitive than PIN diodes, but gain is noisy.

APDs are used for IM/DD systems, because the incident light is generally weak, requiring the amplification effect provided by the APD. A price is paid for the APD's increased sensitivity; multiplicative noise. The gain M has statistical fluctuations that vary as $<M>^x$, where x is typically 0.5. Since heterodyne techniques amplify the light before the photodetector, the gain produced by an APD is not necessary; the light incident in a coherent system is strong enough so that the more efficient and less noisy PIN diode can be used.

4.6. Polarization

An important consideration in coherent optical systems is polarization. Since the state of polarization of a lightwave traveling through a fiber will vary randomly, some means of recovering this polarization is necessary. In some ways, the problem is similar to a random phase variation. And a solution analogous to the PLL would be some sort of polarization tracking system. It has been shown that the rate of polarization fluctuation is small enough so that some of form of feedback tracking should be possible [2]. Another alternative is to break the received lightwave into two orthogonal states of polarization, then process each in a separate branch of the receiver. The performance of the resulting receiver would then in theory be independent of the state of polarization of the received signal [18].

4.7. Heterodyne versus Homodyne

Many of the principles developed in standard RF communication theory can be applied directly to optical communications. For example, the criteria used to choose modulation-demodulation schemes in optical communications are in many cases directly analogous to those developed in RF communications. One interesting difference, however, is this: optical heterodyne reception is 3 dB less sensitive than optical homodyne. There is no such distinction in RF communications. The reason for this difference is not immediately obvious, and to the author's knowledge, there has been no straightforward explanation given in the literature. In fact, one article states [19]

> "Heterodyne reception entails a penalty, however, because the carrier and local oscillator constantly slip out of phase with each other. The receiver is most sensitive at the instant when the signal and local oscillator are in phase. When they are out of phase by 90 degrees, sensitivity approaches zero. The IF signal averages those good and bad conditions, making a heterodyne receiver at least 3 dB less sensitive than homodyne."

Although the argument may be intuitively pleasing, it fails to explain why a 3 dB penalty does *not* occur for RF heterodyne. The inspiration for the discussion that follows comes from a paper by Kazovsky [20], which is dedicated to the heterodyne/homodyne distinction. The purpose of this discussion is to not only verify a 3 dB penalty for optical heterodyne, but to make a direct comparison with microwave systems. What follows is a side by side comparison of optical and RF heterodyne/homodyne techniques.

Optical Homodyne

Consider first optical homodyne reception of a PSK signal, under shot-noise-limited conditions. The ideal receiver is depicted in figure 28(a). The baseband current after dc elimination is

$$i(t) = a_k 2R\sqrt{P_s P_{LO}} + i_{sh}(t), \tag{126}$$

where $S_{sh}(f) = qRP_{LO}$. The output of the integrator is then

$$Y = a_k 2R\sqrt{TP_s P_{LO}} + N, \tag{127}$$

where N is a zero-mean Gaussian random variable with variance qRP_{LO}. The SNR at the input to the slicer is thus

$$\mathrm{SNR}_{opt,hom} = \frac{4R^2 TP_s P_{LO}}{qRP_{LO}}$$

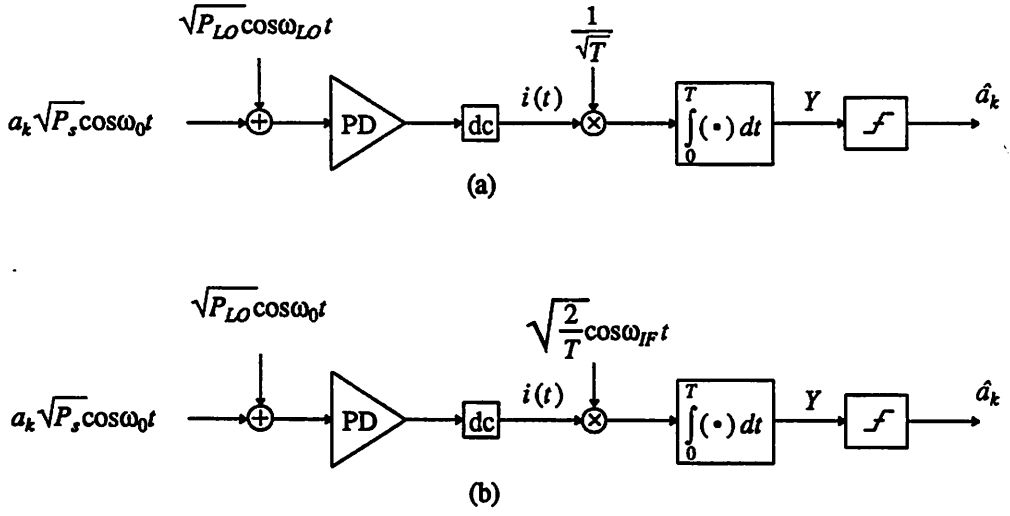

Figure 28. The optimum receiver configurations for optical (a) homodyne, and (b) heterodyne.

$$= 4M. \tag{128}$$

Optical Heterodyne

Next, consider the same PSK signal and a *heterodyne* receiver, as shown in figure 28(b). Here,

$$i(t) = a_k 2R\sqrt{P_s P_{LO}}\cos\omega_{IF}t + i_{sh}(t), \tag{129}$$

where again $S_{sh}(f) = qRP_{LO}$. The basis function for this signal is

$$\sqrt{\frac{2}{T}}\cos\omega_{IF}t, \tag{130}$$

instead of $\frac{1}{\sqrt{T}}$. The resulting correlation Y is then

$$Y = a_k R\sqrt{2TP_s P_{LO}} + N, \tag{131}$$

where again N is a zero-mean Gaussian random variable with variance qRP_{LO}. Thus, the SNR at the slicer input for ideal shot-noise-limited heterodyne PSK is

$$\mathrm{SNR}_{opt,het} = \frac{R^2 2TP_s P_{LO}}{qRP_{LO}}$$

$$= 2M. \tag{132}$$

Comparing (128) and (132), we see that indeed, even for ideal noiseless lasers under shot-noise-limited conditions,

$$\mathrm{SNR}_{opt,het} = \frac{1}{2}\mathrm{SNR}_{opt,hom}. \tag{133}$$

Since the receivers used in the above analysis are optimum, no different or further processing can improve the relationship given by (133).

RF Homodyne.

To clarify the reasons why there is no heterodyne penalty in RF communications, consider the front-end of an RF homodyne PSK receiver shown in figure 29(a).

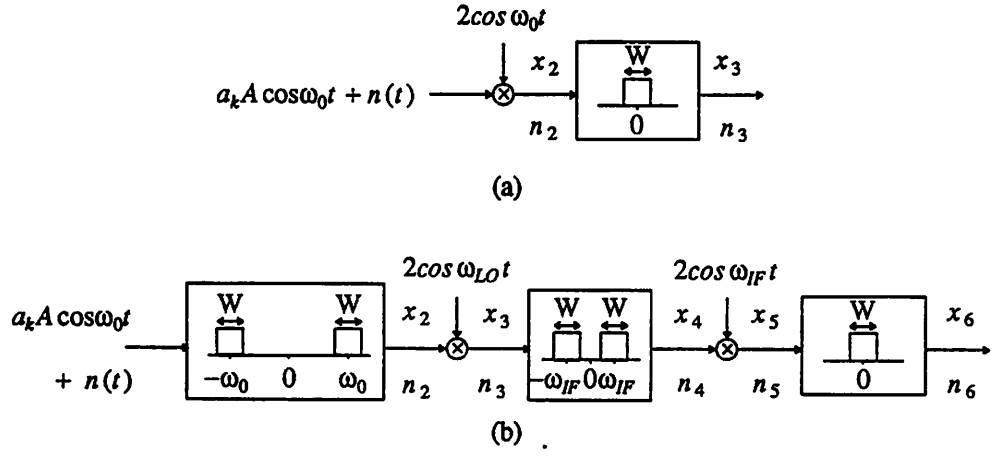

Figure 29. Front-end processing for RF receivers: (a) homodyne, and (b) heterodyne.

The only noise under consideration is $n(t)$, a channel-induced zero-mean white Gaussian process, with two-sided density N_0. The bandwidth of the low-pass filter W is chosen to be large enough to pass the signal undistorted. Computing $x_2(t)$, the output of the mixer:

$$x_2(t) = a_k A + a_k A \cos 2\omega_0 t + n_2(t), \tag{134}$$

where

$$n_2(t) = 2n(t)\cos\omega_0 t \tag{135}$$

is also a zero-mean white Gaussian noise, with density $S_{n_2}(f) = 2N_0$. This result is derived in appendix A. After the low-pass filter, which filters out the high frequency term:

$$x_3(t) = a_k A + n_3(t), \tag{136}$$

where

$$S_{n_3}(f) = 2N_0 rect(f, W/2) = \begin{cases} 2N_0 & for\ |f| < W/2 \\ 0 & for\ |f| > W/2 \end{cases} \tag{137}$$

At this point in the receiver, the SNR is

$$\mathrm{SNR}_{RF,hom} = \frac{A^2}{2N_0 W}. \tag{138}$$

RF Heterodyne.

Finally, consider an RF heterodyne receiver, with the same PSK input, as shown in figure 29(b). Since the noise is introduced in the channel, it is advantageous to pre-filter the received signal with a band-pass filter before demodulation. This would not have helped the homodyne receiver, but the spectral densities shown in figure 30 illustrate its benefit for the heterodyne case. Assuming the band-pass filter passes the expected signal undistorted, its output is

$$x_2(t) = a_k A \cos\omega_0 t + n_2(t), \tag{139}$$

where $S_{n_2}(f)$, along with $S_{n_3}(f)$, $S_{n_4}(f)$, $S_{n_5}(f)$, and $S_{n_6}(f)$, is illustrated in figure 30. After the first mixer,

$$x_3(t) = a_k A \cos\omega_{IF} t + a_k A \cos(\omega_0 + \omega_{LO})t + n_3(t). \tag{140}$$

After the second band-pass filter, which is centered in frequency at ω_{IF},

$$x_4(t) = a_k A \cos\omega_{IF} t + n_4(t). \tag{141}$$

The second mixer results in

$$x_5(t) = a_k A + a_k A \cos(2\omega_{IF})t + n_5(t). \tag{142}$$

After the final low-pass filter, the output is

$$x_6(t) = a_k A + n_6(t). \tag{143}$$

Comparing this result with (136), we find that the RF heterodyne receiver results in the same signal-plus-noise. The corresponding SNR is

$$\begin{aligned} \mathrm{SNR}_{RF,het} &= \frac{A^2}{2N_0 W} \\ &= \mathrm{SNR}_{RF,hom} \end{aligned} \tag{144}$$

which verifies the claim that there is no heterodyne penalty in RF systems.

In case the above analysis is not transparent, here is an explanation in words: Demodulation of a modulated waveform requires a frequency translation of the signal's spectrum from the carrier down to its baseband form, at zero frequency. This frequency translation also applies to the surrounding noise. In homodyne, the translation is done in one step, from f_c to zero. In heterodyne, however, there is an intermediate step: first from f_0 to f_{IF}, then from f_{IF} down to zero. If the noise is white, each translation doubles the relative noise density, making heterodyne 3 dB worse than homodyne. The key difference between RF and optical communications is the source of the white noise. In optical systems, it's from shot noise, which is a result of the demodulation process, while for RF systems, the noise is generated in the channel (or at least before the mixer). This latter fact allows the RF heterodyne receiver to pre-filter the noise before mixing, preventing the noise density from doubling after both translations. This is not possible in optical systems, because the noise is generated *after* the mixing, and thus the 3 dB penalty for optical heterodyne.

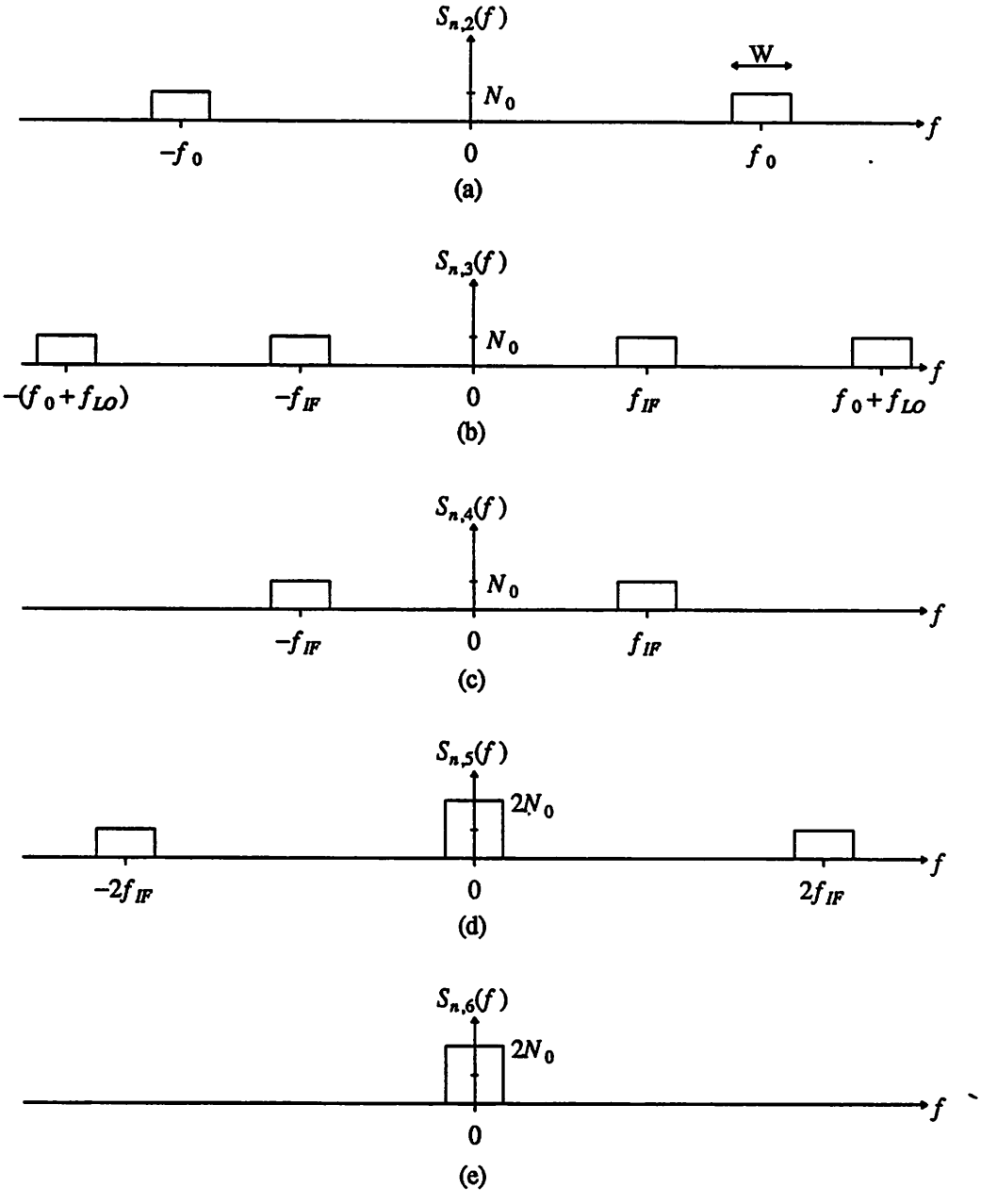

Figure 30. Spectral densities for the noises at various points of the RF heterodyne receiver of figure 29(b). Notice that without the pre-filter, $S_{n,3}(f)$ would have height $2N_0$ instead of N_0.

4.8. Laser Phase Noise

Of all the obstacles preventing coherent reception techniques from making a smooth transition into the optical domain, laser phase noise is the most critical. Ideally, the spectrum of a sinusoidal signal is a delta function centered at the carrier frequency. When the phase of the sinusoid is a random process, the spectrum spreads out so that it has nonzero width. For modern RF oscillators, the width of its spectrum is < 1 Hz. For

typical semiconductor lasers, however, this width is 5–100 MHz. And even though the carrier frequency of an optical oscillator is ~ 10^5 times that of an RF oscillator, the relative quality is still deficient by a factor of ~ 10^3.

To deal with this phase noise analytically, it must be accurately modeled. Define the *linewidth* $\Delta\nu$ of a laser as the 3 dB bandwidth of its spectrum (see figure 31). The laser phase noise process $\phi(t)$ is commonly characterized by a Wiener process, such that its time derivative $\dot{\phi}(t)$ is a zero-mean white Gaussian process with density

$$S_{\dot{\phi}}(f) = 2\pi\Delta\nu. \tag{145}$$

At low frequencies, the density of $\dot{\phi}(t)$ is not flat, due to $1/f$ noise, but it is expected that the resulting variations in frequency are slow enough to be adequately tracked [1]. There is also a peak in $S_{\dot{\phi}}(f)$ at the relaxation frequency of the laser, but this is usually beyond the bandwidth of the receiver. Thus the approximation of $\dot{\phi}$(t) as a white process is a good one. As will be shown, it leads to a spectral shape called *Lorentzian* that has been observed experimentally [21].

It is not obvious that choosing $\dot{\phi}$(t) to be white with density $2\pi\Delta\nu$ results in a sinusoidal signal whose spectrum has bandwidth $\Delta\nu$. To show this, consider the representation of a laser source as

$$x(t) = \sqrt{P_s}\cos[\omega_0 t + \phi(t) + \theta]. \tag{146}$$

The random variable θ is uniform over $[0, 2\pi]$, and independent of $\phi(t)$. It is introduced to to make $x(t)$ stationary, and can be justified by realizing that the time origin is arbitrary, not absolute. That is, since

$$\phi(t) \triangleq \int_0^t \dot{\phi}(u)du, \tag{147}$$

$x(t = 0) = \sqrt{P_s}\cos\theta$ should still be random. To find the spectral density of $x(t)$, its autocorrelation is needed:

$$\begin{aligned} R_x(t_1,t_2) &= E[x(t_1)x(t_2)] \\ &= P_s E[\cos(\omega_0 t_1 + \phi(t_1) + \theta)\cos(\omega_0 t_2 + \phi(t_2) + \theta)] \\ &= \frac{1}{2}P_s E[\cos(\omega_0(t_1 - t_2) + \phi(t_1) - \phi(t_2))] \\ &\quad + \frac{1}{2}P_s E[\sin(\omega_0(t_1 + t_2) + \phi(t_1) + \phi(t_2) + 2\theta)]. \end{aligned} \tag{148}$$

The second term in (148) is zero when the expectation over θ is taken. Thus

$$\begin{aligned} R_x(t_1,t_2) &= \frac{1}{2}P_s E[\cos\omega_0(t_1 - t_2) + \Phi] \\ &= \frac{1}{2}P_s\cos\omega_0(t_1 - t_2)E[\cos\Phi] - \frac{1}{2}P_s\sin\omega_0(t_1 - t_2)E[\sin\Phi], \end{aligned} \tag{149}$$

where for a given t_1 and t_2,

$$\begin{aligned} \Phi &\triangleq \phi(t_1) - \phi(t_2) \\ &= \int_0^{t_1}\dot{\phi}(u)du - \int_0^{t_2}\dot{\phi}(u)du \end{aligned}$$

$$= \int_{t_2}^{t_1} \dot{\phi}(u)du. \tag{150}$$

Therefore, since $\dot{\phi}(t)$ is a zero-mean white Gaussian process, Φ is a zero-mean Gaussian random variable with variance

$$\begin{aligned}
\sigma_\Phi^2 &= E[\Phi^2] \\
&= E[\int_{t_2}^{t_1}\int_{t_2}^{t_1} \dot{\phi}(u)\dot{\phi}(v)\, dudv\ ] \\
&= 2\pi\Delta\nu \int_{t_2}^{t_1}\int_{t_2}^{t_1} \delta(u-v)\, dudv \\
&= 2\pi\Delta\nu\, |t_1 - t_2|.
\end{aligned} \tag{151}$$

It is shown in appendix B that for a zero-mean Gaussian random variable Φ with variance σ_Φ^2,

$$\begin{aligned}
E[\cos\Phi] &= e^{-\sigma_\Phi^2/2}, \\
E[\sin\Phi] &= 0.
\end{aligned} \tag{152}$$

Thus (149) becomes

$$R_x(t_1,t_2) = \frac{P_s}{2} e^{-\pi\Delta\nu|t_1-t_2|} \cos\omega_0(t_1 - t_2), \tag{153}$$

and replacing $t_1 - t_2$ by τ gives

$$R_x(\tau) = \frac{P_s}{2} e^{-\pi\Delta\nu|\tau|} \cos\omega_0\tau. \tag{154}$$

The Fourier transform of this autocorrelation function yields the spectral density of a laser source corrupted by phase noise:

$$S_x(f) = \frac{\frac{P_s}{2\pi\Delta\nu}}{1 + \left[\frac{f \pm f_0}{\Delta\nu/2}\right]^2}. \tag{155}$$

This spectral shape is referred to as *Lorentzian,* and is sketched in figure 31.

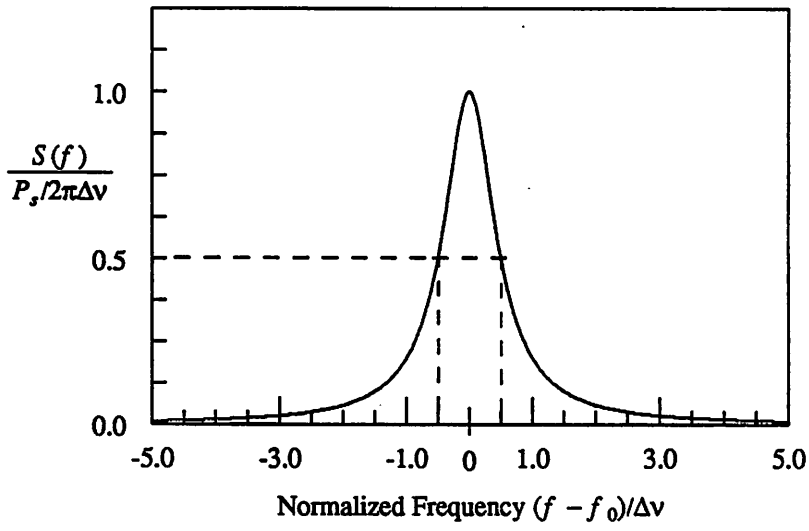

Figure 31. The Lorentzian spectral shape of a sinusoid corrupted by a Wiener phase noise process.

The effect of this nonzero linewidth is the subject of the next section.

5. Laser Phase Noise Analysis

Much of the recent research in semiconductor lasers has concentrated on reducing laser linewidth. Nevertheless, the linewidths are still significant, and it looks like this impairment will be around for a while. It is important, then, to analyze its effects on the performance of various coherent receivers, as the phase noise must be understood before its overall impact can be diminished. There are numerous papers available which study the effects of laser phase noise; see e.g., [22] [1] [23] [24]. Most of the proceeding analysis is based on the works by Kazovsky [22] [25].

All of the analysis in the previous sections assumed zero linewidth ($\Delta\nu = 0$). When this assumption is not valid, the BER curves as shown in figure 12 are affected in two ways:

Power penalty:

If $\Delta\nu > 0$, then the Lorentzian shape of the light source tends to smear the spectrum of the modulated signal. To recover the signal, then, the IF bandwidth must be increased, which subsequently passes more shot noise. This power penalty results in a shift of the BER curve to the right. Note that by increasing the received power P_s, this penalty can be overcome (see figure 32).

BER floor:

For angle modulation schemes (PSK, FSK), a nonzero linewidth can be devastating. It should be clear that if the phase process were allowed to take a "random walk," the performance of a DPSK receiver, which relies on the coherence of the phase in two successive bit intervals, would rapidly deteriorate. For a given $\Delta\nu$, then, one must choose the bit rate R large enough so that there still is some phase correlation

in successive pulses. If not, then *no* amount of received power P_s can overcome this effect. The result is a *BER floor*, and is illustrated in figure 32.

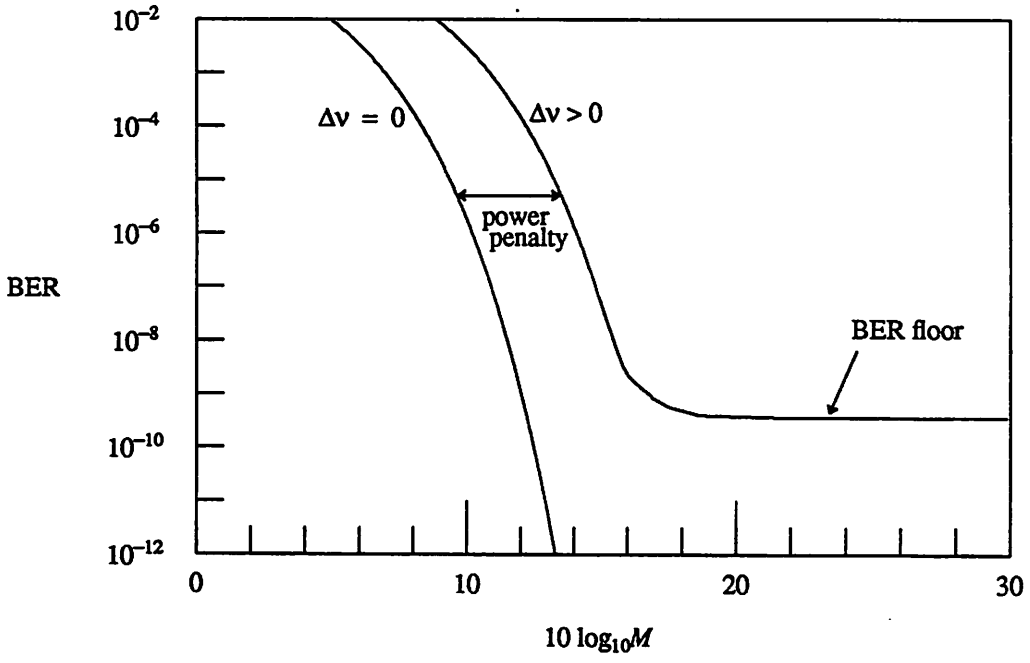

Figure 32. Phase noise has two effects on a BER curve, a power penalty, which causes a shift to the right, and a BER floor, which is a lower limit on the BER.

The power penalties are in general more difficult to obtain analytically, because they require simultaneous consideration of both phase noise and shot noise. First, we will compute the BER floors for angle-modulation schemes.

5.1. Effect of Phase Noise on FSK

The received lightwave corrupted by laser phase noise for FSK can be represented as

$$r(t) = \sqrt{P_s}\cos[(\omega_0 \pm \omega_d)t + \phi_s(t)]. \tag{156}$$

where $\phi_s(t)$ is the phase noise process, and $\dot{\phi}_s(t)$ is a zero-mean white Gaussian process with with spectral density $S_{\dot{\phi}_s}(f) = 2\pi\Delta\nu_s$. Similarly, the LO lightwave is

$$l(t) = \sqrt{P_{LO}}\cos[\omega_{LO}t + \phi_{LO}(t)], \tag{157}$$

where this time $S_{\dot{\phi}_{LO}}(f) = 2\pi\Delta\nu_{LO}$. When the two lightwaves are combined and directed towards a PIN diode, the resulting IF current after dc elimination is

$$i(t) = 2R\sqrt{P_s P_{LO}}\cos[(\omega_{IF} \pm \omega_d)t + \phi(t)], \tag{158}$$

where

$$\phi(t) \triangleq \phi_s(t) - \phi_{LO}(t) \tag{159}$$

is again a Wiener process, and

$$S_{\dot{\phi}}(f) = 2\pi(\Delta\nu_s + \Delta\nu_{LO}). \tag{160}$$

Throughout the rest of this analysis, assume $\Delta\nu_s + \Delta\nu_{LO} = 2\Delta\nu$, so that $S_{\dot{\phi}}(f) = 4\pi\Delta\nu$, and $\Delta\nu$ can be thought of as the average of the transmitter and LO linewidths.

Consider the FSK receiver shown in figure 33.

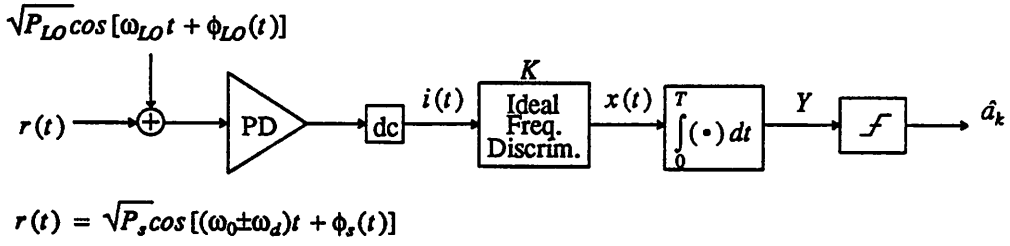

Figure 33. A practical FSK-discriminator receiver.

An ideal frequency discriminator is used, whose output is the instantaneous frequency of its input with respect to ω_{IF}, scaled by a conversion factor K. Thus

$$x(t) = K\left[\begin{array}{c}\text{instantaneous}\\ \text{frequency}\\ \text{of } i(t)\end{array} - \omega_{IF}\right] \tag{161}$$

$$= K[\pm\omega_d + \dot{\phi}(t)].$$

Since $\dot{\phi}(t)$ is a white noise process, we see that for the case of FSK with a frequency discriminator receiver, the laser phase noise problem has reduced to a binary antipodal scheme in additive white Gaussian noise. The optimum receiver integrates $x(t)$ over one bit period, then applies a zero threshold test, as shown in figure 33. The output of the integrator Y is

$$\begin{aligned} Y &= \int_0^T x(t)\,dt \\ &= K\int_0^T (\pm\omega_d + \phi(t))\,dt \\ &= K(\pm\omega_d T + N), \end{aligned} \tag{162}$$

where

$$N = \int_0^T \dot{\phi}(t)\,dt \tag{163}$$

is a zero-mean Gaussian random variable with variance

$$E[N^2] = \mathrm{E}[\int_0^T\int \dot{\phi}(u)\dot{\phi}(v)\,dudv]$$

$$= 4\pi\Delta\nu \int_0^T \int \delta(u - v)\, du dv$$

$$= 4\pi\Delta\nu T. \tag{164}$$

The BER is then

$$\begin{aligned} P_e &= \Pr[Y < 0 \mid \textit{ONE}] \\ &= Q\left[\frac{K\omega_d T}{K\sqrt{4\pi\Delta\nu T}}\right] \\ &= Q\left[\frac{2\pi f_d T}{\sqrt{4\pi\Delta\nu T}}\right] \\ &= Q\left[\frac{\sqrt{\pi} f_d / R}{\sqrt{\Delta\nu / R}}\right]. \end{aligned} \tag{165}$$

Define the *deviation ratio*

$$h \triangleq \frac{2f_d}{R}, \tag{166}$$

where f_d is half the difference between the *ONE* and *ZERO* frequencies, and R is the bit rate. Then from (165), we find the BER floor due to laser phase noise for FSK discrimination detection:

$$P_e = Q\left[\frac{\sqrt{\pi/4}\, h}{\sqrt{\Delta\nu / R}}\right]. \tag{167}$$

The BER floors for FSK are plotted in figure 34, parameterized by h. Laser linewidth requirements can be obtained from these curves. For instance, suppose you want to build an MSK ($h = 0.5$) system to run at 100 Mbps. From figure 34 (or more accurately from equation (167)), you see that to get a BER floor of less than 10^{-9}, you need

$$\Delta\nu / R < 0.00546, \tag{168}$$

so that $\Delta\nu$ must be less than

$$\begin{aligned} \Delta\nu &< 0.00546 R \\ &< 0.00546\,(100 \text{ Mbps}) \\ &< 546 \text{ KHz}. \end{aligned} \tag{169}$$

Thus to build such a system, the average of the linewidths of your transmitter and LO lasers must be less than 546 KHz. Note that equation (167) was derived assuming no shot noise. For $h \lesssim 3$, the signal spectrum is compact enough so that the shot noise admitted through the IF filter will cause a penalty of no more than 1 dB [22]. For larger deviation, however, the IF bandwidth must be so large that the amount of shot noise let in negates any gain obtained with respect to the phase noise. In this respect, equation (167) must be viewed as being valid (to within 1 dB) only for $h \lesssim 3$.

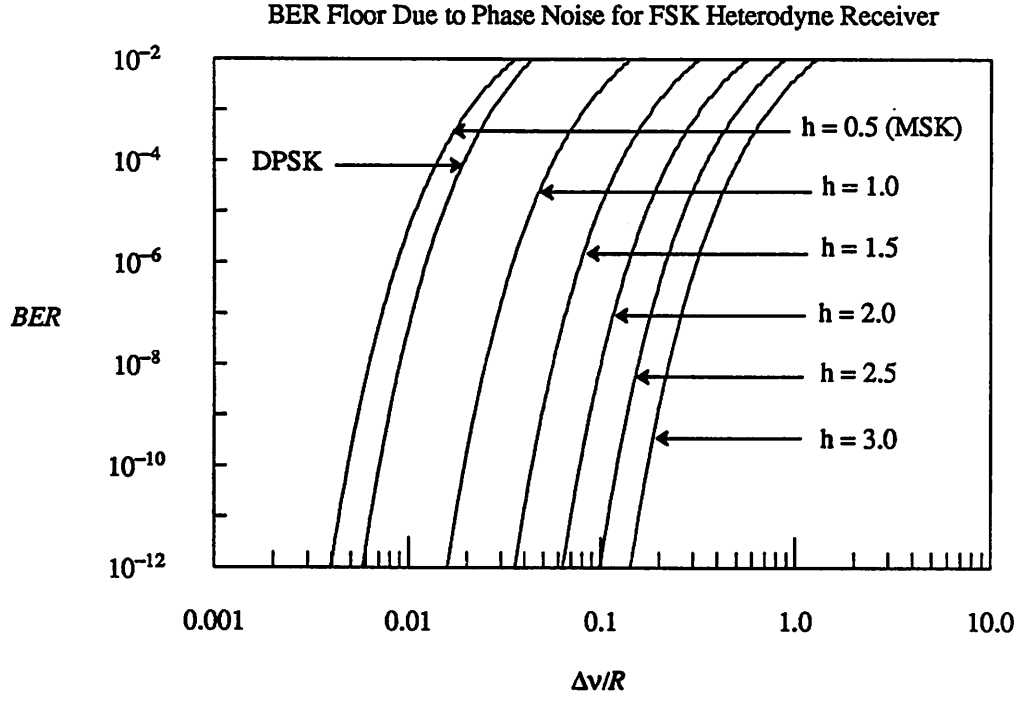

Figure 34. BER floors for FSK discriminator reception due to phase noise: plotted versus $\Delta\nu/R$, where $\Delta\nu$ is the average linewidth of the transmitter and LO lasers, and R is the bit rate. The parameter is h, the deviation ratio, which is the frequency difference between a *ONE* and *ZERO* divided by the bit rate. Also included is the BER floor for DPSK, derived in the next section.

5.2. Effect of Phase Noise on DPSK

The DPSK receiver structure was analyzed for zero phase noise in section 3.2.1. It was shown that the ML receiver (constrained to the delay-demodulation technique) found the correlation of two successive pulses and applied the result to a zero threshold test. The BER was then found by recognizing the equivalence of this correlation test with a signal space projection test. When $\Delta\nu > 0$, the analysis is not so simple. However, observe that an ideal DPSK receiver would determine the phase of $i(t)$ and $i(t-T)$, and average their difference. Thus a lower bound on the BER can be obtained from this idealized model [26] [22] shown in figure 35.

The IF current of the DPSK receiver will be

$$i(t) = 2R\sqrt{P_s P_{LO}}\cos[\omega_{IF}t + \phi(t) + \phi_k] \tag{170}$$

where $\phi_k \in \{0, \pi\}$ is the phase modulation, and $\phi(t)$ is the Wiener phase noise process. The estimates of the phase of $i(t)$ and $i(t-T)$ will therefore be

$$\theta(t) \triangleq \phi(t) + \phi_k$$

and

$$\theta_D(t) \triangleq \phi(t-T) + \phi_{k-1} \tag{171}$$

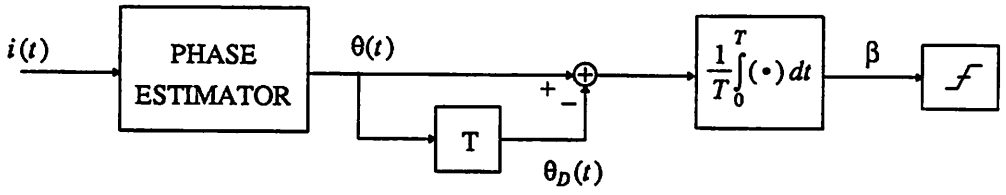

Figure 35. The performance of a DPSK receiver can do no better than this idealized receiver, which averages the phase difference between $i(t)$ and $i(t-T)$.

The BER floor for DPSK is the probability that the phase noise causes the average phase difference

$$\beta \triangleq \frac{1}{T}\int_0^T \Big[\theta(t) - \theta_D(t)\Big]\, dt \tag{172}$$

to vary by more than $\pi/2$:

$$\begin{aligned} P_e &= \Pr[\text{miss}] \\ &= \Pr[|\beta|_{\text{mod } 2\pi} > \frac{\pi}{2} \,|\, \textit{ONE}]. \end{aligned} \tag{173}$$

But from (171),

$$\beta = \frac{1}{T}\int_0^T \Big[\phi(t) - \phi(t-T)\Big]\, dt \;+\; \phi_k \;-\; \phi_{k-1}. \tag{174}$$

To evaluate the BER using (173), assume a *ONE* was sent, which from (104) implies that $\phi_k - \phi_{k-1} = 0$. Thus β becomes

$$\beta = \frac{1}{T}\int_0^T \Big[\phi(t) - \phi(t-T)\Big]\, dt. \tag{175}$$

Since $\phi(t)$ is defined as the integral of $\dot{\phi}(t)$, and since $\dot{\phi}(t)$ is a zero-mean Gaussian process, β is a zero-mean Gaussian random variable. Its variance, derived in appendix C, is

$$\sigma_\beta^2 \triangleq \mathrm{E}[\beta^2] = \frac{8}{3}\pi\Delta\nu/R. \tag{176}$$

For reasonably small $\Delta\nu/R$, the probability of $|\beta| > 3\pi/2$ is negligible, so that the *mod* 2π on the absolute value in (173) can be removed. Thus the BER floor due to laser phase noise for DPSK is

$$\begin{aligned} P_e &= \Pr[|\beta| > \pi/2 \,|\, \textit{ONE}] \\ &= 2Q\left[\frac{\pi/2}{\sigma_\beta}\right] \\ &= 2Q\left[\frac{\pi/2}{\sqrt{(8/3)\pi\Delta\nu/R}}\right] \end{aligned}$$

$$= 2Q\left[\sqrt{\frac{3\pi}{32\Delta\nu/R}}\right]. \tag{177}$$

Comparing this result with (167), we see that this BER floor for DPSK is equivalent to the BER for FSK (except for the factor of 2), with a deviation ratio of $h = \sqrt{\frac{3}{8}}$. The plot of this floor is included with the FSK floors in figure 34. It can be seen from the graph that for a BER floor of less that 10^{-9}, one must have

$$\Delta\nu/R < 0.00819. \tag{178}$$

Thus, for example, to build a DPSK system at 100 Mbps, the linewidths from the transmitter and LO lasers must average less than 819 KHz. The graph in figure 34 also reveals that the laser linewidth requirements for DPSK are less stringent than for MSK, but more stringent than for large deviation FSK.

5.3. Effect of Phase Noise on ASK

If there is no phase noise ($\Delta\nu = 0$), then the signal constellation for ASK lies completely on the ψ_c axis, as shown in figure 36(a).

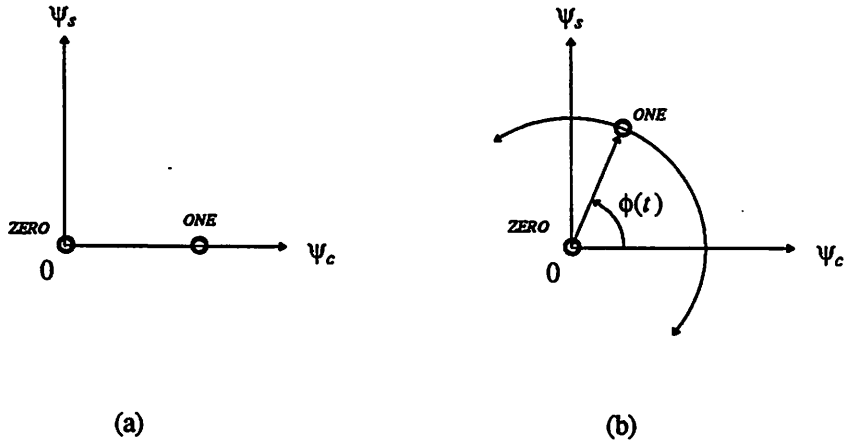

Figure 36. Signal space representation for ASK. If there is no phase noise, then the constellation lies completely on ψ_c axis (a). Phase noise causes the constellation to rotate randomly, as shown in (b).

IF $\Delta\nu > 0$, however, then the constellation rotates randomly, as depicted in figure 36(b). The axes for this discussion are defined as

$$\psi_c \leftrightarrow \sqrt{\frac{2}{T}}\cos\omega_0 t$$

$$\psi_s \leftrightarrow \sqrt{\frac{2}{T}}\sin\omega_0 t. \tag{179}$$

The conventional ASK receiver, as developed in section 3.2.1, found the correlation of the received lightwave with ψ_c, and applied the result to a threshold test. But if there is some phase noise, this method is no longer optimum. For example, if the phase noise

$\phi(t)$ lingers about $\pi/2$, then one can see from figure 36(b) that the projection onto ψ_c would always be zero. A better receiver would look at both dimension in signal space, which leads to the receiver called a *two-port* or *phase-diversity* receiver.

It can be shown [11] that for a received signal $i(t)$ with a constant unknown phase, the ML ASK receiver (accounting for both phase noise and shot noise) computes

$$Y = \langle \mathbf{i}, \psi_c \rangle^2 + \langle \mathbf{i}, \psi_s \rangle^2 \tag{180}$$

and compares the result to a threshold. For the more general case when the phase is not constant but a random process $\phi(t)$, deriving the ML receiver is more difficult. However, one is always free to choose a receiver structure, then analyze its performance and hope it is acceptable. Extending the optimal receiver for unknown constant phase to the case of random time-varying phase, consider the phase-diversity receiver shown in figure 37 (adapted from [25]).

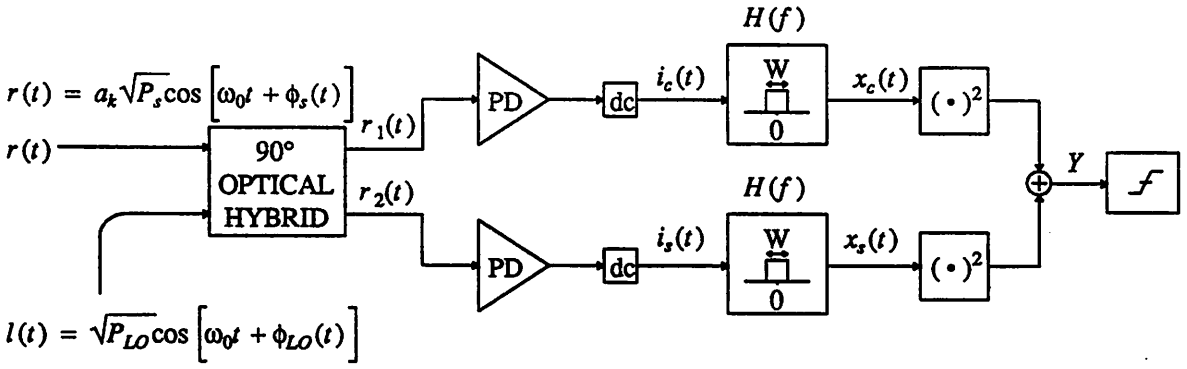

Figure 37. A phase diversity homodyne ASK receiver, following [25].

The 90° hybrid is a device whose output are the in-phase and quadrature sums of it inputs [27]. That is,

$$r_1(t) = \frac{1}{\sqrt{2}} \mathrm{Re}\{ r(t) + l(t) \}$$

and

$$r_2(t) = \frac{1}{\sqrt{2}} \mathrm{Re}\{ jr(t) + l(t) \}. \tag{181}$$

The resulting baseband currents are

$$\begin{aligned} i_c(t) &= a_k I \cos\phi(t) + i_{sh,c}(t) \\ i_s(t) &= a_k I \sin\phi(t) + i_{sh,s}(t) \end{aligned} \tag{182}$$

where

$$I \triangleq R\sqrt{P_s P_{LO}} \tag{183}$$

is the baseband current amplitude and

$$\phi(t) \triangleq \phi_s(t) - \phi_{LO}(t) \tag{184}$$

is the combined phase noise process. Its time derivative $\dot{\phi}(t)$ is white Gaussian with

density $S_{\dot{\phi}}(f) = 4\pi\Delta\nu$. The terms $i_{sh,c}(t)$ and $i_{sh,s}(t)$ are the zero-mean white Gaussian shot noises in branches one and two, respectively, with spectral density $S_{sh,c}(f) = S_{sh,s}(f) = N_0 \triangleq qRP_{LO}$. Note that since they arise from physically separate photodetectors, $i_{sh,c}(t)$ and $i_{sh,s}(t)$ are independent. The low-pass filter in figure 37 has a bandwidth W, which is chosen large enough to pass without distortion the signal corrupted by phase noise. Under this assumption, the filtered currents in (182) become

$$\begin{aligned} x_c(t) &= a_k I \cos\phi(t) + n_c(t) \\ x_s(t) &= a_k I \sin\phi(t) + n_s(t) \end{aligned} \tag{185}$$

where n_c and n_s are zero-mean filtered white noise processes with spectral densities

$$S_{n_c}(f) = S_{n_s}(f) = N_0 \text{rect}(f, W/2). \tag{186}$$

The input to the slicer is $Y = x_c^2 + x_s^2$. Note that if there were no shot noise, then

$$\begin{aligned} Y &= x_c^2 + x_s^2 \\ &= a_k^2 I^2 \cos^2\phi(t) + a_k^2 I^2 \sin^2\phi(t) \\ &= a_k I^2. \end{aligned} \tag{187}$$

Thus without the shot noise, the receiver recovers the signal exactly. The fact that there *is* shot noise, however, complicates the matter, causing Y to be a non-Gaussian random variable. It has some n–cross–n terms, s–cross–n terms, and its mean and variance depend on the transmitted symbol.

To obtain a numerical BER, make two approximations:

- Y is nearly Gaussian
- $P_e = Q(\rho)$,

where

$$\rho \triangleq \frac{m_1 - m_0}{\sigma_1 + \sigma_0} \tag{188}$$

$$m_1 \triangleq \text{E}[Y \mid \textit{ONE}] \tag{189}$$

$$m_0 \triangleq \text{E}[Y \mid \textit{ZERO}] \tag{190}$$

$$\sigma_1^2 \triangleq \text{var}[Y \mid \textit{ONE}] \tag{191}$$

$$\sigma_0^2 \triangleq \text{var}[Y \mid \textit{ZERO}] \tag{192}$$

The second approximation made above was discussed in section 2.4, and is valid for large $I^2/(N_0 W) \sim M$. For the sake of brevity, the derivations of the conditional means and variances of Y are relegated to appendix D. The results are

$$m_1 = I^2 + 2N_0 W \tag{193}$$

$$m_0 = 2N_0 W \tag{194}$$

$$\sigma_1^2 = 4I^2 N_0 W + 4N_0^2 W^2 \tag{195}$$

$$\sigma_0 = 4N_0^2 W^2. \tag{196}$$

Computing the SNR defined in (188),

$$\begin{aligned}\rho &= \frac{m_1 - m_0}{\sigma_1 + \sigma_0} \\ &= \frac{I^2}{\sqrt{4I^2N_0W + 4N_0^2W^2} + 2N_0W} \\ &= \frac{\frac{1}{\alpha}M}{\sqrt{\frac{2}{\alpha}M + 1} + 1}\end{aligned} \tag{197}$$

where

$$\alpha \triangleq \frac{W}{R} \tag{198}$$

is the normalized bandwidth of $H(f)$. The bandwidth must be large enough so that $a_k I\cos\phi(t)$ and $a_k I\sin\phi(t)$ are passed undistorted. The larger $\Delta\nu$ is, the larger W must be. An expression for W in terms of $\Delta\nu$ is difficult to obtain; a slight modification of a result by Kazovsky [22] yields

$$W = R\sqrt{1 + 161.3(\Delta\nu/R)^2}, \tag{199}$$

so that

$$\alpha = \sqrt{1 + 161.3(\Delta\nu/R)^2}. \tag{200}$$

Notice that for $\Delta\nu = 0$, $\alpha = 1$, which corresponds to a matched filter ($W = R$). Combining equations (197) and (200), the BER for a homodyne phase-diversity receiver is

$$P_e = Q\left[\frac{\frac{M}{\sqrt{1 + 161.3(\Delta\nu/R)^2}}}{\sqrt{1 + \frac{2M}{\sqrt{1 + 161.3(\Delta\nu/R)^2}}} + 1}\right]. \tag{201}$$

The BER curves from this result are plotted in figure 38. Note that for the special case of $\Delta\nu = 0$, then $\alpha = 1$ and the BER reduces to

$$P_e = Q\left[\frac{M}{\sqrt{2M + 1} + 1}\right]. \tag{202}$$

Furthermore, for large peak received power M,

$$P_e = Q\left[\sqrt{\frac{1}{2}M}\right]. \tag{203}$$

Thus even under ideal conditions ($\Delta\nu = 0$, large M), the phase-diversity homodyne ASK receiver performs 3 dB worse than homodyne synchronous ASK, and 6 dB worse than the quantum limit. Nevertheless, phase diversity techniques may be important in multi-gigabit applications [28]. In this regime, the IF bandwidth needed for a heterodyne receiver makes homodyne systems more attractive. And unlike most homodyne

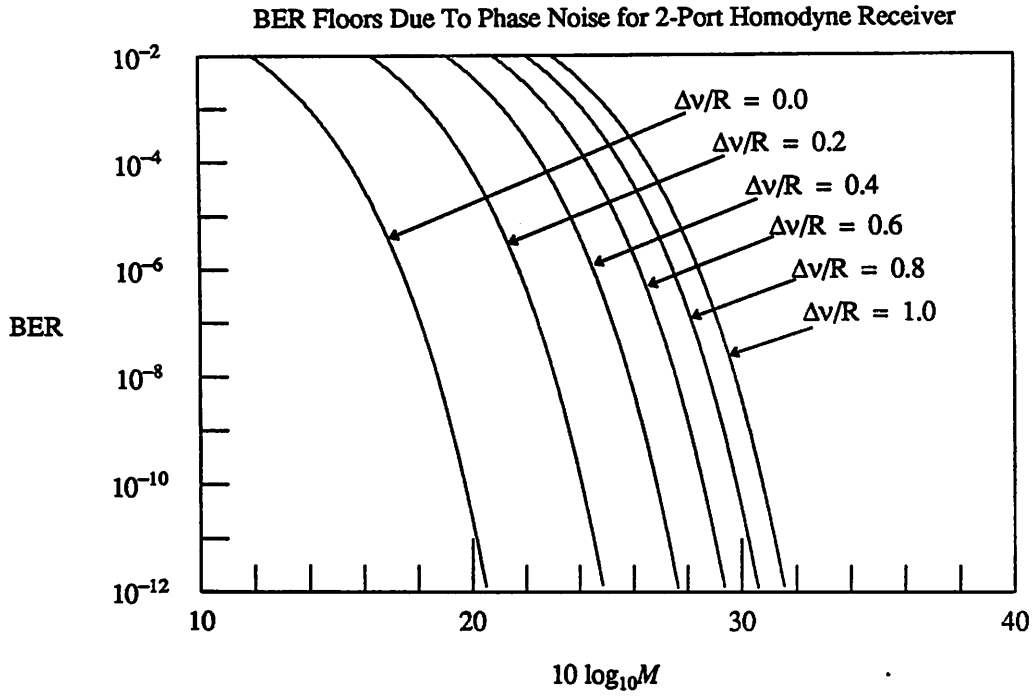

Figure 38. BER curves for 2-port ASK homodyne receiver (figure 37). The parameter is $\Delta\nu/R$, where $\Delta\nu$ is the average laser linewidth, and R is the bit rate. For $\Delta\nu = 0$, the performance of this receiver is 3 dB worse than the homodyne synchronous receiver under shot-noise-limited conditions.

schemes, the phase diversity receiver does not need to track phase.

From inspection of the BER curves in figure 38, we see that nonzero linewidth results in a power penalty- not a BER floor- for the homodyne phase diversity receiver. This is an indirect result of the increased bandwidth necessary to accommodate the signal contaminated by phase noise.

6. Discussion

The analysis of optical receivers in the preceding text had two major themes: the computation of the BER under shot-noise-limited conditions for various receivers, and the effects of laser phase noise on receiver performance.

The BER computations yielded a complete set of equations for various heterodyne, homodyne, synchronous, and asynchronous receivers. These equations were derived assuming all noises were negligible with respect to the shot noise, and thus represent a lower bound on the BER. The results are tabulated in table 1. In practice, laser phase noise [4] [22], intersymbol interference, chromatic and polarization dispersion, imperfect modulation [2], dark current and thermal noise [6] will all degrade performance.

Modulation Format	Receiver Type	BER	Sensitivity
PSK	homodyne	$Q\left[\sqrt{4M}\right]$	9
PSK	heterodyne-synchronous	$Q\left[\sqrt{2M}\right]$	18
IM/DD	(Quantum Limit)	$\frac{1}{2}e^{-M}$	20
DPSK	delay-demodulation	$\frac{1}{2}e^{-M}$	20
ASK	homodyne	$Q\left[\sqrt{M}\right]$	36
FSK	heterodyne-synchronous	$Q\left[\sqrt{M}\right]$	36
FSK	heterodyne-asynchronous	$\frac{1}{2}e^{-M/2}$	40
ASK	heterodyne-synchronous	$Q\left[\sqrt{M/2}\right]$	72
ASK	heterodyne-asynchronous	$\frac{1}{2}e^{-M/4}$	80

Table 1. BERs and sensitivities of various shot-noise-limited receivers.

The equations are expressed in terms of $M \triangleq \eta P_s T/(h\nu)$, the number of photons per *ONE* bit. It should be emphasized that M is the *peak* power per bit. Some European authors prefer to talk in terms of the average power per bit. To compare their results with those of table 1, one must realize that for ASK and IM/DD, M is *twice* the average power per bit.

From the equations in table 1, one can compute receiver sensitivity: the number of photons M required for a BER of 10^{-9}. The sensitivity of each receiver considered is included in table 1. Here we see that homodyne PSK is the most sensitive. As was discussed in section 4.7, and as the tabulated results verify, heterodyne reception is 3 dB

less sensitive than homodyne. Another advantage of homodyne systems is the reduced IF bandwidth needed, thus easing the electrical processing requirements. The fact that homodyne receivers generally need to track the phase of the incoming signal is a major disadvantage. One possible solution, discussed in section 5.3, is a phase-diversity receiver.

Figure 12 shows a plot of the BER versus M, using the expressions in table 1. From this plot, or from table 1, it is evident that the asynchronous receivers incur an additional penalty of ~ 0.5 dB with respect to their synchronous counterparts. This penalty is often seen as a reasonable compromise between receiver performance and complexity, especially since asynchronous receivers are in general less sensitive to laser phase noise [23] [22] [24].

Also included in table 1 is the BER for the ideal IM/DD receiver, and the accompanying sensitivity (called the quantum limit) of 20 photons per *ONE* bit. In section 2.4 it was demonstrated that extraneous noises such as thermal noise and dark current in a realistic IM/DD receiver caused a 26 dB penalty with respect to the quantum limit. In an attempt to overcome these extraneous noises, typical IM/DD receivers use APDs. Although these sensitive photodetectors improve receiver sensitivity somewhat, detailed analysis [7] reveals that IM/DD receivers still require anywhere from 800-2000 photons per *ONE* bit. This corresponds to a 16-20 dB power penalty with respect to the quantum limit.

In section 4.8, laser phase noise was characterized as a Wiener process. Section 5 then went on to analyze the effect of this phase noise on FSK, DPSK, and ASK receivers. The resulting analysis revealed a BER floor for the angle-modulation schemes, namely FSK and DPSK. The expressions for these floors (equations (167) and (177)) were derived assuming that the power of the received signal was sufficient to recover the phase unambiguously. It should be clear that if the received power P_s were small, then the thermal and shot noises would in effect alter the phase, resulting in reduced sensitivity.

Comparing the BER floor plots in figure 34, we see that MSK ($h = 0.5$) requires the narrowest laser linewidth: to obtain a floor of 10^{-9}, the average laser linewidth of the transmitter and LO must be less than 0.55% of the bit rate. In contrast, wide deviation FSK ($h = 3$) can achieve this BER floor with an average linewidth as high as 20% of the bit rate. For even larger deviation FSK ($h > 3$), the linewidth requirements are even less stringent [22]. This cannot be inferred from (167) directly, because when the separation between *ONE* and *ZERO* frequencies gets very large, the shot noise admitted by the widened IF bandwidth results in a power penalty. This secondary effect is not reflected in (167). And although (167) is still a lower bound on the BER for ($h > 3$), it is no longer a very tight bound. Again from figure 34, it is seen that DPSK is is almost as sensitive to phase noise as MSK. A DPSK receiver requires an average linewidth of no more than 0.82% of the bit rate to obtain a BER of 10^{-9}.

The analysis of the phase diversity receiver resulted in figure 38, which shows the power penalties for nonzero laser linewidths. The derivation of the expression used for this plot (equation (201)) required numerous approximations. This is because both phase noise and shot noise were accounted for, significantly complicating the analysis. Nevertheless, the results imply the possibility of a homodyne, phase insensitive coherent receiver- an intriguing combination. Recent experiments verify the feasibility of the

phase-diversity receiver [29] [30].

7. Conclusion

An analysis of coherent optical fiber receivers was presented. A review of direct detection techniques was provided, and the quantum limit for an ideal IM/DD receiver was found to be 20 photons per *one* bit. The photodetector current was shown to be contaminated by shot noise. When the optical power incident to the detector is large enough, the resulting shot noise process can be approximated as a Gaussian white noise process, thus simplifying analysis. Because of extraneous noises such as dark current and thermal noise, a practical IM/DD receiver using a PIN diode was shown to be 26 dB less sensitive than the quantum limit. Coherent techniques were shown to alleviate these extraneous noise effects by maintaining a high local oscillator power. This high power can cause the shot noise to dominate all other noise, a condition referred to as shot-noise-limited operation.

The BER for various heterodyne, homodyne, synchronous, and asynchronous receivers was derived for shot-noise-limited conditions. It was found that homodyne PSK was 3.5 dB more sensitive than the quantum limit, requiring only 9 photons per bit for a BER of 10^{-9}. The next most sensitive receiver was heterodyne PSK, followed by DPSK, homodyne ASK, heterodyne synchronous FSK, heterodyne asynchronous FSK, heterodyne synchronous ASK, and heterodyne-asynchronous ASK. A straightforward explanation of the 3 dB penalty for heterodyne receivers was presented. Then laser phase noise was modeled as a Wiener process, and its effect on FSK, DPSK, and ASK receivers was examined. A BER floor of 10^{-9} required that the average laser linewidth of the transmitter and LO be less than 0.55%, 0.82%, and 20% of the bit rate for MSK, DPSK, and FSK ($h = 3$), respectively. Finally, the effect of phase noise on a homodyne phase-diversity ASK receiver was examined. It was found that nonzero laser linewidth resulted in a power penalty only, implying that its effect could be reduced by increasing the transmitted power.

8. Appendix

Appendix A. Spectrum of a Modulated Random Process.

What follows is a derivation of the spectral density of the product of a stationary random process with a sinusoid. Suppose $n(t)$ is a zero-mean stationary process, with autocorrelation function $R_n(\tau)$ and spectral density function $S_n(f)$. Next let

$$y(t) = n(t)\cdot(2\cos(\omega_0 t + \theta)). \tag{A-1}$$

The random variable θ is uniform over $[0,2\pi]$ and independent of $n(t)$, and is needed to make $y(t)$ stationary. Then to compute the spectral density of $y(t)$, first obtain its autocorrelation function:

$$R_y(\tau) = \mathrm{E}[y(t)y(t+\tau)] \tag{A-2}$$

$$= 4\mathrm{E}[n(t)n(t+\tau)\cos(\omega_0 t + \theta)\cos(\omega_0 t + \omega_0\tau + \theta)] \tag{A-3}$$

$$4R_n(\tau)\mathrm{E}[\frac{1}{2}\cos\omega_0 t + \frac{1}{2}\cos(2\omega_0 t + \omega_0\tau + 2\theta)] \tag{A-4}$$

$$= 2R_n(\tau)\left[\cos(\omega_0\tau) + \mathrm{E}[\cos(2\omega_0 t + \omega_0\tau + 2\theta)]\right] \tag{A-5}$$

Since θ is uniform over $[0,2\pi]$, the expectation over θ is zero:

$$\mathrm{E}[\cos(2\omega_0 t + \omega_0\tau + 2\theta)] = \frac{1}{2\pi}\int_0^{2\pi}\cos(2\omega_0 t + \omega_0\tau + 2\theta)\,d\theta \tag{A-6}$$

$$= 0.$$

Thus

$$R_y(\tau) = 2R_n(\tau)\cos\omega_0\tau. \tag{A-7}$$

Taking the Fourier transform gives the spectral density of $y(t)$:

$$S_y(f) = S_n(f + f_0) + S_n(f - f_0). \tag{A-8}$$

Therefore, if $n(t)$ is a white noise process with density N_0, then $y(t)$ will also be a white noise process, but with density $2N_0$.

Appendix B. The Mean of cos Φ, where Φ is a Gaussian random variable.

Suppose Φ is a zero-mean Gaussian random variable with variance σ^2. First, calculate

$$\mathrm{E}\left[e^{j\Phi}\right] = \int_{-\infty}^{\infty} e^{j\phi} p_\Phi(\phi)\,d\phi \tag{B-1}$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(\phi^2 - j2\sigma^2\phi)/2\sigma^2}d\phi \tag{B-2}$$

$$= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(\phi - j\sigma^2)^2/2\sigma^2}e^{+(j\sigma^2)^2/2\sigma^2}d\phi \tag{B-3}$$

$$= e^{-\sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\phi - j\sigma^2)^2/2\sigma^2} d\phi \tag{B-4}$$

This is an integral of a Gaussian probability density function (pdf) along a line at $j\sigma^2$ parallel to the real axis in the complex plane. Since the Gaussian pdf has no poles in the region bounded by the real axis and a line parallel to it at $j\sigma^2$, its integral on the contour bounding this region is zero. Therefore the integral at $j\sigma^2$ in one direction is the negative of the integral on the real axis in the other:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\phi - j\sigma^2)^2/2\sigma^2} d\phi = -\int_{+\infty}^{-\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\phi^2/2\sigma^2} d\phi \tag{B-5}$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\phi^2/2\sigma^2} d\phi \tag{B-6}$$

This is just the total integral of the Gaussian pdf, which is one. Thus

$$\mathrm{E}\left[e^{j\Phi}\right] = e^{-\sigma^2/2}. \tag{B-7}$$

And since $-\Phi$ is also a zero-mean Gaussian random variable with variance σ^2, we must also have

$$\mathrm{E}\left[e^{-j\Phi}\right] = e^{-\sigma^2/2}. \tag{B-8}$$

Therefore,

$$\mathrm{E}[\cos\Phi] = \frac{1}{2}\mathrm{E}\left[e^{j\Phi}\right] + \frac{1}{2}\mathrm{E}\left[e^{-j\Phi}\right] \tag{B-9}$$

$$= e^{-\sigma^2/2}, \tag{B-10}$$

and

$$\mathrm{E}[\sin\Phi] = \frac{1}{2j}\mathrm{E}\left[e^{j\Phi}\right] - \frac{1}{2j}\mathrm{E}\left[e^{-j\Phi}\right] \tag{B-11}$$

$$= 0. \tag{B-12}$$

Appendix C. Derivation of variance of β.

From equation (113),

$$\beta \triangleq \frac{1}{T}\int_{o}^{T} \Big[\phi(t) - \phi(t-T)\Big] dt \tag{C-1}$$

and

$$\phi(t) \triangleq \int_{0}^{t} \dot{\phi}(u) du, \tag{C-2}$$

where $\dot{\phi}(u)$ is a zero-mean white Gaussian process with density $S_{\dot{\phi}}(f) = 4\pi\Delta\nu$. First, define the phase difference

$$\phi_T(t) \triangleq \phi(t) - \phi(t-T) \tag{C-3}$$

$$= \int_0^t \dot{\phi}(u)\,du - \int_0^{t-T} \dot{\phi}(u)\,du \tag{C-4}$$

$$= \int_{t-T}^{t} \dot{\phi}(u)\,du. \tag{C-5}$$

Thus $\phi_T(t)$ is a zero-mean Gaussian random process with autocorrelation function

$$R_{\phi_T}(\tau) = E\left[\phi_T(t)\phi_T(t+\tau)\right] \tag{C-6}$$

$$= E\left[\int_{(t-T)}^{t}\int_{(t+\tau-T)}^{(t+\tau)} \dot{\phi}(u)\dot{\phi}(v)\,dudv\right] \tag{C-7}$$

$$= 4\pi\Delta\nu \int_{(t-T)}^{t}\int_{(t+\tau-T)}^{(t+\tau)} \delta(u-v)\,dudv \tag{C-8}$$

$$= 4\pi\Delta\nu T\ \mathrm{tria}(\tau, T) = \begin{cases} 4\pi\Delta\nu(T-|\tau|) & \text{for } |\tau| < T \\ 0 & \text{for } |\tau| > T \end{cases} \tag{C-9}$$

Now equation (C-1) becomes

$$\beta = \frac{1}{T}\int_0^T \phi_T(u)du \tag{C-10}$$

and its variance is

$$\mathrm{var}[\beta] = E[\beta^2] \tag{C-11}$$

$$= E\left[\frac{1}{T^2}\int_0^T\int \phi_T(u)\phi_T(v)\,dudv\right] \tag{C-12}$$

$$= \frac{1}{T^2}\int_0^T\int R_{\phi_T}(u-v)\,dudv \tag{C-13}$$

$$= \frac{4\pi\Delta\nu}{T^2}\int_0^T\int \left[T - |u-v|\right]dudv \tag{C-14}$$

$$= \frac{4\pi\Delta\nu}{T^2}\int_0^T\left[\int_0^{\nu}\left[T+u-v\right]du + \int_{\nu}^{T}\left[T-u+v\right]du\right]dv \tag{C-15}$$

$$= \frac{4\pi\Delta\nu}{T^2}\int_0^T\left[T\nu + \frac{1}{2}\nu^2 - \nu^2 + T^2 - T\nu - \frac{1}{2}T^2 + \frac{1}{2}\nu^2 + T\nu - \nu^2\right]dv \tag{C-16}$$

$$= \frac{4\pi\Delta\nu}{T^2}\int_0^T \left[T\nu - \nu^2 + \frac{1}{2}T^2\right] d\nu \tag{C-17}$$

$$= \frac{4\pi\Delta\nu}{T^2}\left[\frac{1}{2}T^3 - \frac{1}{3}T^3 + \frac{1}{2}T^3\right] \tag{C-18}$$

$$= \frac{8}{3}\pi\Delta\nu/R. \tag{C-19}$$

Appendix D. Derivation of Conditional Means and Variances of Y.

From equation (120),

$$x_c(t) = a_k I\cos\phi(t) + n_c(t) \tag{D-1}$$

$$x_s(t) = a_k I\sin\phi(t) + n_s(t) \tag{D-2}$$

where $\phi(t)$ is a Wiener process, $n_c(t)$ and $n_s(t)$ are independent zero-mean Gaussian random processes with an ideal low-pass spectral density

$$S_{n_c}(f) = S_{n_s}(f) = N_0\text{rect}(f, W/2). \tag{D-3}$$

Note that for ASK, $a_k \in \{1,0\}$. The derivation of the mean and variance of Y, defined by

$$Y = x_c^2(t) + x_s^2(t) \tag{D-4}$$

requires careful bookkeeping. Proceeding,

$$Y = \left[a_k I\cos\phi(t) + n_c(t)\right]^2 + \left[a_k I\sin\phi(t) + n_s(t)\right]^2 \tag{D-5}$$

$$= a_k{}^2 I^2\cos^2\phi(t) + 2a_k I n_c(t)\cos\phi(t) + n_c^2(t)$$

$$+ a_k{}^2 I^2\sin^2\phi(t) + 2a_k I n_s(t)\sin\phi(t) + n_s^2(t) \tag{D-6}$$

$$= a_k I^2 + 2a_k I\left[n_c(t)\cos\phi(t) + n_s(t)\sin\phi(t)\right] + n_c^2(t) + n_s^2(t) \tag{D-7}$$

But $n_c(t)$, $n_s(t)$, and $\phi(t)$ are all pairwise independent. And since $n_c(t)$ and $n_s(t)$ have zero mean, the expectation of the bracketed term is zero. Furthermore,

$$\text{E}[n_c^2(t)] = \text{E}[n_s^2(t)] = N_0 W. \tag{D-8}$$

Thus

$$\text{E}[Y] = a_k I^2 + 2N_0 W. \tag{D-9}$$

Substituting $a_k = 1$ and 0 results in the conditional means of equation (122). To compute the variance of Y, first compute its mean-square:

$$E[Y^2] = E\Big[\Big(a_k I^2 + 2a_k I \Big[n_c(t)\cos\phi(t) + n_s(t)\sin\phi(t)\Big]\Big)^2$$
$$+ 2\Big[n_c^2(t) + n_s^2(t)\Big]\Big[a_k I^2 + 2a_k I \Big[n_c(t)\cos\phi(t) + n_s(t)\sin\phi(t)\Big]\Big]$$
$$+ n_c^4(t) + 2n_c^2(t)n_s^2(t) + n_s^4(t)\ \Big] \qquad \text{(D-10)}$$

$$= E\Big[a_k I^4 + 4a_k I^3 \Big[n_c(t)\cos\phi(t) + n_s(t)\sin\phi(t)\Big]$$
$$+ 4a_k I^2 \Big[n_c^2(t)\cos^2\phi(t) + 2n_c(t)n_s(t)\cos\phi(t)\sin\phi(t) + n_s^2(t)\sin^2\phi(t)\Big]$$
$$+ 2a_k I^2 n_c^2(t) + 4a_k I \Big[n_c^3(t)\cos\phi(t) + n_c^2(t)n_s(t)\sin\phi(t)\Big]$$
$$+ 4a_k I \Big[n_c(t)n_s^2(t) + n_s^3(t)\sin\phi(t)\Big] + n_c^4(t) + 2n_c^2(t)n_s^2(t) + n_s^4(t)\ \Big] \qquad \text{(D-11)}$$

Since $n_c(t)$ and $n_s(t)$ are zero-mean Gaussians,

$$E[n_c(t)] = E[n_c^3(t)] = E[n_s(t)] = E[n_s^3(t)] = 0. \qquad \text{(D-12)}$$

Therefore,

$$E[Y^2] = a_k I^4 + 4a_k I^2 E\Big[n_c^2(t)\cos^2\phi(t) + n_s^2(t)\sin^2\phi(t)\Big]$$
$$+ 2a_k I^2 E\Big[n_c^2(t) + n_s^2(t)\Big] + E\Big[n_c^4(t) + n_s^4(t)\Big] + 2E\Big[n_c^2(t)n_s^2(t)\Big] \qquad \text{(D-13)}$$

But $n_c(t)$, $n_s(t)$, and $\phi(t)$ are all pairwise independent, and [3]

$$E[n_c^4(t)] = E[n_s^4(t)] = 3N_0^2W^2. \qquad \text{(D-14)}$$

Thus

$$E[Y^2] = a_k I^4 + 4a_k I^2 N_0 W\, E\Big[\sin^2\phi(t) + \cos^2\phi(t)\Big]$$
$$+ 2a_k I^2(2N_0W) + 6N_0^2W^2 + 2N_0^2W^2 \qquad \text{(D-15)}$$

$$= a_k I^4 + a_k 8I^2 N_0 W + 8N_0^2W^2. \qquad \text{(D-16)}$$

Therefore, the variance of Y is

$$\text{var}[Y] = E[Y^2] - E[Y]^2 \qquad \text{(D-17)}$$
$$= \Big(a_k I^4 + a_k 8I^2 N_0 W + 8N_0^2W^2\Big) - \Big(a_k I^4 + 4a_k I^2 N_0 W + 4N_0^2W^2\Big)$$
$$= a_k 4I^2 N_0 W + 4N_0^2W^2. \qquad \text{(D-18)}$$

Substituting 1 and 0 for a_k results in the conditional variances of Y given in equation (122).

References

1. J. Salz, "Coherent Lightwave Communications," *AT&T Technical Journal*, vol. 64, no. 10, pp. 2153-2209, December 1985.
2. Tingye Li, "Advances in Lightwave Research," *AT&T Technical Journal*, vol. 66, no. 1, pp. 5-18, January/February 1987.
3. A. Papoulis, *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill, New York, 1984.
4. W.G. Chambers, *Basics of Communications and Coding,* Clarendon Press, Oxford, 1985.
5. C.H. Henry, "Phase Noise in Semiconductor Lasers," *IEEE J. Lightwave Technology*, vol. LT-4, March 1986.
6. S.D. Personick, "Receiver Design for Digital Fiber Optic Communications Systems, I and II," *Bell System Technical J.*, vol. 52, no. 6, pp. 843-886, July-August 1973.
7. T. Okoshi, K. Emura, K. Kikuchi, and R. Th. Kersten, "Computation of Bit-Error Rate of Various Heterodyne and Coherent-Type Optical Communication Schemes," *J. Optical Communications*, vol. 2, no. 3, pp. 89-96, 1981.
8. S. Saito, Y. Yamamoto, and T. Kimura, "S/N and Error Rate Evaluation for an Optical FSK-Heterodyne Detection System Using Semiconductor Lasers," *IEEE J. Quantum Electronics*, vol. QE-19, no. 2, pp. 180-193, February 1983.
9. G.L. Turin, *Notes on Digital Communication,* Van Nostrand Reinhold, New York, 1969.
10. E.A. Lee and D.G. Messerschmitt, *Digital Communications: The Physical Layer,* Kluwer Academic Press. To be published.
11. J.M. Wozencraft and I.M. Jacobs, *Principles of Communication Engineering,* John Wiley & Sons, New York, 1965.
12. S. Haykin, *Communication Systems,* John Wiley & Sons, New York, 1983.
13. Y. Yamamoto, "Receiver Performance Evaluation of Various Digital Optical Modulation-Demodulation Systems in the 0.5-10 μm Wavelength Region," *IEEE J. Quantum Electronics*, vol. QE-16, no. 11, pp. 1251-1259, November 1980.
14. T. Okoshi, "Polarization-State Control Schemes for Heterodyne or Homodyne Optical Fiber Communications," *IEEE J. Lightwave Technology*, vol. LT-3, no. 6, pp. 1232-1237, December 1985.
15. B.P. Lathi, *Modern Digital and Analog Communications Systems,* pp. 437-525, Holt, Reinhart and Winston, New York, 1983.
16. W.D. Gregg, *Analog and Digital Communication,* pp. 399-430, John Wiley & Sons, 1977.
17. J. Jackel, Bell Communications Research, Red Bank, NJ, August 1987. Private communication.
18. B. Glance, "Polarization Independent Coherent Optical Receiver," *J. Lightwave Technology*, vol. LT-5, no. 2, pp. 274-276, February 1987.

19. R.A. Linke and P.S. Henry, "Coherent Optical Detection: A Thousand Calls on One Circuit," *IEEE Spectrum*, pp. 52-57, February 1987.

20. L.G. Kazovsky, "Optical Heterodyning Versus Optical Homodyning: A Comparison," *J. Optical Communications*, vol. 6, no. 1, pp. 18-24, 1985.

21. S. Saito and Y. Yamamoto, "Direct Observation of Lorentzian Lineshape of Semiconductor Laser and Linewidth Reduction with External Grating Feedback," *Electronic Letters*, vol. 17, no. 9, pp. 325-327, April 1981.

22. L.G. Kazovsky, "Impact of Phase Noise on Optical Heterodyne Communication Systems," *J. Optical Communications*, vol. 7, no. 2, pp. 66-78, 1986.

23. I. Garrett and G. Jacobsen, "Theoretical Analysis of Heterodyne Optical Receivers for Transmission Systems Using (Semiconductor) Lasers with Nonnegligible Linewidth," *IEEE J. Lightwave Technology*, vol. LT-4, pp. 323-334, March 1986.

24. K. Kikuchi, T. Okoshi, M. Nagamatsu, and N. Henmi, "Degradation of Bit-Error Rate in Coherent Optical Communications Due to Spectral Spread of the Transmitter and the Local Oscillator," *IEEE J. Lightwave Technology*, vol. LT-2, pp. 1024-1033, December 1986.

25. L.G. Kazovsky, "ASK Multiport Optical Homodyne Receivers," *J. Lightwave Technology*, vol. LT-5, no. 6, pp. 770-791, June 1987.

26. J. Salz, "Modulation and Detection for Coherent Lightwave Communications," *IEEE Communications Magazine*, vol. 24, no. 6, pp. 38-49, June 1986.

27. L.G. Kazovsky, L. Curtis, W.C. Young, and N.K. Cheung, "All-Fiber 90° Optical Hybrid for Coherent Communications," *Applied Optics*, vol. 26, pp. 437-439, February 1, 1987.

28. D. Smith, "Techniques for Multigigabit Coherent Optical Transmission," *IEEE J. Lightwave Technology*, vol. LT-5, no. 10, pp. 1466-1478, October 1987.

29. R. Welter and L.G. Kazovsky, Phase Diversity Homodyne Receiver with the Linewidth-to-Bitrate Ratio of 0.5, Bell Communications Research, Red Bank, NJ, August 1987.

30. T.G. Hodgkinson, R.A. Harmon, and D.W. Smith, "Demodulation of Optical DPSK Using In-Phase and Quadrature Detection," *Electronic Letters*, vol. 21, no. 19, pp. 867-868, September 12, 1985.