AT&T Technical Journal
Vol. 64, No. 10, December 1985
Printed in U.S.A.

Coherent Lightwave Communications

By J. SALZ*

(Manuscript received March 29, 1985)

The chief objective of this paper is to develop a fundamental understanding of the effects of laser phase noise on the performance of coherent lightwave communication systems. A comprehensive treatment applicable to a wide variety of coherent receiver designs under a broad range of conditions is provided. Our models and analytical tools are developed in sufficient detail to encompass a broad range of applications. Formulas are derived for the bit error rate in homodyne and heterodyne Phase Shift Keying (PSK), Differential Phase Shift Keying (DPSK), Frequency Shift Keying (FSK) and on-off keying. Estimates are provided of the penalties accrued due to phase noise. Based on detailed mathematical analysis and estimates, we made several findings. Near quantum-limited receiver sensitivity can be achieved with PSK using homodyne detection only at signaling rates 3000 times greater than the laser linewidth. A receiver sensitivity 3 to 6 decibels poorer than the quantum limit can be achieved with heterodyne rather than homodyne detection. DPSK, for example, can operate at rates only 300 times greater than the laser linewidth. At lower rates, FSK is an attractive candidate. It can be designed to be extremely tolerant of phase noise by using wide frequency deviations.

I. INTRODUCTION

Despite the rapid advance of lightwave technology over the past decade, the basic operation of an optical fiber communications link has remained essentially unchanged. Direct modulation of the source (on-off keying) and direct detection at the receiver using a pin diode or Avalanche Photodiode (APD) have been the mainstays of lightwave

* AT&T Bell Laboratories.

systems since their infancy. Recent advances in lightwave components, however, now permit significant improvements on this time-honored approach. For example, external modulation of the laser with an electro-optic device has recently produced better long-distance performance at very high bit rates ($\gtrsim$4 Gb/s) than direct modulation of the laser itself. Another promising technique now being pursued in research laboratories around the world is the use of coherent lightwave—the optical analog of superheterodyne radio reception. Here we provide a comprehensive analytical treatment applicable to a wide variety of coherent receiver designs under a broad range of conditions. Recognizing that not all contingencies can be covered explicitly, we have endeavored to develop our model and analytical tools in sufficient detail so they can be applied to other cases of interest.

Unlike direct detection, where the optical signal is converted directly into a demodulated electrical output, the coherent receiver first adds to the signal a locally generated optical wave and then detects the sum. The resulting photocurrent is a replica of the original signal, translated down in frequency from the optical domain ($\sim 10^5$ GHz) to the radio domain ($\lesssim$ few GHz), where conventional electronic techniques can be used for further signal processing, such as filtering and demodulation. This method offers significant improvements in receiver sensitivity and wavelength selectivity compared with direct detection. In the 1.3- to 1.6-μm lightwave band, for example, an ideal coherent receiver requires a signal energy of only 10 to 20 photons per bit to achieve a bit error rate of 10^{-9}—far less than the roughly 1000 photons required by today's APDs. And because of its improved selectivity, a coherent receiver might permit wavelength-division-multiplexed systems with channel spacings of only, say 100 MHz, instead of the 100 GHz required with conventional optical multiplexing technology. A further advantage of coherent reception, not often cited but potentially very important, is that it allows the use of electronic equalization to compensate for the effects of optical pulse dispersion in the fiber.

The possible advantages of coherent optical communications have been explored for numerous applications. Much of the earlier work emphasized space communications, where highly collimated laser beams could be used to span enormous distances.[1,2] More recently, the use of coherent techniques in optical fiber systems has received considerable attention. Especially at bit rates above 2 Gb/s, where APD performance begins to deteriorate, the high sensitivity and potentially broad bandwidth of coherent receivers is a powerful stimulus to further research. As part of this effort, theroretical analyses of several types of coherent receivers have been published in the literature.[3-5] (Since these investigations are generally based on fundamental equations and

hardware designs closely resembling those encountered in the microwave domain, it is perhaps not too surprising that their general conclusions are also similar. In both cases, for example, the most energy-efficient binary system uses phase shift keying and coherent demodulation.) The performance of several experimental receivers has been compared with theory, and the agreement is generally good, especially when HeNe or YAG lasers are used for the optical sources. In practical coherent lightwave systems, however, it is expected that semiconductor injection lasers will be used; when these have been employed in coherent receiver experiments, measured sensitivity is almost always degraded. The effect is most pronounced in angle-modulation experiments, where receiver performance is often so poor that low error rates ($<10^{-9}$) cannot be achieved at all.[6,7] The cause of this degradation has been identified as laser phase noise, an impairment that is particularly severe in semiconductor devices. The effect of this noise mechanism is to impress random phase modulation on the otherwise monochromatic output of the laser, thereby impairing its performance in angle-modulation experiments. A fundamental understanding of this impairment and its effects on performance is the primary objective of this paper.

Laser phase noise is usually characterized in terms of the linewidth of the laser emission spectrum, (a readily measurable quantity that is directly proportional to the spectral density of the underlying phase noise process.) As was implied above, the linewidths available with today's distributed feedback semiconductor lasers, typically 5 to 50 MHz, are too broad to take full advantage of coherent techniques. Consequently the realization of a stable, reliable narrow-linewidth source is an extremely high priority in lightwave research. Several promising techniques have been demonstrated in the laboratory, but their usefulness under actual field conditions has yet to be established. Since reducing laser linewidth appears to be a difficult task, it is important to understand the effects of this impairment in order to establish precisely how much reduction is required.

The paper begins with an executive summary. Section II provides a brief review of direct detection methods and fundamental limits. Properties of phase noise in lasers are reviewed in Section III. Analysis of phase-lock technique are presented in Section IV. Frequency Shift Keying (FSK) is treated in Section V while Differential Phase Shift Keying (DPSK) and on-off-keying are treated in Sections VI and VII.

1.1 Executive summary

A coherent lightwave receiver is the optical analog of a superheterodyne radio set. Instead of detecting photons directly, the coherent receiver first converts the incoming signal from the optical regime

down to the radio regime, and then uses conventional electronic circuitry to perform various signal processing operations, such as amplification and demodulation. In principle, this technique can yield large increases (~20 dB) in receiver sensitivity compared with direct detection using today's avalanche photodiodes. Indeed, in the 1.3- to 1.6-μm lightwave band, coherent receivers offer the only realistic hope of approaching the so-called "quantum limit" of receiver sensitivity: ~10 photons/bit at 10^{-9} error rate. To date, however, the performance of experimental coherent receivers (especially those employing semiconductor lasers) has fallen short of the idealized theoretical predictions. One of the prime causes of this degradation has been identified as laser phase noise, a phenomenon that is known to be particularly serious in semiconductor devices. And since semiconductor lasers are, at present, the preferred candidates for coherent systems applications, it is imperative to develop an understanding of their noise properties. Our goal in this paper is to present a comprehensive treatment of the deleterious effects of laser phase noise in digital lightwave systems, so that the resulting degradation can be understood and predicted.

Laser phase noise is a random process driven by spontaneous emissions within the laser cavity, which cause the phase of the optical output wave to execute a random walk away from the value it would have had in the absence of spontaneous emission. This random phase process manifests itself as a broadening of the laser emission spectrum; it is the cause of the broad linewidth (typically 5 to 100 MHz) of today's InGaAsP Distributed Feedback (DFB) lasers. In communication systems, phase noise degrades performance because unwanted phase fluctuations in the received wave impair the demodulation process, especially when Phase Shift Keying (PSK) is used. At low signaling rates, the accumulated phase "wander" during a signaling interval might be so great that PSK cannot be used at all. In general, however, as the bit rate is increased, the impairment due to phase noise can be made negligibly small.

1.2 The central question

The central question addressed in this paper is, How high must the signaling rate be in order to ensure tolerable system degradation due to phase noise? The answer, not surprisingly, depends on system design constraints. For example, if one requires quantum-limited receiver sensitivity, then PSK with homodyne detection must be used. Based on detailed mathematical analysis and estimates we conclude that the degradation or penalty due to phase noise can be kept small (<1 dB) only if the ratio of signaling rate to laser linewidth, R/B_L, is greater than 3000. For a laser with B_L = 10 MHz, this condition implies a signaling rate of 30 Gb/s—well outside the range of current

technology. To operate at lower rates, one might use any of several techniques available for narrowing laser linewidth, but only at the price of a substantial increase in complexity. With the theoretical guidelines presented in this paper, the design engineer can strike a balance between the cost of linewidth reduction and the value of improved system performance.

If a system design can tolerate a receiver sensitivity 3 to 6 dB poorer than the quantum limit, then considerable robustness against phase noise can be achieved by using heterodyne, rather than homodyne, detection. With heterodyne differential detection of PSK, for example, the phase-noise penalty is less than 1 dB for $R/B_L \gtrsim 300$, an order-of-magnitude improvement over the homodyne case. Finally, we consider the intriguing case of FSK, which can be made extremely tolerant of phase noise by using very wide transmitter frequency deviation. At moderate bit rates ($\lesssim$500 Mb/s), where direct-deviation laser FSK transmitters operate fairly well, this modulation technique appears to be most attractive.

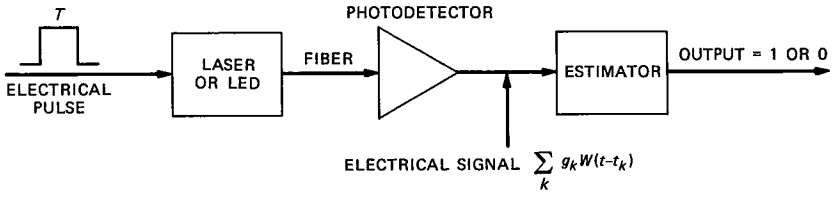
II. REVIEW OF DIRECT DETECTION AND FUNDAMENTAL LIMITS

2.1 Direct detection

Before commencing our principal investigation of coherent techniques, we briefly review some results related to direct detection of light signals.

Direct detection of light pulses implies a photodetector that converts light energy to electrical signals. The detection mechanism is based upon photon counting, which is subject to statistical fluctuations. More specifically, the photon counting process is a time-varying Poisson process whose intensity function $\lambda(t)$ is directly proportional to the information-bearing data wave.

In the case of binary transmission, the choice between a one or a zero is translated into the presence or absence of a burst of optical energy. As an illustration, consider the passage of a single pulse through an ideal transmission model depicted in Fig. 1. In the case of a one being transmitted, a square electrical signal turns on the laser or LED and energy is sent into the fiber. In the photodetector, light will be detected due to the electromagnetic energy present. Exactly when in time the photons register on the detector is random. The actual electrical current at the output of this device caused by a photon is a wideband pulse $g \cdot w(t)$ (which is very narrow compared with the signal duration T), where g (gain) is an integer-valued random variable or $g = 1$, depending on whether an (APD) or a pin diode is used. In practical systems where amplification of weak signals is required, APD's are invariably used.



(t_k) ARE RANDOM ARRIVAL TIMES, POISSON DISTRIBUTED WITH A MEAN ARRIVAL RATE $\lambda(t)$ PHOTONS/s

(g_k): AVALANCHE GAIN OF PHOTODETECTOR

$E(n) = \lambda T$ = EXPECTED NUMBER OF PHOTONS/BIT

$Pe = \frac{e^{-\lambda T}}{2} = \frac{1}{2} e^{-2P}$, P = OPTICAL ENERGY

Fig. 1—On-off direct detection.

Assuming that superposition holds for optical fiber transmission, the single-pulse description may be extended to an entire data wave. If one transmits a sequence of on or off pulses, then the *received signal*, defined as the electrical output of the photodetector on which processing is performed, is written as

$$I(t) = \sum_n g_n w(t - t_n), \tag{1}$$

where the time points $\{t_n\}$ form a Poisson process having intensity function $\lambda(t)$, with

$$\lambda(t) = \sum_n a_n h(t - nT) \tag{2}$$

and $h(t)$ is a square pulse, $\{a_n\} = 0$ or 1 are the data levels, $\{g_n\}$ is avalanche gain, T = signaling interval, and $w(t)$ = output pulse of the photodetector.

In this simple model, to detect the jth bit, one integrates the output of the photodetector over the jth T-second interval and compares the random variable with a threshold. If the output is greater than the threshold, a one is declared; if it is less, a zero is declared.

In the ideal situation, when a pin diode is used, $g_n = 1$ and when the threshold is set at zero, the average output of the integrator will yield $\int_0^T \lambda(t)dt = \lambda T$ when a one is sent and zero output when a zero is sent. Since the number of counts n with intensity λT is Poisson distributed

$$p(n) = \frac{(\lambda T)^n e^{-\lambda T}}{n!}, \tag{3}$$

and the chance of making an error is just $1/2\, p(n = 0)$ or,

$$Pe = \frac{1}{2} e^{-\lambda T}. \tag{4}$$

The average optical energy, photons per bit, is just $P = 1/2\ (\lambda T) + 1/2(0)$ and so (4) is written

$$Pe = \frac{1}{2}\ e^{-2P}. \tag{5}$$

This is a fundamental limit on the bit error rate and is commonly referred to as the "quantum limit."

Equation 5 implies that in order to obtain an error rate of 10^{-9}, about 10 photons/bit are required. This of course is the error rate achieved in the absence of coding. It has been shown recently[8] that by employing coding the number of required photons per bit is on the order of 2 to 3 provided the information rate is less than a characteristic rate called channel capacity.

When an avalanche detector is used to gain optical amplification, the average value of $\{g_n\}$ may be large but the fluctuations are also large causing amplitude jitter. The penalties incurred by avalanche detectors have been extensively studied.[9] Depending on the type of avalanche detectors used, the loss can be anywhere from 10 to 20 dBs from the quantum limit (see, for example, Ref. 10). Thus one of the chief motivations for turning to coherent techniques is to minimize this tremendous loss in detector sensitivity.

2.2 Homodyne direct detection and the super quantum limit

We begin the discussion of coherent techniques and their possible merits by assuming that the electromagnetic wave at the output of a laser can be represented as

$$s(t) = A \cos \omega_0 t, \tag{6}$$

where A^2 is proportional to the optical power. Now suppose that this wave is phase modulated so that a one results in $A \cos \omega_0 t$ and a zero results in $-A \cos \omega_0 t$. An ideal homodyne detector adds to the received wave a local carrier wave of amplitude equal to exactly A. So, the sum is

$$s_0(t) = (A \pm A)\cos \omega_0 t. \tag{7}$$

When the sum is detected by a photodetector (pin diode) and the output integrated for T seconds, one obtains for the average number of counts λT, either $4A^2T$ or 0. The average transmitted optical energy in this case is $P = A^2T$ and so the probability of a bit error is

$$Pe = \frac{1}{2} e^{-4P}.$$

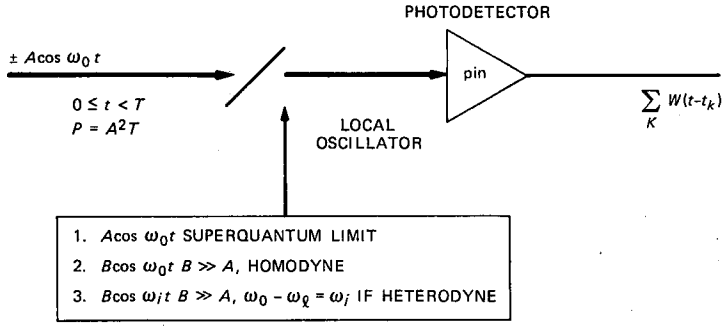


Fig. 2—Ideal homodyne and heterodyne techniques.

This result indicates a 3-dB improvement over the quantum limit, and it is often referred to as the "super quantum" limit.[11] Reviewing briefly, to achieve the super quantum limit, the local laser had to know exactly the frequency, phase, and the magnitude of the transmitter laser—a rather ambitious requirement. This detector is depicted in Fig. 2 with alternative No. 1 used as the input to the photodetector.

Now suppose that we relax the requirements on the local laser and permit its intensity to be any value B, but still requiring knowledge of the transmitted carrier frequency and phase. Now the combined waves become

$$s_0(t) = (B \pm A)\cos \omega_0 t. \tag{8}$$

Again (8) is detected by a photodetector and consequently the average number of counts at the output after integration is now $(B \pm A)^2T$, where B is the amplitude of the local laser and it is assumed that $B \gg A$.

To estimate the resulting bit error rate in this situation we invoke a limit theorem. The theorem has to do with the conditions under which a "shot noise" process—the output current from the photodetector—is well approximated by a "white gaussian" noise process. The main requirement is that the rate of photon arrivals be large. Since B in (8) can be made as large as one desires, the average number of photons is proportional to $\lambda T = (B^2 + A^2 \pm 2AB)T$. If the common bias term $(B^2 + A^2)T$ is subtracted from λT, there is left an antipodal signal pair $\pm 2ABT$ for the net average counts corresponding to reception of binary ones and zeroes. The variance of the resulting Poisson process also equals λT and since $B \gg A$ by hypothesis, the variance is essentially TB^2. Now in the limit of large number of counts due to the addition of the local laser to the incoming signal, the output electrical signal can be modeled by

$$s_0(t) = \pm 2AB + n(t), \qquad 0 \le t \le T, \tag{9}$$

where $n(t)$ is a white Gaussian noise process with double-sided spectral density equal to B^2. Integrating (9) from 0 to T, results in a Gaussian random variable. The resulting bit error rate is then

$$P_e = \frac{1}{2}\operatorname{erfc}\sqrt{2A^2T} \sim e^{-2A^2T}$$

$$= e^{-2P}, \tag{10}$$

which is asymptotically (large P) the quantum limit. We have thus demonstrated that an ideal homodyne detector using a pin photodiode achieves the quantum limit. This is made possible by the availability of large "local" optical power that provides indirect amplification of the incoming weak optical signal. While providing amplification, the procedure also produces additive noise. This mode of detection is depicted in Fig. 2 with alternative No. 2 for the input to the photodetector.

2.3 Ideal heterodyne detection

Finally we consider a detection technique where, instead of translating the incoming optical wave directly to baseband, it might be advantageous in some cases to make a frequency translation to an Intermediate Frequency (IF). This procedure is called heterodyne reception and it is depicted in Fig. 2 with alternative No. 3 for the input to the photodetector.

To understand the consequences of this approach we proceed as follows. Let the local laser frequency be denoted by ω_ℓ and the incoming optical frequency by ω_0 such that the IF frequency is $\omega_i = \omega_0 - \omega_\ell$. The addition of the two waves now results in

$$s(t) = \pm A \cos \omega_0 t + B \cos \omega_\ell t, \qquad 0 \le t \le T, \tag{11}$$

where we denote the phase modulation by ± 1 and again require that $B \gg A$.

Expressing $s(t)$ in terms of the envelope and phase about ω_ℓ results in the representation

$$s(t) = E(t)\cos(\omega_\ell t + \beta(t)), \tag{12}$$

where

$$E(t) = \sqrt{(B \pm A \cos \omega_i t)^2 + A^2 \sin^2 \omega_i t}, \tag{13}$$

and

$$\beta(t) = \tan^{-1} \frac{\pm A \sin \omega_i t}{B \pm A \cos \omega_i t}. \tag{14}$$

Table I—Ideal Performance

1. Super homodyne	e^{-4P}
2. Homodyne	e^{-2P}
3. Heterodyne	e^{-P}

$P = A^2T$ = Energy per bit.

The response of the photodiode to the wave (12) is again a shot-noise process with intensity function, λ_0, equal to the envelope squared.

$$\lambda_0(t) = B^2 + A^2 \pm 2AB \cos \omega_i t \tag{15}$$

Using the same limit arguments as in the previous section, we first subtract $B^2 + A^2$ from $\lambda_0(t)$, which retains the antipodal signal pair

$$\pm 2AB \cos \omega_i t. \tag{16}$$

Because $B \gg A$ the fluctuating noise is white Gaussian with double-sided spectral density $\cong B^2$. Denoting the additive noise by $n(t)$, the equivalent signal-in-noise problem after heterodyning becomes

$$s_0(t) = \pm AB \cos \omega_i t + n(t), \qquad 0 \le t \le T. \tag{17}$$

This is a standard elementary detection problem and deciding whether a plus or a minus was sent is accomplished by multiplying (17) by $\cos \omega_i t$, integrating for T seconds, and comparing the result with a threshold set to zero. The decision statistic is

$$\pm ABT + \int_0^T n(t) \cos(\omega_i t) dt, \tag{18}$$

where we have neglected the double frequency term. Since the random variable $\int_0^T n(t) \cos(\omega_i t)\, dt$ has variance equal to $B^2T/2$, the bit error rate in this case is asymptotically

$$P_e \sim e^{-\frac{A^2B^2T^2}{B^2T}} = e^{-A^2T} = e^{-P}. \tag{19}$$

The exponent is seen to be a factor of 2 smaller than in (10) and because of this heterodyne detection is 3 dB inferior to the quantum limit. The asymptotic performance of the ideal frequency translation methods just discussed are summarized in Table I.

With these preliminaries we are now in a position to discuss more realistic detection systems, where laser phase noise must be taken into account. Before doing this however, we briefly review the origins of this noise.

III. PHASE NOISE IN LASERS

Phase or frequency noise in lasers is a well-known and documented phenomenon that sets fundamental limitations on the performance of

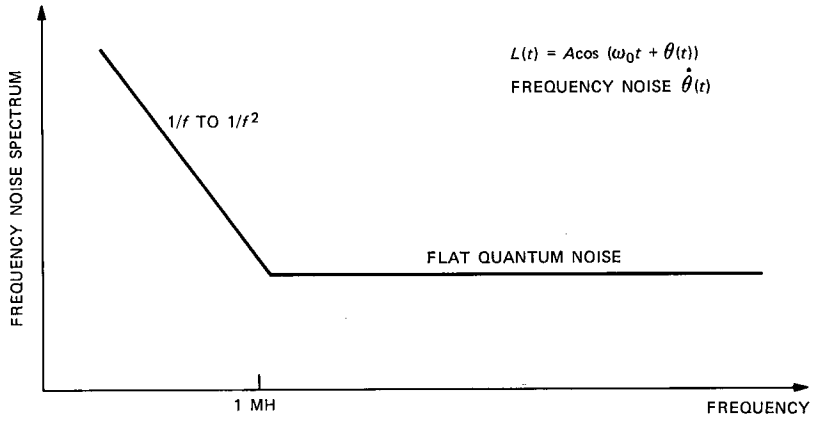


Fig. 3—Laser phase noise.

coherent optical communications.[12-14] It has been observed that the spectral density of this frequency noise has a $1/f$ to $1/f^2$ characteristic up to around 1 MHz, and is flat for frequencies above 1 MHz,[15] as shown in Fig. 3. The flat, or white, component is associated with quantum fluctuations and is the principal cause for line broadening. From a communications theory point of view, the relatively low-frequency components can be easily tracked, and so we shall not dwell on this part of the noise.[16] Our main focus here will be on the white component.

Laser phase noise is caused by randomly occurring spontaneous emission events, which are an inevitable aspect of laser operation. Each event causes a sudden jump (of random magnitude and sign) in the phase of the electromagnetic field generated by the device. As time evolves the phase executes a random walk away from the value it would have had in the absence of spontaneous emission. The mean-squared phase deviation grows approximately linearly with time, and since the average time between steps in the random walk becomes vanishingly small, the random phase $\theta(t)$ becomes in the limit a Wiener process characterized by a zero-mean, white Gaussian frequency noise $\mu(t)$ with two-sided spectral density N_0.[17] Thus, the phase process is represented as

$$\theta(\tau) = 2\pi \int_0^\tau \mu(t)dt, \tag{20}$$

and the mean-squared phase deviation is

$$E\theta^2(\tau) = E\left[2\pi \int_0^\tau \mu(t)dt\right]^2 = (2\pi)^2 N_0 \tau, \tag{21}$$

where E denotes mathematical expectation.

To determine the parameter N_0 (which is a function of both the laser structure and the operating conditions), one can measure the spectral density of the frequency fluctuations in the emitted light and hence determine N_0 directly. Experiments of this sort have shown that the representation in (20) using the white-noise approximation for $\mu(t)$ is reasonably accurate for 0.1 ns $\lesssim \tau \lesssim$ 1 μs, which is adequate for our present purposes. Another technique for measuring N_0 makes use of the fact that phase noise causes an observable broadening of the laser emission spectrum. In effect, the accumulated phase error given by (21) limits the duration of temporal coherence of the laser radiation to an interval of roughly $1/(2\pi)^2 N_0$; the corresponding linewidth is therefore proportional to the noise density N_0. The following discussion makes this relationship more precise.

Consider the sine-wave random process,

$$s(t) = A \cos(2\pi f_0 t + \theta(t) + \phi), \tag{22}$$

where the innocuous inclusion of the uniform phase ϕ renders $s(t)$ a stationary process with correlation function,

$$\begin{aligned} R(\tau) &= Es(t)s(t+\tau) \\ &= \frac{A^2}{2} R_e \left\{ e^{i2\pi f_0 \tau - \frac{(2\pi)^2}{2} N_0 |\tau|} \right\}. \end{aligned} \tag{23}$$

A simple calculation reveals that the Fourier transform of (23), the power spectrum, is

$$G(f) = \frac{A^2}{4\pi^2 N_0} \left\{ \left(1 + \left(\frac{f+f_0}{\pi N_0} \right)^2 \right)^{-1} + \left(1 + \left(\frac{f-f_0}{\pi N_0} \right)^2 \right)^{-1} \right\}, \tag{24}$$

and a quick sanity check yields

$$\int_{-\infty}^{\infty} G(f) df = \frac{A^2}{2}, \tag{25}$$

as it should. A sketch of the baseband spectrum, commonly referred to as *Lorentzian* is shown in Fig. 4. The parameter characterizing $G(f)$, N_0, can be experimentally determined by measuring the 3-dB bandwidth of the spectrum around f_0. Denoting the total (two-sided) 3-dB bandwidth by B_L, it is seen from (24) that

$$N_0 = \frac{B_L}{2\pi}, \tag{26}$$

when $B_L \ll f_0$.

We will see later that seemingly modest amounts of phase noise can seriously degrade coherent system performance; thus it is imperative

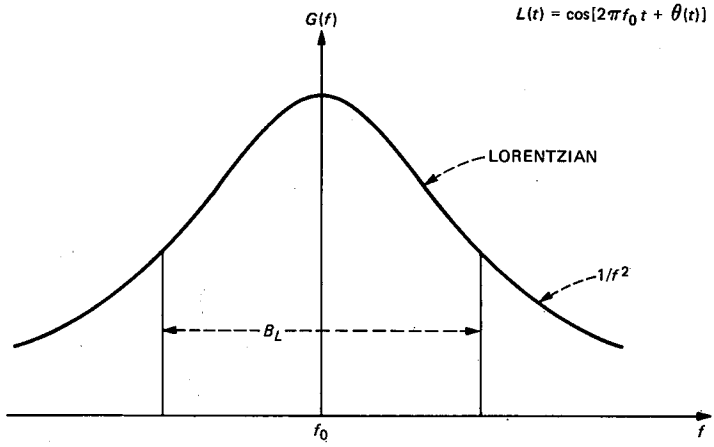


Fig. 4—Power spectrum of laser line.

to make lasers with the narrowest possible linewidth. Unfortunately, the semiconductor injection laser designs likely to be used in the 1.3- to 1.6-μm lightwave band typically have linewidths in the range 5- to 50-MHz, which is too broad for many potentially important coherent lightwave applications.[18-20] (For comparison, the reader should note that microwave oscillators, which are widely used in coherent radio applications, have linewidths on the order of 1 Hz.) To reduce laser linewidth, experimenters have exploited the fact that the noise density N_0 is inversely proportional to P_0Q^2, where P_0 is the laser output power and Q is the quality factor of the "cold" laser cavity resonance; thus high-power, high-Q lasers tend to have narrow linewidths. The most impressive line-narrowing experiments have been performed using a mirror or diffraction grating external to the laser chip to produce a composite cavity of very high Q.[21] Under relatively benign laboratory conditions, linewidths of tens of kilohertz have been obtained: an improvement of three orders of magnitude! Whether this approach will prove practical under harsh field conditions remains to be seen. In any case, it appears that phase noise will be an important consideration for the foreseeable future, so we turn now to developing a clear understanding of its consequences.

IV. PHASE-LOCK TECHNIQUES

We saw in Section III that homodyne detection of an optical PSK data wave makes it possible to achieve the quantum limit. However, to gain the full benefit of this approach the local laser must have perfect knowledge of the transmitted optical center frequency and phase. In this section we explore the possibility of deriving these

crucial parameters from either the optical data wave directly or from a heterodyned version.[22] At microwave frequencies, carrier recovery techniques are well established, while at optical frequencies, the available methodologies are still limited. For example, it is very difficult to directly multiply two optical signals or square an optical wave, which makes it difficult to wipe off the binary modulation. In attempting to derive carrier in the optical frequencies, it is, therefore, necessary to resort to data-aided techniques.[23]

4.1 Optical phase-locked loop

The question we explore here is the possibility of estimating or tracking the phase of the incoming optical wave so that it can be used to coherently demodulate PSK. Making use only of direct intensity detection, a proposed homodyne optical detector is depicted in Fig. 5.

At the input to the detector, the power of the incoming optical data signal is split so that a fraction, A^2k^2, is devoted to the estimation of the phase process $\theta(t)$ by a phase-locked loop, and the remaining portion of the power, $A^2(1-k^2)$, is used for demodulation. The power division that is determined by the choice of the constant $0 \leq k \leq 1$ is a parameter that must be optimized. In the phase-locked loop we presume that the local Voltage Controlled Oscillator (VCO) can be

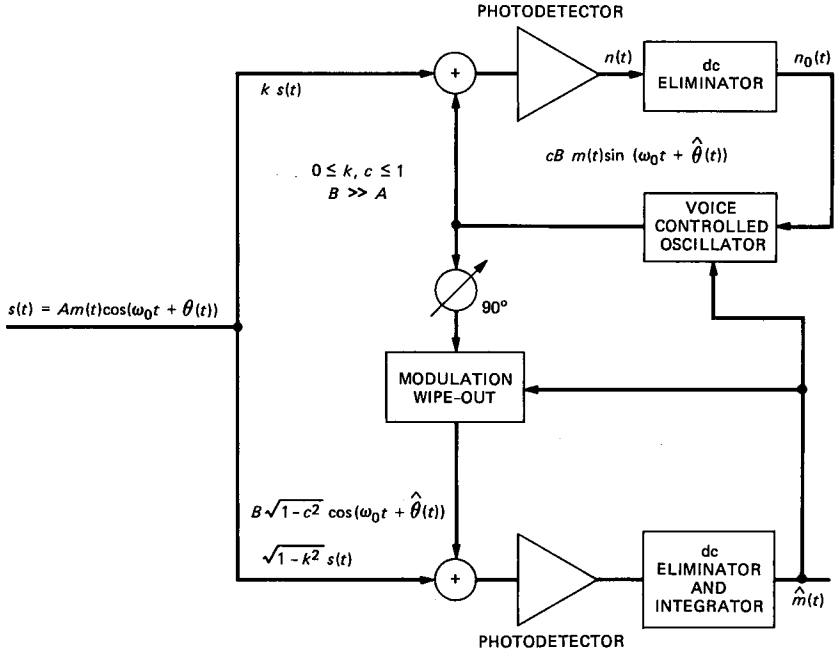


Fig. 5—An optical homodyne detector.

modulated by the estimated data $\hat{m}(t)$ using decision directed techniques, and consequently the modulation can be locally wiped off and a modulation-free carrier thus made available in the lower portion of the figure to perform the homodyne demodulation function. We also assume that the optical center frequency can be identically matched by the local laser. In any event, if this is not the case, an additional phase-locked loop may be needed to track this mismatch.[16] Making these assumptions, it is possible to analyze the performance of this detector and to obtain tight bounds on the degradation from ideal homodyne performance.

For the subsequent analysis we refer to Fig. 5. Adding a fraction c of the VCO output to a fraction of the incoming optical data wave we obtain

$$V(t) = Akm(t)\cos(\omega_o t + \theta(t)) + cB\hat{m}(t)\sin(\omega_o t + \hat{\theta}(t)), \quad (27)$$

where $\hat{m}(t)$ at the output of the VCO is the reconstructed data wave from past decisions, and it is assumed to be devoid of errors. The squared envelope of $V(t)$, which is the average value of the Poisson counting process $n(t)$, at the output of the photodetector is

$$E^2(t) = A^2k^2 + B^2(1 - c^2) + 2ABkc \sin \psi(t), \quad (28)$$

where $\psi(t)$ is the phase-error process

$$\psi(t) = \theta(t) - \hat{\theta}(t), \quad (29)$$

and we used the fact that $\hat{m}(t)m(t) = m^2(t) = 1$.

After dc elimination, the "signal" portion of the "shot noise" process, which in the limit becomes a Gaussian process since B can be made large is

$$2ABk \cdot c \cdot \sin \psi(t), \quad (30)$$

and the resulting zero-mean white Gaussian noise process is denoted by $\nu(t)$ with spectral density equal to B^2c^2.

Thus the equivalent signal-plus-noise process, which controls the frequency of the local laser VCO is

$$u(t) = 2ABkc \sin \psi(t) + \nu(t), \quad (31)$$

and so, because of feedback, we must satisfy the equation

$$\hat{\theta}(t) = \phi(t) + K \int_0^t u(t')dt', \quad (32)$$

where $\phi(t)$ is the phase noise process of the local laser, and K is a proportionality constant to be determined later.

Subtracting from both sides of (32), $\theta(t)$, the phase of the incoming wave, and differentiating, we get the stochastic differential equation

$$\frac{d\psi}{dt} = -\frac{d}{dt}[\theta(t) - \phi(t)] + K[2ABkc \sin \psi(t)] + K\nu(t)$$

$$= K_0 \sin \psi(t) + \nu_0(t), \tag{33}$$

where

$$K_0 = 2ABKkc \tag{34}$$

and $\nu_0(t)$ is now a white Gaussian process with double-sided spectral density,

$$D = B^2K^2c^2 + 2\pi(B_{L1} + B_{L2}), \tag{35}$$

where the Lorentzian bandwidth of the transmitter laser is B_{L1} and the local laser VCO bandwidth is B_{L2}.

It is well known[24,25] that (33) obeys a Fokker-Plank equation yielding the steady-state probability density function (mod2π) for the phase-error process $\psi(t)$,

$$p(\psi) = \frac{\exp[\alpha \cos \psi]}{2\pi I_0(\alpha)}, \qquad -\pi \le \psi \le \pi, \tag{36}$$

where $I_0(\cdot)$ is the zeroth-order modified Bessel function, and where

$$\alpha = \frac{2K_0}{D} = \frac{4AkK_e}{K_e^2 + 2\pi B_L}, \tag{37}$$

$$K_e = BK\sqrt{1 - c^2}, \tag{38}$$

and

$$B_L = (B_{L1} + B_{L2}).$$

The probability density (36) is sharply peaked at $\psi = 0$ when α is large and becomes flat, or uniform when α is small. One strives therefore to design the phase-locked loop so that α is as large as needed to obtain minimum degradation from ideal ($\psi = 0$). For fixed A and k, (37) reveals that α cannot be made arbitrarily large because of the finite Lorentzian bandwidth B_L. However, there exist a maximum value of α (when $K_e^2 = 2\pi B_L$) given by,

$$\alpha_0 = \frac{2Ak}{\sqrt{2\pi B_L}} = k\sqrt{\frac{2PR}{\pi B_L}} \tag{39}$$

where $P = A^2T$—the transmitted optical energy in the received signal. Equation (39) reveals that for fixed optical energy and k, α_0 can be made large only by increasing the ratio R/B_L.

4.2 Performance

We now examine the performance of this optical homodyne detector using the phase-locked loop output as the reference carrier wave.

In the lower portion of Fig. 5 the sum signal $W(t)$ is

$$W(t) = A\sqrt{1-k^2}\, m(t)\cos(\omega_0 t + \theta(t)) + B\sqrt{1-c^2}\cos(\omega_0 t + \hat{\theta}(t)). \quad (40)$$

The squared envelope of $W(t)$—the response of the photodetector—is therefore

$$E^2(t) = A^2(1-k^2) + B^2(1-c^2) + 2AB\sqrt{1-k^2}\sqrt{1-c^2}\, m(t)\cos\psi(t). \quad (41)$$

The resulting shot-noise process at the output of the photodetector again becomes in the limit a Gaussian process with average value (after dc elimination) equal to

$$2AB\sqrt{1-k^2}\sqrt{1-c^2}\, m(t)\cos\psi(t), \quad (42)$$

and zero-mean white Gaussian noise $\nu(t)$ with a spectrum equal to

$$B^2(1-c^2).$$

Thus the equivalent signal plus noise prior to integration is

$$S(t) = 2A\sqrt{1-k^2}\, m(t)\cos\psi(t) + \nu_0(t), \quad (43)$$

where we have divided signal plus noise by $B\sqrt{1-c^2}$ thus normalizing the spectrum of $\nu_0(t)$ to unity.

Integrating (43) over a T-second interval results in the decision statistic

$$s_0 = \pm 2A\sqrt{1-k^2}\int_0^T \cos\psi(t)dt + \int_0^T \nu_0(t)dt, \quad (44)$$

or

$$s_0 = \pm\rho\xi + \bar{\nu}_0, \quad (45)$$

where $\bar{\nu}_0$ is now a zero-mean Gaussian random variable with unit variance, the random variable

$$\xi = \frac{1}{T}\int_0^T \cos\psi(t)dt,$$

and $\rho = 2AT^{1/2}\sqrt{1-k^2}$.

Because of symmetry, the probability of error is

$$Pe = \Pr[-\rho\xi + \bar{\nu}_0 \geq 0]. \quad (46)$$

The exact evaluation of this probability is intractable even numer-

ically since it requires knowledge of the nth-order probability distribution of the phase error process $\psi(t)$. The most we have, however, is the first-order distribution, and therefore we must resort to upper bounds using the only information we have.

In Appendix A, an exponential upper bound on (46) is developed with the result

$$Pe \leq \begin{cases} g(\alpha)e^{-\frac{\rho^2}{2}}, & \alpha/\rho^2 > 1 \\ g(\alpha)e^{-\rho^2/2[2\alpha/\rho^2-(\alpha/\rho^2)^2]}, & \alpha/\rho^2 \leq 1, \end{cases} \tag{47}$$

where the coefficient, $g(\alpha)$, is defined in Appendix A, eq. (178).

When $\alpha/\rho^2 = 1$, (47) is reduced to a single bound,

$$Pe \leq g(\alpha)e^{-\frac{\rho^2}{2}}. \tag{48}$$

Recall that

$$\rho^2 = 4P(1 - k^2), \tag{49}$$

and

$$\alpha = k\sqrt{\frac{2PR}{\pi B_L}}, \tag{50}$$

and so the threshold parameter becomes

$$r = \frac{\alpha}{\rho^2} = \frac{1}{4\sqrt{\pi}}\sqrt{\frac{\gamma}{P}}\frac{k}{1-k^2}, \tag{51}$$

where γ is the ratio of the signaling rate to the average Lorentzian laser bandwidth

$$\gamma = \frac{2R}{B_L}. \tag{52}$$

It is seen from (47) that the error exponent assumes two crucially different forms depending on whether $r \geq 0$ or $r < 0$. For fixed γ and k, the exponent is linear in P when $r \geq 1$, while it behaves as the square root of P for $r < 1$. Thus the probability of error decays much more slowly with P when $P >$ constant γ [eq. (51)], indicating a threshold for Pe versus P. For fixed γ and P, however, r is a monotonically increasing function of k, $0 < k < 1$, and so there exists a value $k = k_o$, such that $r(k_o) = 1$, given by

$$k_o = \sqrt{x^2 + 1} - x, \tag{53}$$

where

$$x = \frac{1}{8\sqrt{\pi}} \sqrt{\frac{\gamma}{P}}.$$

At this value of k the exponent is

$$E(k_o, P) = -2P(1 - k_o^2). \tag{54}$$

This value of k is not, however, the optimum value that makes the negative exponent the greatest, or the probability of error the smallest. The optimum value of k, or the power division, is found by setting the derivative of the exponent in (47) for $r < 1$ to zero. From (47) the negative exponent is

$$E(k) = (1 - k^2)[2r(k) - r^2(k)], \tag{55}$$

and

$$\frac{dE}{dk} = 2(1 - k^2)\left[\frac{dr}{dk} - r\frac{dr}{dk}\right] - 2k[2r - r^2]. \tag{56}$$

Note that

$$\left.\frac{dE}{dk}\right|_{k=k_o} = -2k_o,$$

indicating that there exist a $k_{op} \leq k_o$ which increases the exponent. Setting (55) to zero, we get the formula for the optimum k,

$$k_{op}^2 = 1 - r(k_{op}). \tag{57}$$

Substituting (51) into (57) with the definition of x in (53) we obtain explicitly,

$$k_{op}^2 = 1 - 2x\frac{k_{op}}{1 - k_{op}^2}. \tag{58}$$

One can now proceed to solve (58) numerically and use this optimum value to plot the upper bound on the probability of error versus P for different values of γ. A more incisive way to exhibit the behavior of the error rate, however, is to use the suboptimum value of $k_o > k_{op}$ given in (53), which renders $r = 1$, and then define a penalty, which is the reduction of the exponent relative to ideal performances. Since $k_{op} < k_o$, this still provides an upper bound on the error rate. It can be seen from (58) that when the term $1/(1 - k_{op}^2)$ is neglected, the resulting equation is identical to (53) and so, to this degree of approximation, $k_{op} \sim k_o$ (this is a good approximation when k_{op} turns out to be small, which should be the case when the penalty is small). Following this approach, we solve (53) for x and write the negative exponent (54) as

$$-E(\gamma, P) = 2P(1 - k_o^2) = 4Px[\sqrt{x^2 + 1} - x]. \tag{59}$$

It can be checked that when $\gamma \to \infty$ (zero phase noise), $E(\gamma, P) \to 2P$, as it should.

The penalty incurred due to the finite value of γ may now be defined as follows. Equating (59) to $2P_o$, where P_o is chosen to achieve a desired error rate, say 10^{-9}, in which case $P_o \sim 10$, one obtains a function of P_o/P versus γ, and for each value of γ one can then calculate, $-10 \log P_o/P$), which defines the penalty.

Thus, proceeding in this manner we have,

$$2P_o = 4Px\sqrt{x^2 + 1} - x, \tag{60}$$

and when

$$x = \frac{1}{8\sqrt{\pi}} \sqrt{2/P}$$

is substituted into (60), one gets the penalty function

$$b = \frac{P_o}{P} = \frac{\gamma}{\gamma + 16\pi P_o}, \tag{61}$$

as $\gamma \to \infty$, $G = 1$.

In Fig. 6 we plot (61) versus γ for two different values of P_o. One

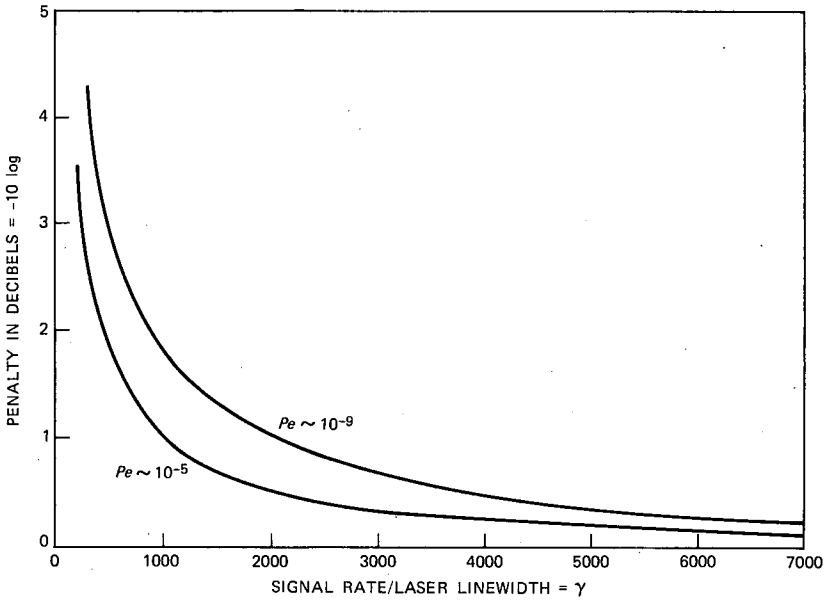


Fig. 6—Penalty versus signaling rate divided by linewidth in homodyne PSK detection at two different ideal error rates for identical laser linewidths.

corresponding to $Pe \sim 10^{-9}$ and the other to $Pe \sim 10^{-5}$. It can be seen that negligible penalty is incurred when $\gamma \geq 3000$ for $Pe \sim 10^{-9}$ and 3 dB is given up at $\gamma \sim 500$. We will see in the next section that a heterodyne Phase-Locked Loop (PLL) also suffers a 3-dB penalty at around $\gamma \sim 500$. This approach is analyzed in the next section.

4.3 Heterodyne phase-locked Loop

Heterodyning the optical data wave down to an IF frequency, and then deriving the carrier from the resulting microwave signal using standard well-known techniques may make it easier to wipe off the modulation, and consequently ease the signal processing burden of the phase-locked loops.[26,27] In the subsequent analysis we presume that the modulation has been eliminated, and as such the phase-locked loop can operate directly on the IF carrier wave. So, after heterodyning the optical signal

$$Am(t)\cos(\omega_o t + \theta(t)) \tag{62}$$

to an IF frequency f_i and wiping off the modulation, we obtain the microwave signal plus noise

$$s(t) = 2A\cos(2\pi f_i t + \theta(t)) + n(t), \tag{63}$$

where A^2 equals optical power, $n(t)$ is again a white Gaussian noise process with unit double sided spectral density, and $\theta(t)$ is now the difference between the transmitting laser's phase noise and the local laser's phase noise. Consequently, the variance of $\theta(t)$ now is

$$\begin{aligned} E\theta^2(t) &= (2\pi)^2(N_{o1} + N_{o2})t \\ &= (2\pi)\bar{B}_L t, \end{aligned} \tag{64}$$

where as before

$$B_L = (B_{L1} + B_{L2}). \tag{65}$$

The signal (63) is now the input to a conventional PLL depicted in Fig. 7. The analysis of the PLL is straightforward. Denote the output of the PLL by

$$S_v(t) = K_1 \sin\theta'(t), \tag{66}$$

where

$$\theta'(t) = 2\pi f_i t + K_2 \int_0^t e(s)ds.$$

The output of the multiplier in Fig. 6 is

$$X(t) = 2A\cos(2\pi f_i t + \theta(t))S_u(t) + S_u(t)n(t), \tag{67}$$

where the function of the Low-Pass Filter (LPF) here is to eliminate

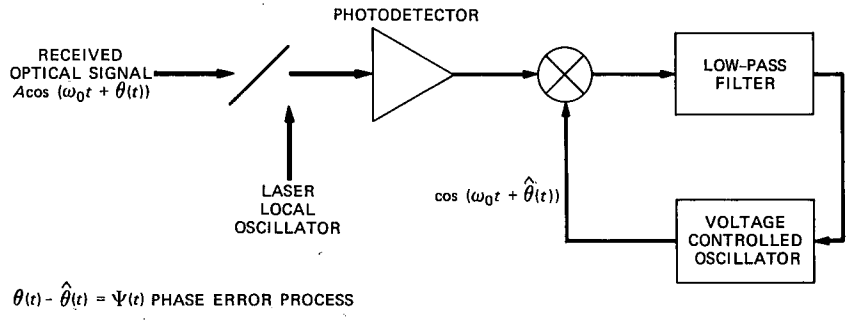


Fig. 7—Phase-lock techniques.

double-frequency terms from its output. Yet, the cut-off frequency is placed high enough so that the output noise can still be regarded as white. Thus, because of feedback, the following equation must be satisfied

$$e(t) = K_1 A \sin(\theta(t) - K_2 \int_0^t e(s)ds) + \nu(t), \tag{68}$$

where now the double-sided spectral density of $\nu(t)$ is equal to $K_1^2/2$. The phase error therefore is

$$\psi(t) = \theta(t) - K_2 \int_0^t e(s)ds, \tag{69}$$

and when this is differentiated one obtains

$$\begin{aligned} \frac{d\psi}{dt} &= \frac{d\theta}{dt} - K_2 e(t) \\ &= 2\pi(\mu_1(t) - \mu_2(t)) - K_2 e(t), \end{aligned} \tag{70}$$

where μ_1 and μ_2 are the frequency noises of the transmitter laser and the laser involved in the local heterodyner, respectively.

When (70) is substituted into (68), one again obtains the well-known stochastic differential equation governing the evolution of the phase-error process,

$$\frac{d\psi(t)}{dt} = -K_1 K_2 A \sin \psi(t), \tag{71}$$

where

$$u(t) = -2\pi(\mu_1 - \mu_2) + \nu(t)K_2 \tag{72}$$

is a white Gaussian noise process with double-sided spectrum equal to

$$D = 2\pi B_L + \frac{K_1^2 K_2^2}{2}. \tag{73}$$

Letting $K_1K_2 = K$, we observe that (71) is identical in structure to (33), which again results in a Fokker-Plank equation yielding the steady-state probability density function (mod 2π) for the phase error

$$p(\psi) = \frac{e^{\alpha}\cos\psi}{2\pi I_o(\alpha)}, \qquad -\pi \le \psi \le \pi. \tag{74}$$

The important system parameter is now given by

$$\alpha = \frac{2AK}{D} = \frac{2AK}{\frac{K^2}{2} + 2\pi B_L}. \tag{75}$$

The algebraic form of the PLL parameter α departs from the conventional form where, by decreasing the loop bandwidth, α can be made as large as possible. Here there is a minimum band resulting from the presence of phase noise. The only way that α can be increased is by increasing the input optical energy, or by decreasing phase noise relative to the signaling rate. Consequently there exists a maximum value of α (when $K^2 = 4\pi B_L$) given by

$$\alpha_{opt} = \sqrt{\frac{RP}{\pi B_L}}, \tag{76}$$

where $R = 1/T$ and $P = A^2T$—optical energy.

4.4 Performance

Now consider the modulated heterodyned wave

$$S_m(t) = \pm 2A\cos(2\pi f_i t + \theta(t)) + n(t), \qquad 0 \le t \le T, \tag{77}$$

and the estimated carrier wave from the PLL

$$\cos(2\pi f_i t + \hat{\theta}(t)). \tag{78}$$

Multiplying (77) by (78), integrating from 0 to T, and eliminating the double-frequency components results in the decision statistic

$$S = \pm A\int_0^T \cos\psi(t)dt + \int_0^T n(t)\cos(2\pi f_i t + \hat{\theta}(t))dt, \tag{79}$$

where $\psi(t)$ is the phase-error process obeying at any instant of time the probability density (74). Rewriting (79) as before,

$$S = \pm AT\xi + \nu, \tag{80}$$

where again

$$\xi = \frac{1}{T}\int_0^T \cos\psi(t)dt,$$

and where now ν is a Gaussian random variable with $E\nu^2 = T/2$.

Dividing (80) by $\sqrt{2/T}$ yields,

$$s_0 = \pm\rho\xi + \nu_0, \tag{81}$$

where $\rho = \sqrt{2T}A$ and ν_0 now has unit variance.

The probability of error is as before,

$$Pe = \Pr[-\rho\xi + \nu_0 \geq 0]. \tag{82}$$

This probability is structurally identical to (46), and we therefore use the same bound from Appendix A.

$$Pe \leq \begin{cases} g(\alpha_o)e^{-\rho^2/2}, & \alpha_o/\rho^2 \geq 1 \\ g(\alpha_o)e^{-\rho^2/2[2\alpha_o/\rho^2+(\alpha_o/\rho)^2]}, & \alpha_o/\rho^2 < 1 \end{cases} \tag{83}$$

With the definition of α_o in eq. (76) we can compute the threshold parameter in this case,

$$r_o = \alpha_o/\rho^2 = \frac{1}{\sqrt{8\pi}} \sqrt{\frac{\gamma}{P}}, \tag{84}$$

where again

$$\gamma = \frac{2R}{B_L}. \tag{85}$$

With these definitions we express (83) as

$$Pe \leq \begin{cases} g(\alpha)e^{-P}, & \gamma \geq 8\pi P \\ g(\alpha)e^{-P[2r_o-(r_o^2)]}, & \gamma \leq 8\pi P. \end{cases} \tag{86}$$

We note that the behavior of this bound is slightly different from the one applicable in the homodyne case. For one, here ideal performance is 3 dB inferior to the quantum limit, which is accounted for by the heterodyning operation, as we have already noted. Next, ideal performance with negligible degradation is attained when

$$\frac{2R}{B_L} > 8\pi P,$$

and for ratios less than this, the degradation is increased gracefully. As an example, when $P = 20$, yielding an ideal error rate $\sim 10^{-9}$, the threshold parameter $\gamma = 500$. For ratios greater than this, no penalty in optical power (from the ideal $P = 20$) is incurred. To assess the penalty at operations at less than this threshold, we use the same definition as before. The probability of error exponent when $r_o \leq 1$ is

$$E(\gamma, P) = P(2r_o - r_o^2) \tag{87}$$

and when this is equated to P_o we get the formula

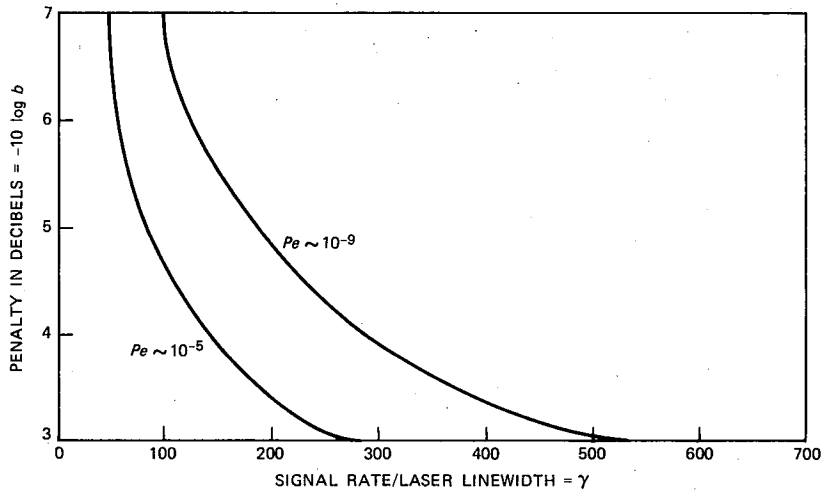


Fig. 8—Penalty versus signaling rate divided by linewidth in heterodyne PSK detection at two different ideal error rates for identical laser linewidths.

$$b = \frac{P_o}{P} = 2r_o - r_o^2 \leq 1. \tag{88}$$

Substituting the definition of r_o [eq. (84)] and solving for γ_o yields

$$\gamma = 8\pi P_o \left[\frac{(1 - \sqrt{1 - b})^2}{b}\right], \tag{89}$$

as can be seen when $b = 1$, $\gamma = 8\pi P_o$—the threshold value.

In Fig. 8 we plot b in decibels versus γ for two different values of P_o. We see that around $\gamma \sim 400$, the asymptotic degradation of 3 dB is approached. Interestingly, at around this ratio the optical PLL also degrades by 3 dB, as we have already mentioned.

In concluding this section we remark that our estimates are only upper bounds, albeit, we feel, tight bounds. The exact evaluation of the error rate is not feasible because of the nonlinear functionals that are involved. If one chooses to ignore the time integrals, than the probability of error can be evaluated numerically as has been done in Refs. 28 and 29.

In the next section we will see that noncoherent techniques such as frequency modulation and differential phase modulation, where knowledge of carrier phase is not essential, yields performance very near to what can be attained with heterodyne phase-lock technique at reasonable signaling rates.

V. BINARY FREQUENCY SHIFT KEYING

Here we discuss and analyze the performance of binary Frequency

Shift Keying (FSK) as one of the modulation options.[30–33] In this modulation method information is conveyed by switching the frequency of the laser between two different values. Thus, during a fixed interval, $T = 1/R$, where R is the signaling rate, the receiver has to decide whether

$$A \cos(\omega_1 t + \theta(t)) \tag{90}$$

or

$$A \cos(\omega_2 t + \theta(t))$$

was transmitted. In (90) A^2 is again proportional to the received optical power and $\theta(t)$ is the phase noise process associated with the laser.

It appears that this modulation method is impervious to the effects of phase noise, since the shifted frequencies ω_1 and ω_2 can be sufficiently separated and if enough bandwidth is available, crosstalk due to the fluctuating phase noise can be minimized. These are the chief reasons, then, for considering FSK. We point out that this is not what is commonly referred to as continuous-phase narrow-band FM. The latter yields the same performance as differential phase modulation treated in the next section.

The first step in the processing of the FSK optical signal is to heterodyne (90) to an IF frequency ω_i. As was already noted, this can be accomplished by adding to (90) a locally generated optical signal and then direct detecting the sum by a photodetector. The sum signal then is

$$S(t) = A \cos(\omega_l t + \theta(t)) + B \cos(\omega_l + \phi(t)), \qquad l = 1, 2, \tag{91}$$

where the IF frequencies are $\omega_i = \omega_l - \omega_f$, and $\phi(t)$ is the phase noise associated with the local laser. The output of the photodetector is again a "shot noise" process

$$I(t) = \sum_n w(t - t_n). \tag{92}$$

The squared envelope of (91) with respect to ω_f is

$$E^2(t) = A^2 + B^2 + 2AB \cos((\omega_l - \omega_f)t + \Delta(t)), \tag{93}$$

where $\Delta(t) = \theta(t) - \phi(t)$.

When the local laser intensity $B \gg A$ (66) approaches a white Gaussian process with average value equal to $\lambda(t)$ and standard deviation also equal to $\lambda(t)$. Thus the dc part of the average value of (93) is

$$S_o(t) = 2AB \cos(\omega_i t + \Delta(t)), \qquad i = 1, 2, \tag{94}$$

while the resultant zero-mean Gaussian noise, $n(t)$, has a double-sided spectral density equal to B^2.

With these preliminaries, we now confront a classical detection

problem. Given the IF signal (94), plus white Gaussian noise of unit spectral density

$$V(t) = 2A\cos(\omega_i t + \Delta(t)) + n(t), \qquad 0 \le t < T, \qquad i = 1, 2, \quad (95)$$

how does one process $V(t)$ so as to attain the least probability of error? While the problem is classical the solution is not tractable in general because of the presence of the phase noise process $\Delta(t)$.

For calibration purposes, let us first review briefly the performance under the assumption that the phase noise is effectively a constant. In the case when $\Delta(t)$ is slowly varying with respect to the rate $R = 1/T$, or when the symbol rate R is much greater than the bandwidth of the laser signal, the optimum detector has a well-known structure depicted schematically in Fig. 9. The results in this case will serve as a benchmark to which the later more general results will be compared. Also assume that the frequency shifted signals are orthogonal, i.e., the two frequencies ω_1 and ω_2 are chosen such that

$$\int_0^T \cos(\omega_1 t + \Delta)\cos(\omega_2 t + \Delta)dt = 0.$$

To proceed with the error rate analysis, assume that ω_1 was sent. In this case, we expect the x output in Fig. 8 to be greater than y, and an error is made when $x - y \le 0$. Because of symmetry, when ω_2 is sent, y is expected to be greater than x and a mistake is now made when $y - x \ge 0$. So, the probability of error is just

$$Pe = \Pr[x - y \le 0]. \quad (96)$$

In order to evaluate this probability, we express the random variables x and y as indicated by the mathematical operations in Fig. 7. It then can be seen that

$$Pe = \Pr[x_1^2 - x_2^2 + x_3^2 - x_4^2 \le 0], \quad (97)$$

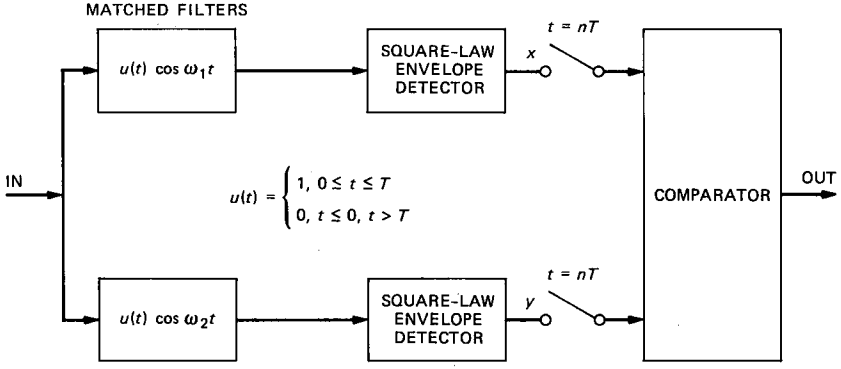


Fig. 9—Structure of the optimum FM detector when linewidth approximates zero.

where

$$x_1 = AT \cos \Delta + \int_0^T n(t)\cos \omega_1 t dt,$$

$$x_2 = \int_0^T n(t)\cos \omega_2 t dt,$$

$$x_3 = AT \sin \Delta + \int_0^T n(t)\sin \omega_1 t dt,$$

and

$$x_4 = \int_0^T n(t)\sin \omega_2 t dt.$$

The desired probability is just the probability that the difference in the lengths of two 2-dimensional Gaussian vectors is less than zero. As can be verified, the x's are independent Gaussian random variables with identical variances, $\sigma^2 = T/2$. For these random variables, (97) can be expressed exactly[34] as

$$Pe = \frac{1}{2} \exp\left[-\frac{A^2T}{2}\right]. \tag{98}$$

Comparing this with the performance of direct detection, we observe that the optical signal (90), after direct detection and integration for T seconds, yields an average photon count equal to A^2T. In this direct detection case, the chance of making an error would be the chance of detecting zero photons in T seconds. From the Poisson distribution, for the number of photons detected, this probability is just $1/2 \exp[-A^2T]$, which is 3 dB worse than the quantum limit. Comparing this with (98), however, reveals an additional 3-dB loss due to heterodyning. So, heterodyne FSK detection is 6 dB inferior to the so-called "quantum limit" provided that phase noise can be neglected.

Returning now to the more realistic situation where phase noise is present and must be included in the performance analysis, we recall that the crucial assumption for the previous analysis was that the symbol rate baud $R = 1/T$ be much greater than the linewidth of the laser, so that the phase process $\Delta(t)$ could be regarded as a constant during the integration period. The inclusion of the phase noise process immediately raises fundamental problems concerning the detector structure.

As is well known,[23] the optimal processor in the presence of phase noise first estimates the phase process, and then uses this estimate to coherently demodulate or detect the IF signal. This is precisely what a PLL does, but we have seen that coherent PLL detection becomes

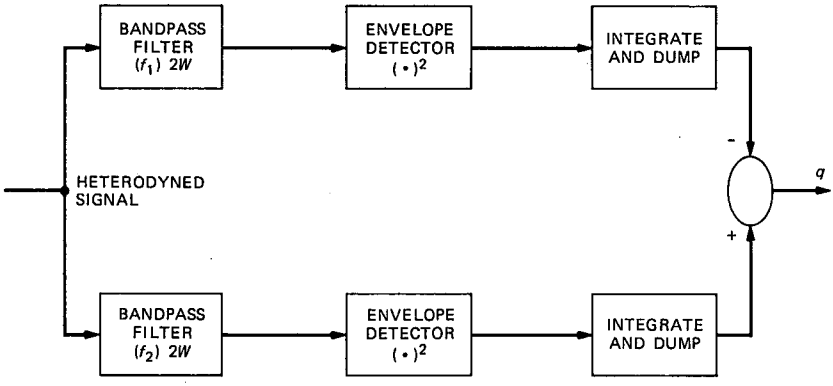


Fig. 10—FSK detector.

feasible only at very high data rates. This situation leads us to postulate a detector that, while not optimum, is reasonable and does not require a phase-locked loop.

The proposed frequency detector structure is shown in Fig. 10. It is essentially an energy detector. It consists of two ideal bandpass filters of total bandwidth equal to $2W$. The purpose of these filters is to limit the added white noise bandwidth as much as possible, while at the same time to retain most of the energy in the information carrying signal. A precise number for the bandwidth that satisfies these two seemingly contradictory requirements is hard to derive because, strictly, the received sine wave with phase noise has infinite bandwidth. The front-end bandwidth must remain as a parameter in our subsequent analysis, and engineering estimates will be attempted later.

Following these bandlimiting filters with square-envelope detectors and an integrator essentially provides an estimate of the energy in each frequency band, and this quantity should be independent of phase noise provided that the front-end bandwidth is sufficiently large. We now proceed to analyze the performance of this structure.

In the representation of signal plus noise, eq. (95), assume that ω_1 was sent. Regarding ω_1 as the center frequency, we represent the signal and noise in terms of in-phase and quadrature components as

$$\begin{aligned} S(t) &= 2A \cos \omega_1 t \cos \Delta(t) - 2A \sin \omega_1 t \sin \Delta(t) \\ &\quad + n_1(t) \cos \omega_1 t + n_2(t) \sin \omega_1 t \\ &= x(t) \cos \omega_1 t + y(t) \sin \omega_1 t, \end{aligned} \tag{99}$$

where

$$x(t) = 2A \cos \Delta(t) + n_1(t),$$

and

$$y(t) = 2A \sin \Delta(t) + n_2(t).$$

At the output of the bandpass filters $x(t)$ and $y(t)$ are bandlimited versions of the input. As we have already stated the signals $x(t)$ and $y(t)$ remain essentially undistorted at the output, and the only effect of the bandpass filters is to limit the noise band. Clearly, as the band W increases, this approximation becomes better.

The output baseband noises with unit input noise intensity now have mean-square values

$$En^2 = En_1^2 = En_2^2 = 4W. \tag{100}$$

Squaring the envelope and integrating as indicated in Fig. 8 yields the quadratic decision statistic q_1,

$$q_1 = \int_0^T [x^2(t) + y^2(t)]dt, \tag{101}$$

where for a given $\Delta(t)$, $x(t)$ and $y(t)$ are independent Gaussian processes with

$$Ex(t) = 2A \cos \Delta(t),$$

and

$$Ey(t) = 2A \sin \Delta(t). \tag{102}$$

The covariance functions of the ideally bandlimited baseband noise processes $n_1(t)$ and $n_2(t)$ are

$$En_1(t)n_1(t + \tau) = En_2(t)n_2(t + \tau) = 4W \frac{\sin 2\pi W\tau}{2\pi W\tau}. \tag{103}$$

As is well known,[35] associated with this covariance kernel are an infinite set of orthonormal eigenfunctions $\{\psi_k(t)\}$ and a set of nonnegative eigenvalues $\{\lambda_k\}$. Using these eigenfunctions, we represent the processes $x(t)$ and $y(t)$ as

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \sum_k \psi_k(t) \begin{pmatrix} x_k \\ y_k \end{pmatrix}, \tag{104}$$

where

$$\begin{pmatrix} x_k \\ y_k \end{pmatrix} = \int_0^T \psi_k(t) \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} dt,$$

and

$$Ex_k y_{k'} = \lambda_k \delta_{kk'}.$$

Using these expansions, the quadratic form q_1 in (101) can be put into the form

$$q_1 = \int_0^T [x^2(t) + y^2(t)] = \sum_k (x_k^2 + y_k^2). \tag{105}$$

In the bottom leg of Fig. 10 we assume that only the "noise" goes through (by previous hypothesis), and so the resulting quadratic form q_2 is now comprised only of noise. Of course, this is a mild assumption since ω_1 can be separated from ω_2 as much as one wishes to provide minimum leakage. Thus,

$$\begin{aligned} q_2 &= \int_0^T (\nu_1^2(t) + \nu_2^2(t))dt \\ &= \sum_k (\nu_{1k}^2 + \nu_{2k}^2), \end{aligned} \tag{106}$$

where $\nu_1(t)$ and $\nu_2(t)$ are quadrature and in-phase bandlimited noises in the lower leg, and are independent of the noises in the upper leg since the spectra occupy nonoverlapping frequency bands. For this reason we denote these noises by $\nu_1(t)$ and $\nu_2(t)$ to distinguish them from n_1 and n_2 in the upper leg. The decision statistic is the difference of the quadratic forms (105) and (106) and consequently the probability of error is

$$Pe = \Pr[q = q_1 - q_2 \leq 0]. \tag{107}$$

This can be expressed in terms of the characteristic function of q,

$$C(\omega) = Ee^{i\omega q}, \tag{108}$$

as the integral[34]

$$Pe = -\frac{1}{2\pi i} \int_{-\infty}^{\infty} \frac{C(\omega)}{\omega + i\epsilon}\, d\omega. \tag{109}$$

The $i\epsilon$, $\epsilon > 0$ in the denominator denotes the fact that in the complex ω plane the contour of integration goes above the singularity at $\omega = 0$.

Using (105) and (106) it is straightforward to calculate (108)

$$\begin{aligned} C(\omega) &= Ee^{i\omega\sum_k (x_k^2 + y_k^2)} Ee^{-i\omega\sum_k (\nu_{1k}^2 + \nu_{2k}^2)} \\ &= E_{\Delta(t)} \left\{ \frac{1}{\prod_k (1 - 2i\omega\lambda_k)} \frac{1}{\prod_k (1 + 2i\omega\lambda_k)} e^{i\omega\sum_k \left(\frac{\bar{x}_k^2 + \bar{y}_k^2}{1 - 2i\omega\lambda_k}\right)} \right\}, \end{aligned} \tag{110}$$

where

$$\begin{pmatrix} \bar{x}_k \\ \bar{y}_k \end{pmatrix} = \int_0^T \psi_k(t) \begin{pmatrix} 2A\cos\Delta(t) \\ 2A\sin\Delta(t) \end{pmatrix} dt,$$

and $E_{\Delta(t)}(\cdot)$ denotes the expectation with respect to the phase process $\Delta(t)$. To proceed further, we invoke an excellent approximation[35] regarding the behavior of the eigenvalues $\{\lambda_n\}$ in this application. Since these are the eigenvalues of the Prolate-Spheroidal wave functions, it is shown in Ref. 35 that

$$\lambda_k \sim \begin{cases} 2, & k \leq n = 2WT \\ 0, & k > n = 2WT, \end{cases} \tag{111}$$

and when this approximation is applied in (110), we get for the characteristic function

$$C(\omega) = \frac{\exp\left(\dfrac{i\omega 4A^2T}{1 - 4i\omega}\right)}{(1 - 4i\omega)^n(1 + 4i\omega)^n}, \tag{112}$$

where $n = 2WT$.

Substituting (112) into (109) we obtain

$$Pe = \frac{-1}{2\pi i}\int_{-\infty}^{\infty} \frac{d\omega}{\omega + i\epsilon} \frac{e^{\frac{Pi\omega}{1-i\omega}}}{(1 - i\omega)^n(1 + i\omega)^n}, \tag{113}$$

where $P = A^2T$ and we set $4\omega \rightarrow \omega$. By letting $i\omega = z$ we write (113) as a contour integral

$$Pe = -\frac{1}{2\pi i}\int_{-i\infty}^{i\infty} \frac{dz}{z} \exp\left(\frac{Pz}{1 - z}\right) \frac{1}{(1 - z)^n(1 + z)^n}, \tag{114}$$

where the indentation is now to the left around the origin.

We note the nth order pole at $z = -1$ and the essential singularity at $z = 1$. When the time-bandwidth product $n = 1$, the contour can be closed in the left-hand plane, and the value of the integral is just the residue of

$$f_1(z) = \frac{e^{P\frac{z}{1-z}}}{(1 - z)(1 + z)}\left(-\frac{1}{z}\right) \tag{115}$$

at $z = -1$. This gives for the probability of bit error

$$Pe(n = 1) = \frac{e^{-\frac{P}{2}}}{2}. \tag{116}$$

This result is identical to (98) and a moments reflection will reveal the reason for the consistency. Note that roughly

$$n = 2TW \sim \frac{1}{R}(R + kB_L) = 1 + \frac{B_L k}{R}, \tag{117}$$

where $R + kB_L = 2W$—the bandpass bandwidth of the incoming signal. This band must equal to or be greater than the signaling rate R, plus kB_L—the bandwidth required to pass the sine-wave signal with the phase noise undistorted (k is a positive integer), and again B_L is the laser linewidth. Clearly, when $B_L/R \ll 1$, no postintegration is required and consequently $n = 1$. Therefore in this case we get the previous result where we regarded the phase noise as a constant—precisely the case when $kB_L/R \ll 1$. When this condition is not necessarily satisfied and so $n \neq 1$, we can still evaluate the contour integral. In this general case, closing the contour in the left-hand plane enclosing the nth order pole gives for (113) the residue and hence the probability of bit error as a function of P and n,

$$Pe(P, n) = -\frac{1}{(n-1)!}\left[\frac{d^{n-1}}{dz^{n-1}}\left(\frac{\exp\left(\frac{Pz}{1-z}\right)}{z(1-z)}\right)\right]_{z=-1}. \tag{118}$$

We show in Appendix B that (118) can be expressed more explicitly as

$$Pe(P, n) = P_n(P/2)\exp\left(-\frac{P}{2}\right), \tag{119}$$

where $P_n(P/2)$ is an nth order polynomial in $(P/2)$ with properties

$$P_1(P/2) = 1/2$$

and

$$\lim_{n\to\infty} P_n(P/2) = \frac{1}{2}\exp\left(\frac{P}{2}\right).$$

As can be seen from (119), the degradation due to excess noise and postintegration manifests itself only in an algebraic coefficient in the bit-error-rate expression and not in the exponent.

From the foregoing analysis we present the curves of Fig. 11 that are plots of Pe versus P in decibels for different values of γ, which is twice the ratio of symbol rate to the sum of laser linewidths. The front-end bandwidths around frequencies f_1 and f_2 in Fig. 9 were selected to be

$$2W = R + 10(B_{L1} + B_{L2}), \tag{120}$$

where B_{L1} and B_{L2} are the laser linewidths, $k = 10$ in (117). The factor

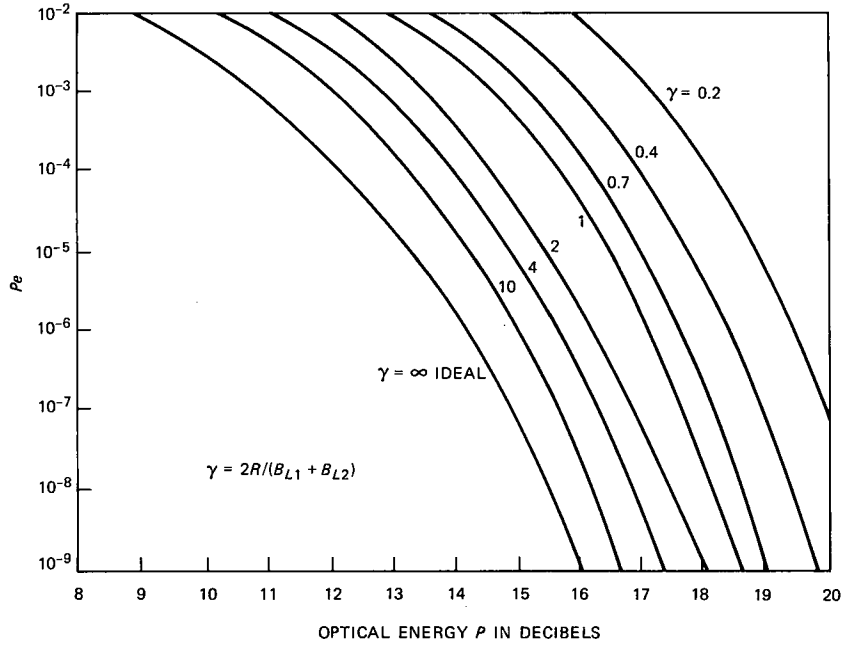


Fig. 11—FSK performance.

of 10 times the laser linewidths is judged to be adequate to pass the incoming FM signal without appreciable distortion.

When $\gamma \to \infty$, $Pe \to \exp\{-P/2\}$, which is the ideal binary FM performance. As already mentioned, this ideal performance is still 6 dB worse than the quantum limit, since 3 dB is lost from heterodyning and 3 additional decibels is lost due to the fact that the heterodyned FM signals are orthogonal rather than antipodal.

Fig. 11 can be used to determine the minimum data rate for efficient performance. Suppose the lasers have identical linewidths of 10 MHz, and one wishes the degradation not to exceed 1 dB. What is the minimum admissible data rate R? From Fig. 11, less than 1-dB degradation yields $\gamma \sim 10$, implying $R \geq 100$ Mb/s.

What is the degradation if one desires to transmit at 20 Mb/s? For the same laser linewidths as before, eq. (98) yields $\gamma = 2$, and from the figure we see that this value of γ yields a degradation of 2 dB.

As is well known, FSK can accommodate more than two frequencies to convey digital information without appreciably altering the form of the error rate expression provided again that bandwidth expansion is not an obstacle. For example, consider using 2^m frequences $m \geq 2$. Generalizing the structure of Fig. 10 to 2^m legs yields a probability of

error that is 2^m times the binary error rate but requires only $P/2m$ photons per bit to deliver the same amount of information as in the binary case. To cite an example, using four frequencies, $m = 2$, the information rate is $2/T$ and so T can be doubled to obtain the same bits per second as when $m = 2$. This comes at a moderate increase of bandwidth and a factor-of-4 increase in the error rate.

Let us consider the previous example where the binary rate $R = 100$ Mb/s. This can be achieved with ~40 photons per bit resulting in an error rate $\sim 10^{-9}$. Suppose one had only 20 photons per bit to expand, how can this data rate be accommodated without increasing the error rate? Suppose we half the signaling rate so that the new T equals twice the old T. To maintain the same rate in bits per second we must use four frequencies rather than two. The new signaling rate has now been halved and so we must recalculate γ from eq. (98) corresponding to $R = 50$. We find it to be five. From Fig. 11 we estimate that this value of γ results in a ~1-dB loss. The upshot is that FSK with four frequencies signaling at a data rate of 100 Mb/s can be achieved at a 4-dB increase of optical power over the quantum limit.

Next we analyze differential phase shift keying.

VI. DIFFERENTIAL PHASE SHIFT KEYING

In the presence of additive white Gaussian noise, Differential Phase Shift Keying (DPSK) is known to be very efficient in terms of the s/n required to achieve an acceptable error rate.[36-38] It is only a fraction of a decibel less efficient than coherent phase shift keying—the most efficient known method. Our objective here is to investigate the performance of this modulation method in optical communications and to assess the incurred penalty due to phase noise.

We begin our treatment by first considering detection at optical frequencies and then analyze the heterodyned version. Processing at optical frequencies may be practically inhibited because of the present lack of efficient (noise free) amplifiers, and therefore we also analyze the heterodyned version and compare performance of these two different approaches.

6.1 Optical processing

In DPSK, information is conveyed by the phase differences in two consecutive signaling intervals. Thus when the optical signal in a particular signaling interval is

$$S_0(t) = a_0 A \cos(\omega_0 t + \theta(t)), \qquad 0 \le t \le T, \tag{121}$$

and in the previous interval it was

$$S_{-1}(t) = a_{-1} A \cos(\omega_0 t + \theta(t)), \qquad -T \le t < 0, \tag{122}$$

where a_0, a_{-1} are ± 1 and $\theta(t)$ is again the phase noise process, information then is conveyed by the product $a_0 a_{-1}$.

Ideally one would measure the time average* of the phase difference in (121) and (122) during the interval $[0, T]$ and decide that $a_{-1}a_0 = 1$ if it lies between $-\pi/2$ and $\pi/2$ and $a_{-1}a_0 = -1$ if it lies outside of this phase range. Consequently, in a practically implemented system a lower bound on the bit error rate would be

$$Pe \geq \Pr\left[\frac{1}{T}\int_0^T [\theta(t) - \theta(t-T)]dt \geq \pi/2\right] \sim \exp\left\{-\left(\frac{\pi}{2}\right)^2 \frac{1}{2\sigma^2}\right\}, \tag{123}$$

where

$$\sigma^2 = E\left(\frac{1}{T}\int_0^T [\theta(t) - \theta(t-T)]dt\right)^2 = 2/3(2\pi)^2 N_0 T = \frac{4\pi}{3}\frac{B_L}{R},$$

and B_L is again the full 3-dB linewidth of the laser.

From this lower bound on bit error rate the ratio B_L/R is determined, setting a floor for the minimum admissible rate in terms of B_L. For example, if one desires an error rate $\sim 10^{-9}$, one equates

$$\left(\frac{\pi}{2}\right)^2 \frac{1}{2\sigma^2}$$

to 20 to obtain

$$R \sim 67 B_L.$$

As an example, for a $B_L = 20$ MHz, R would have to be greater than 1.34 Gb/s to achieve $Pe \sim 10^{-9}$.

The problem, however, is that in the optical processing repertory the phase differential cannot yet be obtained directly nor can two optical waves be multiplied directly. Therefore, one must consider a detection method that provides information about the phase difference in an indirect manner. Thus consider the following approach: first, the incoming signal is delayed by T seconds, then half of the power of the delayed signal plus and minus half of the power of the undelayed versions are passed through separate photodetectors. A schematic representation of this phase detector is shown in Fig. 12.

---

* I am indebted to Leonid G. Kazovsky for pointing out the time average of the phase noise differential provides a better lower bound than just the instantaneous value that appeared in my original manuscript.

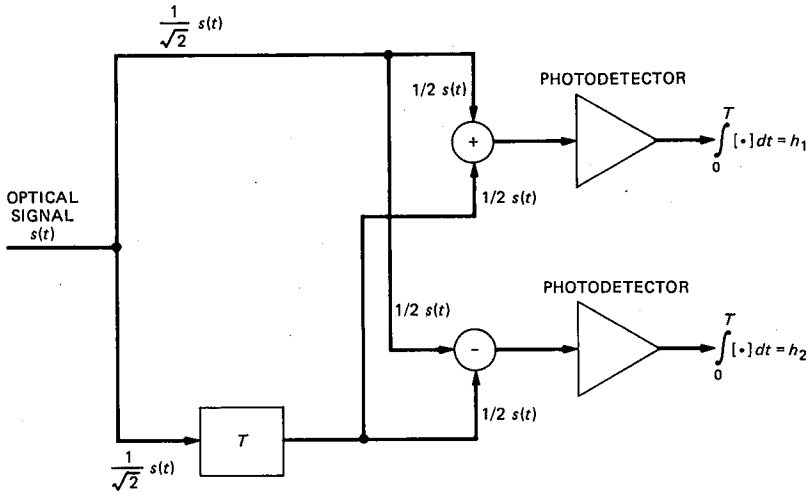


Fig. 12—Optical DPSK detector.

Referring to this figure, we write the sum and difference as

$$V(t) = \frac{a_0 A}{2}\cos(\omega_0 t + \theta(t)) \pm \frac{a_{-1}A}{2}\cos(\omega_0 t + \theta(t-T))$$

$$= \frac{a_0 A}{2}\cos(\omega_0 t + \Delta\theta(t) + \theta(t-T))$$

$$\pm \frac{a_{-1}A}{2}\cos(\omega_0 t + \theta(t-T)$$

$$= \frac{a_0 A}{2}\cos(\Delta\theta(t))\cos(\omega_0 t + \theta(t-T))$$

$$- \frac{a_0 A}{2}\sin(\Delta\theta(t))\sin(\omega_o t + \theta(t-T))$$

$$\pm \frac{a_{-1}A}{2}\cos(\omega_o t + \theta(t-T)), \tag{124}$$

where we set

$$\Delta\theta(t) = \theta(t) - \theta(t-T).$$

The output pulse count from the "sum" photodetector is a doubly stochastic Poisson process n_1 with conditional intensity λ_+ equal to the squared envelope of (124),

$$\lambda_+ = \frac{A^2}{2}(1 + a_0 a_{-1}\cos\Delta\theta(t)), \tag{125}$$

while the "difference" photodetector output is also a Poisson process n_2 independent of n_1 but having intensity

$$\lambda_- = \frac{A^2}{2}(1 - a_0 a_{-1}\cos\,\Delta\theta(t)). \tag{126}$$

Integrating the outputs of the two photodetectors for T seconds yields two average random counts

$$\Lambda_+ = \frac{A^2}{2}\int_0^T (1 + a_0 a_{-1}\cos\,\Delta\theta(t))dt$$

and

$$\Lambda_- = \frac{A^2}{2}\int_0^T (1 - a_0 a_{-1}\cos\,\Delta\theta(t))dt. \tag{127}$$

The detection statistic is the difference between the two counts,

$$n = n_1 - n_2. \tag{128}$$

When $n > 0$, $a_0 a_{-1}$ is taken to be 1, while if it is less than zero, $a_0 a_{-1}$ is taken to be -1.

Because of phase noise, the exact evaluation of the bit error rate is not mathematically tractable, but an exponential upper bound can be obtained from the moment generating function of the differential count, n. The moment generating function of n, conditioned on θ and $a_0 a_{-1}$, is readily calculated from the Poisson distribution

$$M_n(s \mid \theta,\, a_0 a_{-1}) = e^{\Lambda_+(e^s-1)+\Lambda_-(e^{-s}-1)}, \tag{129}$$

and will be used to upper bound the probability of error.

We begin by writing the probability of error

$$Pe = \frac{1}{2}\Pr(n \le 0 \mid a_0 a_{-1} = 1) + \frac{1}{2}\Pr(n \ge 0 \mid a_0 a_{-1} = -1), \tag{130}$$

and because of symmetry,

$$Pe = \Pr[n \le 0, \mid a_o a_{-1} = 1] \le E_{\Delta\theta(t)} e^{\bar{\nu}(e^{-s}-1)} + \bar{u}(e^s - 1)$$
$$\le E_{\Delta\theta} e^{\nu(e^{-s}-1)+u(e^s-1)}, \quad s \ge 0, \tag{131}$$

where

$$\bar{\nu} = \Lambda_+(A_o A_{-1} = 1)$$
$$\bar{u} = \Lambda_-(A_o A_{-1} = 1), \tag{132}$$

and

$$\nu = \frac{P}{2}(1 + \cos \Delta\theta)$$

$$u = \frac{P}{2}(1 - \cos \Delta\theta). \tag{133}$$

In the last two inequalities we used (129) in a Chernoff bound and made use of the convexity of the exponential function as well as the fact that $\Delta\theta(t)$ is stationary.

The tightest upper bound is obtained by selecting an optimum s for a given ν and u. This value of s can be obtained by setting the derivative of the exponent to zero. The set of random s's optimizing the exponent is then found to be

$$s_o = \tfrac{1}{2} \ln \frac{\nu}{u}, \tag{134}$$

and since s_o has to be positive, we require that $\nu > u$, which from (133) implies that $\cos \Delta\theta(t)$ must be positive. For values of $\Delta\theta(t)$ such that $\cos \Delta\theta \leq 0$, the optimum value of s_o is seen to be zero. Thus, in order that the bound (131) be reasonably tight, the average with respect to $\Delta\theta(t)$ indicated in (131) must be carried out over two sets of $\Delta\theta$

$$\Delta\theta \in R_1, \cos \Delta\theta \leq 0$$

and

$$\Delta\theta \in R_2, \cos \Delta\theta > 0. \tag{135}$$

This yields for (131)

$$Pe \leq E_{\Delta\theta \in R_1} e^{\nu(e^{-s}-1)+u(e^{s}-1)} + E_{\Delta\theta \in R_2} e^{\nu(e^{-s}-1)+u(e^{s}-1)}. \tag{136}$$

The first term above can be upper bounded by setting $s = 0$ and further upperbounding the probability that $\cos \Delta\theta \leq 0$, yields for this term $\Pr[\Delta\theta \geq \pi/2]$. So after some calculations and substitutions, (136) becomes

$$Pe \leq \exp\left\{-\left(\frac{\pi}{2}\right)^2 \frac{1}{2\sigma_{\Delta\theta}^2}\right\} + e^{-P} E_{\Delta\theta \in R_2} e^{P|\sin \Delta\theta|}, \tag{137}$$

where

$$\sigma_{\Delta\theta}^2 = 2\pi \frac{B_L}{R} \tag{138}$$

and again $P = A^2T$.

The penalty incurred due to phase noise depends on the behavior of

the expectation in (137). It does not appear feasible to evaluate this expectation exactly and therefore we must resort to asymptotic analysis valid for large P. We defer this analysis to after the discussion of heterodyne DPSK, since the penalty evaluation there involves a similar calculation.

Before embarking an analysis of heterodyne DPSK, however, we remark that even when $\Delta\theta = 0$, the probability of error in optically processed DPSK is 3 dB inferior to the quantum limit. This is seen from (17) since when $\Delta\theta = 0$,

$$Pe = \frac{1}{2} e^{-A^2T} = \frac{1}{2} e^{-P}. \tag{139}$$

The reason for this loss is inherent in the demodulation process. As can be seen, the optical detector turned the optical signal into an "on-off" signal and the explanation for the inefficiency is that twice as much average optical power has been transmitted to detect an on-off signal.

It should also be pointed out that the chief motivation for using differential phase modulation in the microwave region is that it yields an error probability close to coherent demodulation without having to transmit and recover carrier phase.[39] We therefore turn our attention to heterodyne DPSK detection next.

6.2 Heterodyne DPSK

As has already been pointed out, heterodyning an optical signal results in additive white noise and therefore it is imperative as in the FSK case, to bandlimit the heterodyned signal plus noise to a bandwidth just sufficient to pass the received signal with the impressed phase noise undisturbed. In DPSK, a bandlimited signal is multiplied by a delayed version and the product is post integrated in order to eliminate residual noise. This is a standard comparison detector[39] and is depicted in Fig. 13.

As has been done before, the heterodyned signal is represented in the interval $0 \leq t \leq T$ as,

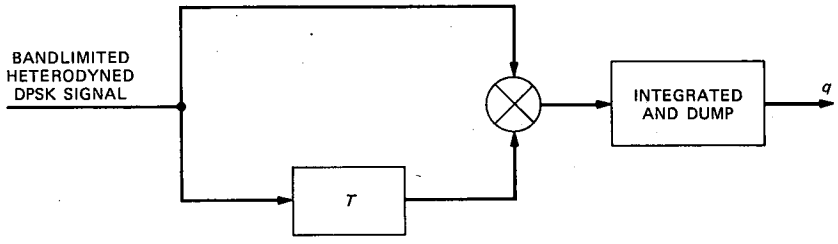


Fig. 13—Heterodyne differential phase shift keying.

$$V(t) = a_0 2A \cos(\omega_i t + \delta(t)) + n(t), \tag{140}$$

$$V_d(t) = a_{-1} 2A \cos(\omega_i t + \delta_d(t)) + n_d(t), \tag{141}$$

where the subscript indicates a delay of T seconds, ω_i is the IF frequency, and $\delta(t) = \theta(t) - \Phi(t)$, ($\theta(t)$ is the transmitter laser phase noise and $\Phi(t)$ is the local laser's phase noise.) Again, $n(t)$ is a white gaussian noise process with unit double sided spectral density.

Expressing (140) and (141) in terms of in-phase and quadrature components, yields

$$\begin{aligned} V(t) &= 2a_0 A \cos \delta(t) \cos \omega_i t + n_1(t) \cos \omega_i t + n_2(t) \sin \omega_i t \\ &\quad - 2a_0 A \sin \delta(t) \sin \omega_i t \\ &= [2a_0 A \cos \delta(t) + n_1(t)] \cos \omega_i t \\ &\quad - [2a_0 A \sin \delta(t) + n_2(t)] \sin \omega_i t, \end{aligned} \tag{142}$$

$$\begin{aligned} V_d(t) &= 2a_{-1} A \cos \delta_d(t) \cos \omega_i t + n_{1d}(t) \cos \omega_i t \\ &\quad + n_{2d} \sin \omega_i t - 2a_{-1} A \cos \delta_d(t) \sin \omega_i t \\ &= [2a_{-1} A \cos \delta_d(t) + n_{1d}(t)] \cos \omega_i t \\ &\quad + [2a_{-1} A \sin \delta_d(t) + n_{2d}(t)] \sin \omega_i t \end{aligned} \tag{143}$$

The baseband noise processes in (142) and (143), $n_1(t)$, $n_{1d}(t)$, $n_2(t)$ and $n_{2d}(t)$ are mutually independent and bandlimited to W hertz. Since the total noise power at the output of the band-pass filter is $4W$, the identical variances of the baseband noises must also equal to $4W$. Now, multiplying $V(t)$ by $V_d(t)$, eliminating double frequency components and integrating, results in the decision statistic q_0,

$$\begin{aligned} q_o &= \int_0^T [2a_o A \cos \delta(t) + n_1(t)][2a_{-1} A \cos \delta_d(t) + n_{1d}(t)] dt \\ &\quad + \int_0^T [2a_o A \sin \delta(t) + n_2(t)] \\ &\quad \cdot [2a_{-1} A \sin \delta_d(t) + n_{2d}(t)] dt \\ &= a_o a_{-1} q, \end{aligned} \tag{144}$$

where

$$\begin{aligned} q &= \int_0^T [2A \cos \delta(t) + n_1(t)][2A \cos \delta_d(t) + n_{1d}(t)] dt \\ &\quad + \int_0^T [2A \sin \delta(t) + n_2(t)][2A \sin \delta_d(t) + n_{2d}(t)] dt. \end{aligned} \tag{145}$$

A detection error is made whenever $a_0a_{-1} = 1$ and $q \leq 0$ or when $a_0a_{-1} = -1$ and $q \leq 0$. So, the bit error rate then is just

$$Pe = \Pr[q \leq 0]. \tag{146}$$

To facilitate the calculation of (146) let,

$$\begin{aligned} x(t) &= 2A \cos \delta(t) + n_1(t) \\ x_d(t) &= 2A \cos \delta_d(t) + n_{1d}(t) \\ y(t) &= 2A \sin \delta(t) + n_2(t) \\ y_d(t) &= 2A \sin \delta_d(t) + n_{2d}(t), \end{aligned}$$

and write

$$\begin{aligned} q = \int_0^T (xx_d + yy_d)dt = \int_0^T \left[\left(\frac{x + x_d}{2}\right)^2 + \left(\frac{y + y_d}{2}\right)^2\right] dt \\ - \int_0^T \left[\left(\frac{x - x_d}{2}\right)^2 + \left(\frac{y - y_d}{2}\right)^2\right] dt. \end{aligned} \tag{147}$$

Further define

$$\begin{aligned} u_1 &= \frac{x + x_d}{2}, \quad u_2 = \frac{y + y_d}{2}, \\ v_1 &= \frac{x - x_d}{2}, \quad v_2 = \frac{y - y_d}{2}, \end{aligned} \tag{148}$$

and conditioned on $\delta(t)$, calculate

$$\begin{aligned} Eu_1 &= A(\cos \delta + \cos \delta_d) \\ Eu_2 &= A(\sin \delta + \sin \delta_d) \\ Ev_1 &= A(\cos \delta - \cos \delta_d) \\ Ev_2 &= A(\sin \delta - \sin \delta_d). \end{aligned} \tag{149}$$

Rewriting (147) in terms of (148) we then obtain

$$\begin{aligned} q &= \int_0^T (u_1^2 + u_2^2) - \int_0^T (v_1^2 + v_2^2), \\ &= \sum_k (u_{1k}^2 + u_{2k}^2 - v_{1k}^2 - v_{2k}^2), \end{aligned} \tag{150}$$

where the u_k and v_k are again the coefficients in the expansions of $v(t)$ and $u(t)$ in the eigenfunctions $\{\psi_n(t)\}$.

Structurally (150) is identical to the quadratic forms obtained in the FSK case and therefore, we can express the bit error rate as

$$Pe = \Pr[q \le 0] = \frac{-1}{2\pi i} \int_{-\infty}^{\infty} \frac{Ee^{i\omega q}}{\omega + i\epsilon}\, d\omega. \tag{151}$$

Now however the characteristic function of the quadratic form (150) is

$$Ee^{i\omega q} = E_{\delta,\delta_d}\left[\frac{\exp\left\{ i\omega \sum_k \frac{(\bar{u}_{1k}^2 + \bar{u}_{2k}^2)}{1 - 2i\omega\lambda_k} - i\omega \sum_k \frac{(\bar{v}_{1k}^2 + \bar{v}_{2k}^2)}{1 + 2i\omega\lambda_k} \right\}}{\prod_k (1 - 2i\omega\lambda_k)(1 + 2i\omega\lambda_k)} \right]. \tag{152}$$

The eigenvalues as before are approximately 1 for $k \le n = 2WT$ and zero beyond and consequently (152) is to a good approximation,

$$\begin{aligned} &E\{e^{i\omega q} \mid \delta, \delta_d\} \\ &\sim \frac{\exp\left\{ \frac{i\omega}{1 - 2i\omega} \int_0^T [(Eu_1)^2 + (Eu_2)^2]dt \right.}{} \\ &\qquad \frac{\left. - \frac{i\omega}{1 + 2i\omega} \int_0^T ((Ev_1)^2 + (Ev_2)^2)dt \right\}}{(1 - 2i\omega)^n (1 + 2i\omega)^n} \\ &= \frac{\exp\left\{ \frac{i\omega 2A^2}{1 - 2i\omega} \int_0^T (1 + \cos(\delta - \delta_d))dt \right.}{} \\ &\qquad \frac{\left. \cdot \frac{-i\omega 2A^2}{1 + 2i\omega} \int_0^T (1 - \cos(\delta - \delta_d))dt \right\}}{(1 - 2i\omega)^n (1 + 2i\omega)^n}. \end{aligned} \tag{153}$$

Substituting this formula into (143), we obtain

$$\begin{aligned} Pe = -\frac{1}{2\pi i} \int_{-i\infty}^{i\infty} \frac{dz}{z}\, E_\phi \\ \cdot \left[\frac{\exp\left\{ \frac{zA^2}{1 - z} \int_0^T (1 + \cos\phi)dt \right.}{} \right. \\ \left. \frac{\left. - \frac{z}{1 + z} A^2 \int_0^T (1 - \cos\phi)dt \right\}}{(1 - z)^n (1 + z)^n} \right], \end{aligned} \tag{154}$$

where $\phi = \delta(t) - \delta_d(t)$.

Unfortunately this integral cannot be expressed more explicitly as in FSK because of the essential singularity at $z = -1$. When $\phi = 0$ (no phase noise), the essential singularity disappears and (154) is identical to the previously encountered integral associated with the error rate in FSK. So in this case we obtain exactly

$$Pe = P_{n-1}(P)e^{-P}, \tag{155}$$

where P_{n-1} is again the $(n-1)$th order polynomial defined in Appendix B and $P = A^2T$.

In the important case when $1/T = R \gg B_L$, we obtain from (154), setting $n = 1$,

$$Pe = \frac{e^{-A^2T}}{2} = \frac{e^{-P}}{2}. \tag{156}$$

This result, again as in the optical case, can be seen to be 3 dB inferior to the "quantum limit" that is solely attributed to the 3-dB loss in the heterodyning process.

We now return to the central problem of assessing the penalty incurred by DPSK due to the presence of phase noise. An exact evaluation of the penalty is not mathematically tractable in general, and, as in the previous case, we must resort to upper bounds using the moment generating function of the quadratic form (153) or (154).

From (154) the moment generating function of q is

$$\begin{aligned} M_q(z) = Ee^{zq} &= E_\phi \exp \frac{\left(P\bar{\nu}\dfrac{z}{1-z} - P\bar{u}\dfrac{z}{1+z}\right)}{(1-z)^n(1+z)^n}, \\ &\le E \exp\left\{P\nu\frac{1}{1-z} - Pu\frac{z}{1+z}\right\}, \qquad z \ge 0, \end{aligned} \tag{157}$$

where

$$\bar{\nu} = \frac{1}{T}\int_0^T \nu, \qquad \bar{u} = \frac{1}{T}\int_0^T u$$

and

$$\nu = 1 + \cos\phi, \qquad u = 1 - \cos\phi. \tag{158}$$

The inequality in (157) is valid because of convexity and the stationarity of $\phi(t)$. While the form of this moment generating function is different from (129), similar techniques can be used to bound the probability of error. We proceed as follows:

$$Pe = \Pr[q \le 0] \le \frac{E_\phi e^{-Pf(z)}}{(1-z)^n(1+z)^n}, \tag{159}$$

where

$$f(z) = z\left(\frac{\nu}{1+z} - \frac{u}{1-z}\right), \qquad z \ge 0.$$

Splitting the range of ϕ into two parts, R_1 such that $\phi \epsilon R_1$, $\cos \phi > 0$ and R_2 such that $\phi \epsilon R_2$, $\cos \phi \leq 0$, we write (159) as

$$Pe \leq E_{\phi \epsilon R_1} \frac{e^{-Pf(z)}}{(1-z)^n(1+z)^n} + E_{\phi \epsilon R_2} \frac{e^{-Pf(z)}}{(1-z)^n(1+z)^n} \quad (160)$$

By setting the derivative of $f(z)$ to zero reveals that there exists an optimizing $z > 0$ namely,

$$z_0 = \frac{\sqrt{\nu} - \sqrt{u}}{\sqrt{\nu} + \sqrt{u}}, \qquad \cos \phi > 0$$

$$= 0, \qquad \cos \phi \leq 0 \quad (161)$$

When these values of z_0 are substituted into (160) we get

$$Pe \leq \exp\left(-\left(\frac{\pi}{2}\right)^2 \left(\frac{1}{2\sigma_\phi^2}\right) + E_{\phi \epsilon R_2}\left(\frac{(1+|\sin \phi|)^n}{2^n |\sin \phi|^n} e^{-P(1-|\sin\phi|)}\right), \quad (162)$$

where

$$\sigma_\phi^2 = 2\pi \frac{(B_{L1} + B_{L2}}{R}, \quad (163)$$

and where B_{L1} and B_{L2} are again the linewidths of the two lasers.

From (162) and (137) we see that the degradation from ideal performance, $\exp(-P)$, is essentially determined by the average value over the range $\phi \epsilon R_2$ of

$$F(\phi) = Ge^{P|\sin\phi|}, \quad (164)$$

where

$$G(\phi) = \left(\frac{1 + |\sin \phi|}{2 \sin \phi}\right)^2,$$

or

$$= 1,$$

depending on whether it is used in eq. (162) or (137), respectively.

Thus we write

$$EF(\phi) = \frac{1}{\sqrt{2\pi}\sigma_\phi} \int_{R_2} G(\phi) e^{P\epsilon(\phi)} d\phi, \quad (165)$$

where

$$\epsilon(\phi) = \sin \phi - \frac{\phi^2}{2C}, \quad (166)$$

and

$$C = \sigma_\phi^2 P. \tag{167}$$

We now evaluate (167) asymptotically valid for $P \to \infty$ and fixed C. Setting the derivative of (166) to zero, we solve the transcendental equation for the saddle point, $0 \le \phi_0 \le \pi/2$.

$$\epsilon'(\phi_0) = \cos \phi_0 - \frac{\phi_0}{C} = 0, \tag{168}$$

and observe that

$$\epsilon''(\phi_0) = -\left(\sin \phi + \frac{1}{C}\right) < 0. \tag{169}$$

With these results we can express (165) as

$$EF(\phi) \sim \frac{1}{\sqrt{2\pi}\sigma_\phi} G(\phi_0) e^{P\epsilon(\phi_0)} \int_{-\infty}^{\infty} e^{\frac{P\epsilon''(\phi_0)}{2}(\phi-\phi_0)^2} d\phi$$

$$\sim \frac{G(\phi_0)}{\sigma_\phi \sqrt{|\epsilon''(\phi_0)|P}} e^{P\epsilon(\phi_0)}. \tag{170}$$

Using (170) we can summarize the error rate estimates developed in the foregoing analysis.

1. Optical DPSK. Substituting (170) into (137) with the definition of $G(\phi)$ in (167), the probability of error is

$$Pe \le \exp\left\{-\left(\frac{\pi}{2}\right)^2 \frac{1}{2\sigma_{\Delta\theta}^2}\right\} + [C \sin \phi_0 + 1]^{-1/2} \cdot \exp\{-P(1 + \phi_0^2/2C - \sin \phi_0)\}, \tag{171}$$

where

$$\sigma_{\Delta\theta}^2 = 2\pi \frac{B_L}{R}.$$

2. Heterodyne DPSK. Now substituting (170) into (162) yields,

$$Pe \le \exp\left\{-\left(\frac{\pi}{2}\right)^2 \frac{1}{2\sigma_\phi^2}\right\} + [C \sin \phi_0 + 1]^{-1/2} \left[\frac{1 + \sin \phi_0}{\sin \phi_0}\right]^n \cdot \exp\left\{-P\left(1 + \frac{\phi_0^2}{2C} - \sin \phi_0\right)\right\}, \tag{172}$$

where

$$\sigma_\phi^2 = 2\pi \frac{B_{L1} + B_{L2}}{R}.$$

Ignoring unimportant coefficients, the probability of error is seen to be dominated by the maximum of the two terms in either (171) or (172). These are seen to be identical expressions barring the coefficients. It is observed from (171) and (172) that the exponents in the second term approach the exponents of the first term as $P \to \infty$. This can be verified from (168) since when $P \to \infty$, $C \to \infty$ for fixed σ_ϕ^2 and so the solution, $\phi_0 \to \pi/2$. As a consequence of this limit, our estimate of the error rate versus P has a threshold, or floor, at

$$\exp\left\{-\left(\frac{\pi}{2}\right)^2\left(\frac{1}{2\sigma_\phi^2}\right)\right\}.$$

As an example, at $Pe \sim 10^{-9}$,

$$\left(\frac{\pi}{2}\right)^2\left(\frac{1}{2\sigma_\phi^2}\right) \sim 20$$

and according to our prediction the floor for optical DPSK is $R/B_L \sim 100$ while in heterodyne DPSK it is $R/B_L \sim 200$, for identical laser linewidths. These floor predictions are slightly pessimistic. The lower bounds in (123) predict a floor of $R/B_L = 67$ for optical DPSK and $R/B_L = 134$ for heterotype DPSK. The discrepancy has to do with our bounding techniques.

It is now possible to define the exponential degradation or penalty, from ideal performance (above the respective floors) in either case by

$$\text{Penalty} = -10 \log\{1 - \sin \phi_0 + \phi_0^2/2C\}. \tag{173}$$

This is seen to be a function of the single parameter C defined in (167) and ϕ_0 the solution to (168).

It is important to observe that (171) and/or both (162) and (137) exhibit an exponential degradation due to phase noise unlike in the case of FSK. In both systems widening the front-end bandwidths of the respective detectors ensures minimal distortion suffered by the received heterodyned signals. However, in DPSK the static phase noise differential manifests itself in an exponential degradation, while in FSK no such degradation occurs.

In Fig. 14 we exhibit the penalty function, (173), as a function of R/B_L for both optical and heterodyne DPSK. We note that the exponential degradation is infinite below these respective floors. The arrows shown on the figures at $R/B_L = 100$ and 200 are aimed to emphasize that according to our estimates, the degradation is infinite at rates less than these values. In other words, no amount of additional optical energy can drive the error rate below these respective floors. Similar curves can be drawn for different Pes and hence different Ps. A striking feature of these curves is that the 3-dB degradation is

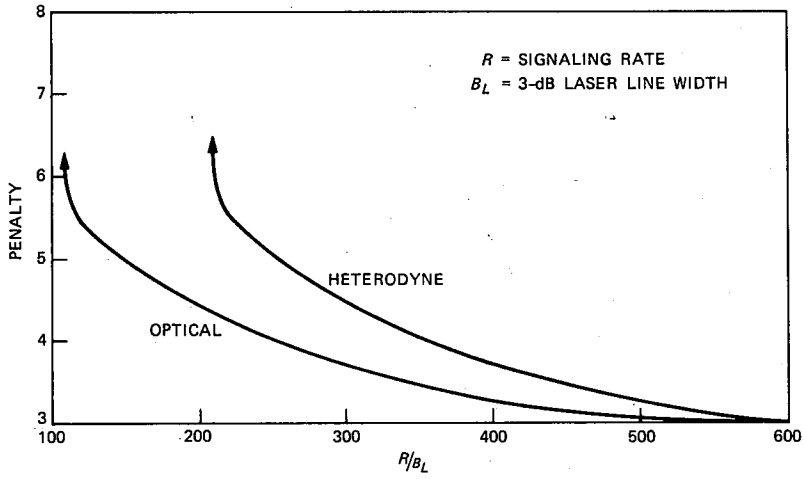


Fig. 14—Penalty in decibels due to phase noise in optical and heterodyne DPSK, $Pe \sim 10^{-9}$, $D = 20$.

approached very rapidly at rates greater than 400 times the laser linewidths (less than 4 dB is given up in either system). Evidently DPSK is very sensitive to phase noise for $R < 300B_L$.

VII. ON-OFF KEYING

We saw that on-off keying of an optical wave using direct detection achieves the quantum limit. Here we wish to analyze the performance of this modulation method when the on-off optical signal is first heterodyned to an IF frequency and then direct detected. Since the microwave version of on-off modulation has the same signal distance properties as FSK, we expect similar performance. There are however some differences and for the sake of completeness we include the following analysis.

Letting the IF frequency be ω_i, the differential phase noise be $\delta(t)$ and the resultant Gaussian noise with unit spectral density be $n(t)$, the observed IF microwave signal is then represented as

$$a2A \cos(\omega_0 t + \delta(t)) + n(t), \qquad 0 \le t \le T, \tag{174}$$

where here $a = 1$ or 0.

A reasonable way to process (121) is to first bandlimit it, then envelope detect the bandlimited version and finally postintegrate the square-envelope to obtain the decision statistic. This processor is depicted in Fig. 15 and the output statistic is

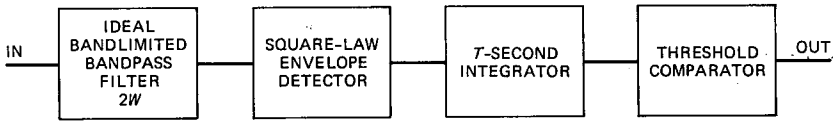


Fig. 15—Envelope detector.

$$q_1 = \int_0^T [2A \cos \delta(t) + n_1(t)]^2 dt + \int_0^T [2A \sin \delta(t) + n_2(t)]^2 dt, \qquad a = 1, \quad (175)$$

and

$$q_0 = \int_0^T [n_1^2(t) + n_2^2(t) dt, \qquad a = 0, \quad (176)$$

where n_1 and n_2 are baseband gaussian noise processes with double-sided spectral densities equal to 2 and are bandlimited to W hertz. This must be so since the bandpass noise process

$$n(t) = n_1(t)\cos \omega_0 t + n_2(t)\sin \omega_0 t,$$

is bandlimited to $2W$ and consequently,

$$En^2 = En_1^2 = En_2^2 = 4W.$$

In writing (122) we assumed as before that W is sufficiently wide to pass the in-phase and quadrature signals in the presence of phase noise undistorted.

Now let

$$x(t) = 2A \cos \delta(t) + n_1(t)$$

and

$$y(t) = 2A \sin \delta(t) + n_2(t), \quad (177)$$

and make an expansion of x and y in terms of the prolate-spheroidal orthonormal set of functions

$$\{\psi_n(t)\}, \qquad 0 \leq t \leq T, \qquad n = 0, 1, 2, \cdots. \quad (178)$$

Then we can write

$$q_i = \sum_n [x_n^2 + y_n^2]$$

and

$$q_0 = \sum_n [n_{1n}^2 + n_{2n}^2]. \quad (179)$$

We note that $\{x_n\}$, $\{y_n\}$, $\{n_{1n}\}$ and $\{n_{2n}\}$ are mutually independent conditional Gaussian random variables (conditioned on $\delta(t)$) with the following parameters:

$$Ex_n = \int_0^T 2A \cos \delta(t)\psi_n(t)dt, \text{ all } n$$

$$Ey_n = \int_0^T 2A \sin \delta(t)\psi_n(t)dt, \text{ all } n,$$

and

$$En_{1n} = En_{2n} = 0, \text{ all } n.$$

Also

$$\text{Var } x_n = \text{Var } y_n = \text{Var } n_{1n} = \text{Var } n_{2n} = \lambda_n,$$

where $\{\lambda_n\}$ are the eigenvalues associated with the eigenfunctions $\psi_n(t)$.

Our method of estimating the error rate will be based on a Chernof bounding technique requiring the moment generating functions of q_1 and q_0. These functions are readily evaluated for any θ as follows,

$$Mq_1(\theta) = Ee^{\theta q_1} = \prod_\ell Ee^{\theta x_\ell^2} Ee^{\theta y_\ell^2}$$

$$= \frac{\exp\left\{\theta \sum_\ell \frac{(Ex_\ell)^2 + (Ey_\ell)^2}{1 - 2\lambda_\ell\theta}\right\}}{\prod_\ell [1 - 2\lambda_\ell\theta]}, \tag{180}$$

and

$$Mq_0(\theta) = Ee^{\theta q_0} = \prod_\ell Ee^{\theta n_{1\ell}^2} Ee^{\theta n_{2\ell}^2}$$

$$= \prod_\ell (1 - 2\lambda_\ell\theta)^{-1}. \tag{181}$$

To proceed further with the analysis, we invoke again the excellent approximation regarding the behavior of the eigenvalues λ_ℓ. Here it can be verified that

$$\lambda_\ell \sim 2, \qquad \ell \leq n = 2WT$$

$$= 0, \qquad \ell > n = 2WT. \tag{182}$$

and when this approximation is used in (180) and (181) we obtain for the moment generating functions

$$Mq_1(\theta) = \frac{\exp\left\{P\frac{\theta}{1 - \theta)}\right\}}{(1 - \theta)^n}, \tag{183}$$

and

$$Mq_0(\theta) = (1 - \theta)^n, \tag{184}$$

where $P = A^2T$, the optical energy.

The bit error rate is now upper bounded as follows:

$$Pe = \tfrac{1}{2}\Pr[q_1 \leq \tau] + \tfrac{1}{2}\Pr[q_0 > \tau], \tag{185}$$

where τ is a threshold that will be optimized later. The probabilities in (185) cannot be evaluated exactly; however, tight exponential upper bounds can be obtained as follows.

$$\Pr[q_1 \leq \tau] \leq e^{\theta\tau} M_{q1}(-\theta), \qquad 0 \leq \theta < 1 \tag{186}$$

and

$$\Pr[q_0 > \tau] \leq e^{-\theta\tau} M_{q1}(\theta), \qquad 0 \leq \theta < 1. \tag{187}$$

Using (183) in (186) it can be verified that there exists a $\theta = \theta_0$, which makes the bound tightest. By differentiation we find

$$\theta_0 = \sqrt{\frac{P}{\tau}} - 1, \qquad P > \tau, \tag{188}$$

and when this value of θ_0 is substituted into (186) we get

$$\Pr[q_1 \leq \tau] \leq \frac{\left(\frac{\tau}{P}\right)^n}{2} e^{-P\left(1-\sqrt{\frac{\tau}{P}}\right)^2} \tag{189}$$

Following the same procedure for tightening the bound in (187) we find an optimum θ_0 given by,

$$\theta_0 = 1 - \frac{n}{\tau}, \qquad \tau > n, \tag{190}$$

and when this value of θ is substituted into (187) we obtain,

$$\Pr[q_0 \geq \tau] \leq \left(\frac{\tau}{n}\right)^n e^{-n\left(\frac{\tau}{n}-1\right)}. \tag{191}$$

Equating the exponent in (189) to the one in (191) we determine the best threshold τ given by

$$\tau_0 = \frac{1}{4P}(P+n)^2$$

Substituting this value of τ_o in (189) and (191) we get for Pe in (185)

$$Pe \leq \left(\frac{1}{2}\right)^{2n+1}\left[\left(\frac{P}{n}\right)^n\left(1+\frac{n}{P}\right)^{2n} + \left(1+\frac{n}{P}\right)^n\right] e^{-\frac{P}{4}\left(1-\frac{n}{P}\right)^2} \tag{192}$$

We see from this upper bound that when $n < P$,

$$Pe \sim e^{\frac{A^2T}{4}} = e^{-\frac{P}{2}} \tag{193}$$

which is identical to the FSK performance since here the average optical energy is half that of FSK. The exponential degradation from ideal is seen to be

$$D_0 = -20 \log \left(1 + \frac{n}{P}\right),$$

where $n = 2WT$ is the total band required to pass the modulated signal with phase noise undistorted. A conservative example might be as follows. $P = 80$ will yield an ideal error rate of $\sim 10^{-9}$, $n = 1/R(1 + 10B_L)$ ($10B_L \sim 400$ MHz should be sufficient to pass the signal with phase noise undistorted.) For these parameters a simple calculation reveals that when $R = 0.5$ GB/s the penalty is about 0.2 dB. At rates below this the degradation starts to be substantial. Note however, that this performance is still 3 dB poorer than DPSK and 6 dB poorer than the quantum limit.

VIII. ACKNOWLEDGMENTS

I wish to thank many of my colleagues at AT&T Bell Laboratories for the many fruitful technical interactions during the course of this research. The contributions of Paul S. Henry have been invaluable. He acquainted me with this subject matter, provided continuous guidance, and made significant and substantial contributions to the evolution of the main results. I would have been honored to have him as a coauthor but he modestly refused.

I also had valuable discussions with the following: A. S. Acampora, N. Amitay, B. Glance, D. J. Goodman, L. J. Greenstein, G. J. Foschini, T. Li, L. A. Linke, B. F. Logan, R. W. Lucky, J. E. Mazo, A. A. M. Saleh, G. Vannucci, and A. D. Wyner. Thanks are also due to Vincent W. S. Chan from MIT Lincoln Laboratory for stimulating discussions.

REFERENCES

1. M. Ross, *Laser Receivers*, New York: Wiley and Sons, 1966.
2. W. K. Pratt, *Laser Communication Systems*, New York: Wiley and Sons, 1969.
3. Y. Yamamoto and T. Kimura, "Coherent Optical Fiber Transmission Systems," IEEE J. Quant. Electron., *QE-17* (June 1981), pp. 919–35.
4. T. Okoshi et al., "Computation of Bit-Error Rate of Various Heterodyne and Coherent-Type Optical Communication Schemes," J. Opt. Commun., *2* (September 1981), pp. 89–96.
5. T. Okoshi, "Heterodyne and Coherent Optical Fiber Communications: Recent Progress," IEEE Trans. Micr. Th. and Tech., *MTT-30* (August 1982), pp. 1138–48.
6. F. Faure and D. LeGuen, "Effect of Semiconductor Laser Phase Noise on BER Performance in an Optical DPSK Heterodyne-Type Experiment," Electron. Lett., *18* (October 28, 1982), pp. 964–5.
7. 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. Quant. Electron., *QE-19* (February 1983), pp. 180–93.
8. A. D. Wyner, unpublished work.

9. J. E. Mazo and J. Salz, "On Optical Data Communication via Direct Detection of Light Pulses," B.S.T.J., *55*, No. 3 (March 1976), pp. 347–70.
10. J. C. Campbell et al., "High Performance Avalanche Photodiode with Separate Absorption, Grading and Multiplication Regions," Electron. Lett., *19* (September 29, 1983), pp. 818–9.
11. R. S. Kennedy, private communication.
12. C. H. Henry, "Theory of the Linewidth of Semiconductor Lasers," IEEE J. Quant. Electron., *QE-18* (February 1982), pp. 259–64.
13. M. W. Fleming and A. Mooradian, "Fundamental Line Broadening of Single-Model GaAlAs Diode Lasers," Appl. Phys. Lett., *38* (April 1, 1981), pp. 511–3.
14. C. Harder, K. Vahala, and A. Yariv, "Measurement of the Linewidth Enhancement Factor α of Semiconductor Lasers," Appl. Phys. Lett., *42* (February 15, 1983), pp. 328–30.
15. F. G. Walther and J. E. Kaufmann, "Characterization of GaAlAs Laser Diode Frequency Noise," Sixth Top. Mtg. Opt. Fib. Commun., New Orleans, February–March 1983, Paper TUJ5.
16. J. E. Kaufmann, "Phase and Frequency Tracking Considerations for Heterodyne Optical Communications," Proc. Int. Telemetering Conf., San Diego, Calif., September 1982.
17. S. O. Rice, "Mathematical Analysis of Random Noise, W. Nelson, ed., *Selected Papers on Noise and Stochastic Processes:* Dover Publications Inc., 1954, pp. 145–62.
18. M. Shikada et al., "100 Mb/s ASK Heterodyne Detection Experiment Using 1.3 μm DFB Laser Diodes," Conf. Opt. Fiber Commun., New Orleans, January 1984, Paper TUK6.
19. K. Kikuchi, T. Okoshi, and R. Arata, "Measurements of Linewidth and FM-Noise Spectrum of 1.52 μm InGaAsP Lasers," Electron. Lett., *20* (June 21, 1984), pp. 535–6.
20. T. P. Lee et al., "Measured Spectral Linewidth of Single-Frequency 1.3 and 1.5 μm Injection Lasers," Electon. Lett., *20* (November 22, 1984), pp. 1011–2.
21. R. Wyatt and W. J. Devlin, "10 kHz Linewidth 1.5 μm InGaAsP External Cavity Laser with 55 nm Tuning Range," Electron Lett., *19* (February 3, 1983), pp. 110–2.
22. A. L. Scholtz et al., "Infra-red Homodyne Receiver with Acousto-optically Controlled Local Oscillator," El. Lett., *19* (March 17, 1983), pp. 234–5.
23. D. D. Falconer and J. Salz, "Optimal Reception of Digital Data Over the Gaussian Channel with Unknown Delay and Phase Jitter," IEEE Trans. Inf. Th., *IT-23*, No. 1 (January 1977), pp. 117–26.
24. E. Wong, *Stochastic Processes in Information and Dynamical Systems*, New York: McGraw-Hill, 1971.
25. A. J. Viterbi, *Principles of Coherent Communication*, New York: McGraw-Hill, 1966.
26. R. M. Gagliardi and S. Karp, *Optical Communications*, New York: Wiley and Sons 1976, Chapter 6.
27. R. Wyatt, T. G. Hodgkinson, and D. W. Smith, "1.52 μm PSK Heterodyne Experiment Featuring an External Cavity Diode Laser Local Oscillator," Electron. Lett., *19* (July 7, 1983), pp. 550–2.
28. V. K. Prabhu, "PSK Performance with Imperfect Carrier Phase Recovery," IEEE Trans. Aerospace and Elec. Syst., *AES-12*, No. 2 (March 1976), pp. 275–86.
29. K. Kikuchi et al., "Bit-Error Rate of PSK Heterodyne Optical Communication System and its Degradation Due to Spectral Spread of Transmitter and Local Oscillator," Electron. Lett., *19*, No. 11 (May 26, 1983), pp. 417–8.
30. V. M. S. Chan, L. L. Jeromin, and J. E. Kaufmann, "Heterodyne Lasercom Systems using GaAs Lasers for ISL Applications," IEEE Int. Conf. Commun., Boston, June 1983, Paper E1.5.
31. L. L. Jeromin and V. W. S. Chan, "Performance Estimates for a Coherent Optical Communication System," Conf. Opt. Fiber Commun., New Orleans, January 1984, Paper TUK2.
32. R. Wyatt et al., "140 Mb/s Optical FSK Fibre Heterodyne Experiment at 1.54 μm," Electron. Lett., *20* (October 25, 1984), pp. 912–3.
33. K. Emura et al., "Novel Optical FSK Heterodyne Single Filter Detection System Using a Directly Modulated DFB Laser Diode," Electron. Lett., *20* (November 22, 1984), pp. 1022–3.
34. J. E. Mazo and J. Salz, "Probability of Error for Quadratic Detectors," B.S.T.J., *44* (November 1965), pp. 2165–86.

35. D. Slepian and H. Pollak, "Prolate Sphyroidal Wave Functions—Fourier Analysis and Uncertainty—I," B.S.T.J., *40* (January 1961), pp. 43–63.
36. W. R. Bennett and J. R. Davey, *Data Transmission*, New York: McGraw-Hill, 1965, Chapter 11.
37. M. Shikada, M. Emura, and K. Minemura, "High-Sensitivity Optical PSK Heterodyne Differential Detection Simulation Experiment," Fourth Int. Conf. Int. Opt. and Opt. Fiber Commun., Tokyo, June 1983, Paper 30C3-4.
38. G. Nicholson, "Probability of Error for Optical Heterodyne DPSK System with Quantum Phase Noise," Electron. Lett., *20* (November 22, 1984), pp. 1005–7.
39. R. W. Lucky, J. Salz, and E. J. Weldon, Jr., *Principles of Data Communication*, New York: McGraw-Hill, 1968.

APPENDIX A

An Estimate of Probability of Error in Optical Homodyne Detection

Here we derive an upper bound for the bit error rate which will be used in evaluating performance of optical homodyne as well as heterodyne reception. We need to estimate the following probability (eq. 192),

$$Pe = \Pr[-\rho\xi + \bar{\nu}_0 \geq 0], \tag{194}$$

where ξ and $\bar{\nu}_0$ are independent random variables with joint probability density $p(\xi)p(\bar{\nu}_0)$. By definition, (194) is written,

$$Pe = \int_0^\infty p(x)dx, \tag{195}$$

where $p(x)$ is the density of $\bar{\nu}_0 - \rho\xi$. For any $\lambda \geq 0$, (195) is upper bounded by

$$Pe \leq \int_{-\infty}^{\infty} e^{\lambda x}p(x)dx = Ee^{\lambda x} = Ee^{-\rho\xi\lambda}Ee^{\lambda\bar{\nu}_0} = e^{\frac{\lambda^2}{2}}Ee^{-\rho\xi\lambda} \tag{196}$$

To make progress with (196), we note that since e^{-y} is convex

$$\begin{aligned} Ee^{-\rho\xi\lambda} &= Ee^{-\rho\lambda\frac{1}{T}\int_0^T \cos\psi dt} \\ &\leq \frac{1}{T}\int_0^T Ee^{-\rho\lambda\cos\psi}dt = Ee^{-\lambda\rho\cos\psi}, \end{aligned} \tag{197}$$

since $\psi(t)$ is a stationary process. Using this result, (196) is upper bounded by

$$Pe \leq e^{\frac{\lambda^2}{2}}Ee^{-\lambda\rho\cos\psi} \tag{198}$$

Substituting the stationary probability density of ψ (36) into (198) we get,

$$Pe \leq e^{\frac{\lambda^2}{2}} \frac{1}{2\pi I_0(\alpha)} \int_{-\pi}^{\pi} e^{-\lambda\rho\cos\psi + \alpha\cos\psi} d\psi$$

$$\leq \frac{1}{I_0(\alpha)} e^{\frac{\lambda^2}{2} + |\alpha - \lambda\rho|}. \tag{199}$$

Lower bounding the Bessel function by

$$I_0(\alpha) \geq \frac{e^{\alpha}}{\sqrt{2\pi\alpha}} \left[1 - \frac{1}{\pi} \sqrt{\frac{2}{\alpha}} e^{-\pi^2\alpha/2}\right],$$

we further upper bound (199)

$$Pe \leq \frac{e^{\frac{\lambda^2}{2} - \alpha + |\alpha - \lambda\rho|}}{\left(1 - \frac{1}{\pi} \sqrt{\frac{2}{\alpha}} e^{-\frac{\pi^2\alpha}{2}}\right) (\sqrt{2\pi\alpha})^{-1}}. \tag{200}$$

The exponent in (200) is now optimized with respect to $\lambda > 0$. Let

$$E(\lambda) = \frac{\lambda^2}{2} - \alpha + |\alpha - \lambda\rho|, \tag{201}$$

and note that $E(0) = 0$. Moreover,

$$\frac{d}{d\lambda} E(\lambda)|_{\lambda=0} = (\lambda - \rho \operatorname{sgn}(\alpha - \lambda\rho))_{\lambda=0}$$

$$= -\rho. \tag{202}$$

We are thus assured that a $\lambda = \lambda_0$ exists such that $E(\lambda_0) \leq 0$. To find this optimum value of λ, we examine,

$$\frac{dE(\lambda)}{d\lambda} = \lambda - \rho \operatorname{sgn}(\alpha - \lambda\rho) = 0, \tag{203}$$

and conclude that when $\alpha/\rho \geq \rho$, $\lambda_0 = \rho$. On the other hand when $\alpha/\rho \leq \rho$, the optimum value of λ_0 is on the boundary $\lambda_0 = \alpha/\rho$ and so the optimum exponent in (201) is

$$E(\lambda_0) = \begin{cases} \frac{\rho^2}{2}, & \alpha > \rho^2 \\ \alpha \left[1 - \frac{\alpha}{2\rho^2}\right], & \alpha \leq \rho^2 \end{cases}. \tag{204}$$

Substituting (204) into (200) we finally get

$$Pe \leq \begin{cases} g(\alpha)e^{-\frac{\rho^2}{2}}, & \frac{\alpha}{\rho^2} > 1 \\ g(\alpha)e^{-\alpha\left(1-\frac{\alpha}{2\rho^2}\right)}, & \frac{\alpha}{\rho^2} < 1 \end{cases}, \tag{205}$$

where

$$g(\alpha) = \left(1 - \frac{1}{\pi}\sqrt{\frac{2}{\alpha}}\, e^{-\frac{\pi^2\alpha}{2}}\right)^{-1}\sqrt{2\pi\alpha}.$$

APPENDIX B

Evaluation of Residues

I am indebted to B. F. Logan, Jr. of Department 11219 for supplying material included in this Appendix.

We require the residue of $f_n(z)/(-z)$ at $z = -1$, where

$$f_n(z) = \frac{\exp\left(\frac{Pz}{1-z}\right)}{(1-z)^{n+1}(1+z)^{n+1}}, \qquad P > 0. \tag{206}$$

Setting $P/2 = \lambda$, $z = t - 1$, the probability of error $Pe(P, n)$ in Section V (eq. 118) is the coefficient of t^n in the Taylor series expansion of the function

$$\frac{\exp\left(P\frac{(t-1)}{2-t}\right)}{(1-t)(2-t)^{n+1}} = e^{-\frac{P}{2}}\frac{\exp\left(\frac{Pt}{2(2-t)}\right)}{(1-t)(2-t)^{n+1}}. \tag{207}$$

Now consider the function

$$F_n(x, \lambda) = \frac{1}{2^{n+1}}\frac{\exp\left(\frac{\lambda x}{1-x}\right)}{(1-2x)(1-x)^{n+1}}$$

$$= \frac{1}{2^{n+1}}\sum_{k=0}^{\infty} p_k(\lambda, n)x^k. \tag{208}$$

Then,

$$Pe(n, \lambda) = \frac{e^{-\lambda}}{2^{2n+1}}\, p_n(\lambda, n).$$

The generator function for the generalized Laguere polynomials, defined as

$$L_m^{(n)}(-\lambda) = \sum_{k=0}^{m}\binom{m+n}{n-k}\frac{\lambda^k}{k!}$$

$$= \sum_{k=0}^{m}\binom{m+n}{k}\frac{\lambda^{m-k}}{(m-k)!} \tag{209}$$

is just the function

$$Gn(x, \lambda) = \frac{\exp\left(\frac{\lambda x}{1-x}\right)}{(1-x)^{n+1}}$$

$$= \sum_{k=0}^{\infty} L_k^{(n)}(-\lambda) x^n. \tag{210}$$

Multiplying Gn by $(1-2x)^{-1}$, we have

$$p_n(\lambda, n) = \sum_{k=0}^{m} 2^{m-k} L_k^{(n)}(-\lambda), \tag{211}$$

and collecting coefficients of $\frac{\lambda^k}{k!}$ we find

$$p_m(\lambda, n) = \sum_{k=0}^{n} a_k(m, n) \frac{\lambda^k}{k!}, \tag{212}$$

where

$$a_k(m, n) = \sum_{j=0}^{m-k} \binom{n+k+j}{j} 2^{m-k-j}$$

$$= \sum_{j=0}^{m-k} \binom{n+m-d}{n+k} 2^d.$$

Now setting $a_k(n, n) = p_k(n)$, we have

$$Pe(\lambda, n) = \frac{e^{-\lambda}}{2^{2n+1}} \sum_{k=0}^{n} p_k(n) \frac{\lambda^k}{k!}, \tag{213}$$

where

$$p_k(n) = \sum_{j=0}^{n} \binom{2n-j}{n+k} 2^j. \tag{214}$$

Since (214) is a finite polynomial, (213) can readily be evaluated by computer.

AUTHOR

Jack Salz, B.S.E.E., 1955, M.S.E., 1956, and Ph.D. (Electrical Engineering), 1961, University of Florida; AT&T Bell Laboratories, 1961—. Mr. Salz first worked on the electronic switching system. Since 1968 he has supervised a group engaged in theoretical studies in data communications and is currently a member of the Network Systems Research Department. During the academic year 1967–68, he was on leave as Professor of Electrical Engineering at the University of Florida. He was a visiting lecturer at Stanford University in Spring 1981 and a visiting MacKay Lecturer at the University of California, at Berkeley, in Spring 1983.