Submitted to IEEE Trans. on Communication
November 1992

LIDS-P-2145

Reference Transmission Schemes for Phase Noise Immunity [1]

Murat Azizoğlu
Department of Electrical Engineering and Computer Science
George Washington University

Pierre A. Humblet
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

October 28, 1992

Abstract

We consider an optical communication system where a reference signal is transmitted along with a phase modulated information carrying signal. This system has the potential of alleviating the effect of phase noise considerably. We jointly optimize the bandwidth of the receiver and the powers of the two signals. The scheme achieves a performance that is close to that of ideal phase modulation systems when the ratio of bit rate to laser linewidth is large, and a performance that is identical to that of frequency modulation when the ratio is small. We also introduce a centered filter structure which achieves a far better performance than traditional filters due to increased the phase coherence.

[1] This research was supported by NSF under grants NCR-8802991 and NCR-9206379, by DARPA under grant F19628-90-C-002, and by ARO under contract DAAL03-92-G-0115.

1 Introduction

Phase noise is a major impairment on the performance of coherent optical communication systems. Due to spontaneously emitted photons within the laser cavity, the phase of a semiconductor laser exhibits random fluctuations resulting in phase noise [1].

The incomplete knowledge and time-varying nature of the phase causes a degradation in the bit error rate. Traditional methods to alleviate this performance degradation involve modulation formats that are relatively insensitive to phase uncertainty, e.g. Amplitude Shift Keying (ASK) and Frequency Shift Keying (FSK). Envelope detector structures with widened filter bandwidths are used in conjunction with these modulation formats to reduce the performance degradation to a few dB's [2, 3, 4]. While this receiver modification yields acceptable performance, it does not achieve the full potential of phase noise impaired systems. This is because the signaling mechanisms do not exploit the structure of the phase noise problem.

In this paper we consider an alternative communication scheme which has been designed specifically for its robustness against phase noise. This approach attempts to optimize the signaling mechanism as well as the receiver structure and therefore yields better results than doing the latter alone.

The phase noise problem may be viewed as the lack of a reference signal at the receiver that has the same phase structure as the received signal. Therefore the transmission of a reference signal that is corrupted with the same noisy phase sample as the information carrying signal may help improve the performance. We study one such transmitted reference scheme here and obtain its performance. In Section 2 we describe the phase noise process and its statistical properties. In Section 3 we introduce transmitted reference systems and explain their structure. The performance of such systems with wideband filters is given in Section 4. The reference transmission schemes with optimally adjusted power and bandwidth parameters are described in Section 5 and their performances are analyzed in Sections 6 and 7. The results are presented in Section 8, and then Section 9 concludes the paper.

2 Phase Noise Model

The unmodulated field output of a semiconductor laser is given by

$$s_T(t) = A_T \cos\left(2\pi\nu_0 t + \theta_T(t)\right)$$

where A_T is the amplitude, ν_0 is the optical frequency and $\theta_T(t)$ is the phase noise process. The phase noise is commonly modeled as a Brownian motion process [5] which can be written as

$$\theta_T(t) = 2\pi \int_0^t \mu(t)\, dt$$

where the frequency noise $\mu(t)$ is a white Gaussian process with spectral height $\beta_T/2\pi$. The parameter β_T is the 3 dB bandwidth of the power spectral density of the output field $s_T(t)$. It is called the laser *linewidth* as it provides a measure of the spectral broadening induced by phase noise.

In coherent optical systems, the received field is first processed by an optical heterodyne receiver, which adds a local oscillator signal and photodetects the sum. The local oscillator signal is at frequency ν_1 and is corrupted by its own phase noise process $\theta_{\mathrm{LO}}(t)$. When the local oscillator power is sufficiently high, the photodetection process can be modeled as a downshift of the carrier frequency from optical domain to electrical domain as well as addition of a white shot noise process with Gaussian statistics. Hence, in the absence of modulation, the intermediate frequency (IF) output of the photodetector is

$$r(t) = A\cos\left(2\pi f_c t + \theta(t)\right) + n(t)$$

where $f_c = |\nu_0 - \nu_1|$, $\theta(t) = \theta_T(t) - \theta_{\mathrm{LO}}(t)$ is the combined phase noise process, and $n(t)$ is the additive white Gaussian noise with spectral density $N_0/2$. Since the transmitter and the local oscillator lasers have statistically independent phase noise processes, the combined linewidth β of $\theta(t)$ is the sum of individual laser linewidths.

3 Transmitted Reference Systems

The discussion in the previous section reveals how the phase noise problem may be viewed as a reference problem. If the local oscillator signal had the same phase noise corruption as the transmitted signal, then the two phase noise processes would cancel to result in an IF signal free of phase noise. This is not to suggest, however, that the signal to be used as the local oscillator signal can be transmitted from the transmitter. Such a scheme would result in a weak local signal which would overshadow the advantage of perfect phase noise cancellation.

The reference transmission to be discussed here aims to alleviate the adverse effects of phase noise *after the photodetection.* The goal is to provide the receiver with two IF signals corrupted with the same phase noise process. One of these signals will be the modulated, information-carrying signal, while the other signal will be an unmodulated reference signal. These two signals can be subsequently processed in the IF domain to minimize the error probability.

The first issue that needs to be addressed is the mechanism in which two signals can be transmitted simultaneously. Since these signals will share a common channel provided by the optical fiber, a certain orthogonality must be provided so that the signals do not interfere with each other and can be extracted at the receiver for further signal processing. Two main methods have been suggested to achieve this orthogonality [6, 7]. The first method assigns two different optical carrier frequencies ν_1 and ν_2 to the information and reference signals respectively. This can be accomplished by shifting the frequency of a portion of the transmitter laser output. The signals will occupy nonoverlapping frequency bands if the difference between ν_1 and ν_2 is much larger than the data rate and the linewidth. The receiver can separate them by using coherent detection and appropriate IF filtering, and bring them to the same center frequency.

The second reference transmission method uses two orthogonal polarizations for the two signals. The transmitter laser produces a lightwave that contains both x and y polarization components which are separated by a polarization-sensitive beam splitter. One of the branches is modulated before beam combining and transmission. The receiver can separate the two signals by using another polarization-sensitive beam splitter.

Frequency and polarization based reference systems are identical when viewed at the IF domain in the receiver. Both provide the receiver with two IF signals corrupted with the same phase noise process.

In this paper, we assume that Phase Shift Keying (PSK) modulation format is employed. We then have the IF signals

$$
\begin{aligned}
r_1(t) &= A_1 \cos\left(2\pi f_c t + \theta(t) + \pi m(t)\right) + n_1(t) \\
r_2(t) &= A_2 \cos\left(2\pi f_c t + \theta(t)\right) + n_2(t)
\end{aligned}
$$

where $\theta(t)$ is the combined phase noise process, $n_1(t)$ and $n_2(t)$ are statistically independent white Gaussian processes each with spectral density $N_0/2$, and $m(t)$ is the binary data waveform. If the bit duration is T, the two signals have signal-to-noise ratios (SNR) $\xi_i = A_i^2 T/2N_0$ for $i = 1, 2$. The information carrying signal $r_1(t)$ and the reference signal $r_2(t)$ will be processed by an IF receiver that will reach a bit decision based on the correlation between the two signals. Essentially, the reference signal will be used as a local signal which is perfectly phase-locked to the received signal. Since $r_1(t)$ and $r_2(t)$ have a common phase noise process, the mixed signal $r_1(t)r_2(t)$ contains a baseband term which is free of phase noise, and a double frequency term with doubled phase noise. Therefore, in the absence of additive noise, we could eliminate the phase noise entirely by filtering out the high frequency component. While this observation is promising, it merely points out that there is no error floor[2] in this transmitted reference scheme. Since the SNR is always finite, the signals need to be filtered first to limit the additive noise power. Therefore, the IF receiver is of the form shown in Figure 1. Filters 1 and 2 are bandpass filters centered at f_c, while Filter 3 is a lowpass filter. We will refer to this receiver as a double filter receiver due to two stages of filtering. When the lowpass filter is absent, it will be called a single filter receiver.

The main contribution of this paper is to point out the benefit of jointly optimizing the filter bandwidths and the distribution of the available transmitter power between the signals, and to quantify the performance improvement. Filtering operations in the presence of phase noise must be analyzed carefully since the spectral broadening necessitates wider filters. Matched filters, which are optimal in the absence of phase noise, start deforming the desired signal with the introduction of phase noise. Wider filters introduce more additive noise at their output. The tradeoff between the phase noise and additive noise must be accurately analyzed for transmitted reference systems [6, 7, 8]. Reference transmission has been previously studied only in the context of wideband filters [6, 7], resulting in a suboptimal performance, as outlined in the next section.

4 Performance of Wideband Receiver Structures

4.1 Single Filter Receiver

A wideband single filter receiver is as in Figure 1 with the post-mixing filter absent and the pre-mixing filters having a large passband as explained below. The decision device samples the mixer output at the end of each bit period, and compares the sampled value to 0 to reach the decision.

Let the filter bandwidths for the reference and information signals be W and B respectively. When these bandwidths are large enough to pass the signals undistorted, the standard results of [9] about the probability that the product of two complex Gaussian random variables has a negative real part can be invoked to find the error probability as

$$P_e = \frac{1}{2}\left[1 - Q(a,b) + Q(b,a)\right] \tag{1}$$

where

$$a = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{\xi_1}{BT}} + \sqrt{\frac{\xi_2}{WT}}\right) \tag{2}$$

[2]Error floor is the residual error probability as SNR tends to ∞.

$$b = \frac{1}{\sqrt{2}}\left(\sqrt{\frac{\xi_1}{BT}} - \sqrt{\frac{\xi_2}{WT}}\right)$$

and $Q(\cdot,\cdot)$ is the Marcum Q function. The design goal is to optimize the power distribution subject to the total power constraint $\xi_1 + \xi_2 \le \xi$ for a given bandwidth pair (B, W).

When the two filters are identical, $B = W$ as in [6], the optimal power distribution is symmetric, $\xi_1 = \xi_2 = \xi/2$. To see this, one uses the identity [10]

$$P_e = \frac{1}{2}\left[1 - Q(a,b) + Q(b,a)\right] = \frac{1}{2}e^{-(a^2+b^2)/2}\left[I_0(ab) + 2\sum_{n=1}^{\infty}\left(\frac{b}{a}\right)^n I_n(ab)\right] \tag{3}$$

where $I_n(\cdot)$ is the modified Bessel function of order n. $I_0(x) \ge 1$ with equality only when $x = 0$. On the other hand, the series above is nonnegative, and equals 0 only when $b = 0$. Therefore we obtain the lower bound $P_e \ge 1/2e^{-\xi/2BT}$ through the use of $a^2 + b^2 = (\xi_1 + \xi_2)/BT$. The equality is achieved when $b = 0$, i.e. when $\xi_1 = \xi_2 = \xi/2$. Therefore the optimal power distribution achieves the error probability

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

Since BT is large by definition, this system will have a performance that is far inferior to phase noise free FSK with incoherent IF detection which has the error probability $P_e = \frac{1}{2}e^{-\xi/2}$.

The following fundamental observation is made in [7]. The reference signal $r_2(t)$ occupies a smaller bandwidth than the information signal $r_1(t)$, as it is unmodulated. Hence $r_2(t)$ can be filtered with a narrower passband reducing the additive noise power at the output. This fact breaks the symmetry between the two signals and allows an asymmetric power distribution. The exact calculation of the optimal power distribution appears to be infeasible in this case. If it is assumed that the SNR is high in both channels, (1) can be approximated by $P_e \simeq Q(a - |b|)$ where $Q(\cdot)$ is the complementary distribution function of a standard Gaussian random variable [11]. Then the asymptotically optimal power distribution should maximize the minimum of ξ_1/BT and ξ_2/WT subject to the constraint $\xi_1 + \xi_2 = \xi$. The solution is easily found as

$$\frac{\xi_1}{\xi} = \frac{B}{B+W}$$

which satisfies $b = 0$ in (2). Therefore P_e is obtained from (3) to be simply given as[3]

$$P_e = \frac{1}{2}e^{-\xi/(B+W)T}\ . \tag{5}$$

[3]This simple form is not given in [7].

A comparison of (5) and (4) shows that asymmetric power distribution has the potential of 3 dB performance improvement. The effective drop in the SNR is the sum of filter bandwidths normalized by the bit rate. By making the reference filter narrower, one saves in the SNR. This is particularly significant when the phase noise strength, $\gamma \triangleq 2\pi\beta T$, is small, which implies $W/B \ll 1$.

4.2 Double Filter Receiver

A wideband double filter receiver is also considered in [6]. The first stage of filtering is the same as the single filter receiver. The mixer output, which contains the baseband PSK modulation, double frequency components, signal cross noise and noise cross noise terms, is integrated over the bit duration prior to sampling. When the front end filters are identical with wide passband B, the two signal cross noise terms are Gaussian with flat spectral levels $A_i^2 N_0/4$ over $|f| < B$. The noise cross noise term, which is neglected in [6], has a triangular baseband spectrum over $|f| < B$ with peak $N_0^2 B/2$. As in [6], we assume that the total noise is Gaussian, which is only accurate in the $BT \gg 1$ regime. We then calculate the noise variance making use of the fact that the lowpass filter passband is much narrower than the noise cross noise spectrum. The error probability is then found as $P_e = Q\left(\sqrt{2\xi_1\xi_2/(\xi + BT)}\right)$. The optimal power distribution still satisfies $\xi_1 = \xi_2 = \xi/2$ with the resulting error probability

$$P_e \simeq Q\left(\sqrt{\frac{\xi/2}{1 + BT/\xi}}\right) . \tag{6}$$

If the SNR is very high, then $BT/\xi \ll 1$ and one obtains a performance that is 3 dB worse than phase noise free FSK. However, for practical values of phase noise strength and optimized receiver bandwidths, phase noise does not cause a penalty as large as 3 dB in FSK performance [2, 3]. This clearly shows the undesirability of wideband filters in the receiver.

These conclusions remain valid when $W < B$. B is replaced W in (6) and the accompanying discussion. Consequently, the optimal power distribution is the symmetric one even when the front-end filters are not identical. The advantage due to slightly narrower, but still wideband, reference filtering is not significant with the narrowband filtering after the correlation, as the same poor performance limit is encountered. We will see in the later sections that this will not be true when the front-end filters are not wideband.

4.3 Discussion

The analyses outlined above are useful in bringing out the possible performance improvement by employing nonidentical filters and unequal powers for the information and reference signals. However, the conclusion is imprecise as the effect of phase noise on the system performance is implicit in the selection of the filter bandwidths. These bandwidths B and W are taken to be sufficiently large to pass the phase noisy sinusoids unaltered. A heuristic relation introduced in [12] and used in [6, 7] takes $B = 1/T + k\beta$ and $W = k\beta$, where k is chosen to be large enough, such as 8, to make the filters wideband according to the criterion mentioned above. This relation is motivated by the fact that the unmodulated carrier has a 3 dB bandwidth of β, and the modulation increases the bandwidth by an amount on the order of the bit rate. The predicted performance will depend strongly on the assumed value for k. Furthermore, for a fixed value of k the linewidth can be made small enough to violate the wideband nature of the filters, thus making the performance estimate entirely unrealistic. The presence of a heuristic bandwidth parameter, the setting of which arbitrarily determines the performance, is an undesired feature of these analyses.

The wideband front-end filters considered above do not adequately balance the tradeoff between phase noise and additive noise. The spectral broadening effect of phase noise is overemphasized by forcing these filters to pass the signals undistorted, while the effect of additive noise is underemphasized. It is well understood today that the nondistortion afforded by wideband filters is not necessary for good performance in the presence of phase noise [2, 3, 4]. An interesting example occurs in the floor performance of DPSK modulation where a narrowband receiver outperforms a wideband receiver even when the additive noise is neglected [13].

Figures 2 and 3 demonstrate the poor performance of wideband reference transmission schemes. In Figure 2, the error probability of the single wideband filter receiver from (5) is shown together with that of single filter binary FSK with optimal bandwidth setting [3]. It is seen that although the power is optimally distributed in the wideband system, a much simpler FSK scheme yields a better performance. The parameter k is set to 8 as in [7]. The same observation holds true with double filter receivers as shown in Figure 3 where the value $k = 7$ is used. Again binary FSK with optimal filter bandwidth outperforms the wideband reference scheme by far.

An important observation is that the performance of a transmitted reference system with identical filters and an even power distribution is identical to that of binary orthogonal signaling, e.g. wide deviation FSK, with envelope detection. This is because an equivalent receiver will take the filter outputs $y_1(t)$ and $y_2(t)$, form $(y_1(t) + y_2(t))^2$ and $(y_1(t) - y_2(t))^2$, perform lowpass filtering (integration) on these signals in the case of a double filter receiver, and decide for the larger of the sampled values. The signal component of one of the waveforms formed will be zero when the powers are equal. This effectively results in an orthogonal modulation. Two immediate conclusions can be drawn from this fact:

1. The inferior performance of wideband reference transmission scheme is due to the lack of filter optimization,
2. Without power and filter optimization, reference transmission is limited to the performance of FSK[4].

It is with the joint optimization of filters and powers, as will be shown in the following sections, that the true potential of reference transmission can be uncovered.

5 Reference Transmission with Optimal Signal and Receiver Parameters

In this section, we describe the model for the filters in the receiver, and we provide a performance analysis of the system. Because of the correlation operation that follows the front-end filters, the lowpass equivalents of the filter outputs must be considered. The lowpass equivalent of a bandpass filter is modeled as a finite-time integrator. This filter model is used by Foschini et. al. [2, 14] and many others [3, 15, 16, 17, 18]. It makes a time-domain analysis with random signals tractable.

The integrators for the information signal $r_1(t)$ have a time duration of $T_1 = T/M$, while those for the reference signal $r_2(t)$ have time durations $T_2 = KT/M$, where $K, M \geq 1$ will be found so as to optimize the performance. Thus, the information filter has a *bandwidth expansion factor* of M relative to a matched filter, while the reference filter has a *bandwidth reduction factor* of K relative to the information filter. This reflects the previous observation that the reference signal can be filtered more tightly due to the absence of modulation. For analytical convenience we assume that M and K are both integers.

We will consider two forms of timing alignment between the filter outputs, as shown in Figure 4. In the first form, the output of an integrator at time t is its input integrated over $(t - T_i, t)$ where $i = 1$ or 2. If T_1 is viewed as a time unit, the first $K - 1$ units of the integration window of the reference filter precedes that of the information filter. For

[4]This also implies that the performance of wideband double filter receiver approximated by (6) can be more accurately estimated using the analysis of [2].

large K the distance in time between portions of the phase noise process that affect the filter outputs will be large. Therefore the filter outputs may lose the phase coherence of the inputs. To increase this coherence, one has to minimize the maximum distance between the respective integration windows. This means that the windows must be centered: If the information filter integrates over $(t - T_1, t)$, then the reference filter must integrate over $(t - (T_1 + T_2)/2, t + (T_2 - T_1)/2)$, as in Figure 4.c. This constitutes the second timing form.

The two filter forms above will be referred to as noncentered and centered filters respectively. At the output of the first stage of filtering we have the information signal $x(t)$ and the reference signal $y(t)$ which are subsequently mixed and lowpass filtered. This mixing is equivalent to finding the inner product of two signal vectors. The mixed signal is passed through a lowpass filter which we model as a discrete-time adder after [2]. The output of this filter is the sum of M samples of its input where the samples are taken with period T_1.

It is convenient to employ a complex envelope notation where $x(t) = \mathrm{Re}(\tilde{x}(t)e^{j2\pi f_c t})$ and $y(t) = \mathrm{Re}(\tilde{y}(t)e^{j2\pi f_c t})$. The complex envelopes are given by

$$\begin{aligned} \tilde{x}(t) &= d\frac{A_1}{2}\int_{t-T_1}^{t} e^{j\theta(\tau)}\,d\tau + \int_{t-T_1}^{t} n_1(\tau)e^{j2\pi f_c \tau}\,d\tau \\ \tilde{y}(t) &= \frac{A_2}{2}\int_{t-T_2}^{t} e^{j\theta(\tau)}\,d\tau + \int_{t-T_2}^{t} n_2(\tau)e^{j2\pi f_c \tau}\,d\tau \end{aligned} \tag{7}$$

for the noncentered filter, where $d = \pm 1$ depending on the data bit. We have assumed that $f_c T_i \gg 1$ $(i = 1, 2)$ so that the double frequency components do not appear at the integrator outputs. The centered filter will have the integral limits in $\tilde{y}(t)$ appropriately modified. The decision variable is then given by $Y = \sum_{k=1}^{M} Y_k$ where $Y_k = \mathrm{Re}[\tilde{x}(kT_1)\tilde{y}^*(kT_1)]$, and the error probability is $P_e = \Pr(Y \le 0 | d = 1)$.

Exact calculation of the error probability is complicated by two phenomena. The conditional error probability given the phase noise process $\theta(t)$ is the probability that the complex inner product of two Gaussian vectors, $[\tilde{x}(kT_1) : k = 1, 2, \ldots, M]$ and $[\tilde{y}(kT_1) : k = 1, 2, \ldots, M]$, has a negative real part. The evaluation of this probability is an involved task even when the entries of the vectors are statistically independent [19]. In our case, however, $[\tilde{y}(kT_1)]$ has dependent entries for $K > 1$, due to overlapping integration windows at successive sampling times. A further problem is the removal of the conditioning on phase noise. Even for the simple case of $M = 1$, where the second filtering stage vanishes and the vectors reduce to scalars, one gets a conditional error probability of the same form as (1) with the arguments containing two correlated random variables that depend on phase noise. Therefore, the exact calculation of the error probability seems to be infeasible.

The approach we take here for predicting the performance and for finding the optimal parameters involves two steps. First, we find a Chernoff bound to the conditional error probability by taking the correlation of additive noise samples into account. Next, we remove the conditioning on phase noise by an approximation that will be described in Section 7.

6 The Conditional Chernoff Bound

In this section, we obtain the Chernoff bound to the error probability conditioned on the phase noise process, and we assess the tightness of the bound by considering some special cases of our formulation.

6.1 General Development

The Chernoff bound to the bit error probability conditioned on the phase noise process $\{\theta(t)\}$ is given by

$$P_e(\theta(t)) \triangleq \Pr(Y < 0 \mid \{\theta(t)\}, d = 1) \leq E\left(e^{-sY} \mid \theta(t)\right) \quad s \geq 0$$

where the expectation in the right hand side is to be taken over the additive noise samples. We first obtain this bound for the noncentered filter and then modify it for the centered filter.

We define the normalized phase noise integrals

$$X_i(k) \triangleq \frac{1}{T_i} \int_{(k-1)T_i}^{kT_i} e^{j\theta(t)}\, dt \quad k = 1, \ldots, M, \quad i = 1, 2 \tag{8}$$

which satisfy $X_2(k) = \frac{1}{K}\sum_{l=0}^{K-1} X_1(k-l)$ since $T_2 = KT_1$. The sampled bandpass filter outputs, $\tilde{x}(kT_1)$ and $\tilde{y}(kT_1)$, are conditionally Gaussian random variables with means $\tilde{a}(k) \triangleq \frac{A_1 T_1}{2} X_1(k)$ and $\tilde{b}(k) \triangleq \frac{A_2 T_2}{2} X_2(k)$ respectively. The additive noise components of $\tilde{x}(kT_1)$ are statistically independent, while those of $\tilde{y}(kT_1)$ are not independent due to the overlapping integration windows. In fact, we have

$$\tilde{x}(kT_1) = \tilde{a}(k) + \tilde{w}(k)$$
$$\tilde{y}(kT_1) = \tilde{b}(k) + \sum_{l=0}^{K-1} \tilde{z}(k-l)$$

where $\tilde{w}(k) = \int_{(k-1)T_1}^{kT_1} n_1(t) e^{j2\pi f_c t}\, dt$, $\tilde{z}(k) = \int_{(k-1)T_1}^{kT_1} n_2(t) e^{j2\pi f_c t}\, dt$ are independent, complex Gaussian random variables with zero mean and component-wise variance $\sigma^2 \triangleq N_0 T_1/4$.

We define the complex M-dimensional column vectors $\vec{a}$, $\vec{b}$ and $\vec{w}$ in the generic form $\vec{\tilde{\alpha}} = [\tilde{\alpha}(1), \ldots, \tilde{\alpha}(M)]^T$ and the $(M+K-1)$-dimensional complex vector $\vec{z} = [\tilde{z}(-K+$

$2), \ldots, \vec{z}(M)]^T$. The latter vector must be of a larger dimension because of the larger time span in the reference integrators. Finally, the $M \times (M + K - 1)$ matrix A is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } 0 \leq j - i \leq K - 1 \\ 0 & \text{otherwise,} \end{cases}$$

hence A is a Toeplitz matrix with K consecutive 1's in each row starting with the diagonal entry.

With these definitions at hand, the decision variable can be written as

$$Y = \mathrm{Re}\left[(\vec{\mathrm{a}} + \vec{\mathrm{w}})^{\mathrm{H}}(\vec{\mathrm{b}} + \mathrm{A}\vec{\mathrm{z}})\right]$$

where the superscript H denotes the complex conjugate transpose. The following lemma provides the result necessary to obtain the Chernoff bound explicitly.

Lemma: Let $\vec{x}$ and $\vec{y}$ be mutually independent, real Gaussian vectors with means $\vec{m}_x$ and $\vec{m}_y$, and covariance matrices $\Lambda_x = \sigma^2 I$, $\Lambda_y = B$ for some positive definite matrix B. Then the moment generating function of $\vec{x}^T\vec{y}$ is given by

$$E\left(\exp(-s\vec{x}^T\vec{y})\right) = |I - s^2\sigma^2 B|^{-1/2} \exp\left(-s\vec{m}_x{}^T\vec{m}_y + \frac{s^2\sigma^2}{2}\|\vec{m}_y\|^2 + \frac{1}{2}\vec{\alpha}^T\Lambda^{-1}\vec{\alpha}\right)$$

where $\vec{\alpha} = -s\vec{m}_x + s^2\sigma^2\vec{m}_y$, $\Lambda = B^{-1} - s^2\sigma^2 I$, and $|\cdot|$ denotes the determinant of the argument.

Proof: Since $\vec{x}$ has statistically independent elements, it is convenient to first condition the expectation on $\vec{y}$. Then it is easily observed from the moment generating function of Gaussian random variables that

$$E\left[\exp(-s\vec{x}^T\vec{y}) \mid \vec{y}\right] = \exp\left(-s\vec{m}_x{}^T\vec{y} + \frac{s^2\sigma^2}{2}\|\vec{y}\|^2\right) .$$

Then the desired expectation is given by

$$(2\pi)^{-M/2}|B|^{-1/2}\int \exp\left(-s\vec{m}_x{}^T\vec{y} + \frac{s^2\sigma^2}{2}\|\vec{y}\|^2 - \frac{1}{2}(\vec{y} - \vec{m}_y)^T B^{-1}(\vec{y} - \vec{m}_y)\right) d\vec{y}$$

which can be evaluated by completing the exponent to a quadratic to obtain the desired result. □

Using the lemma with $B = \sigma^2 AA^T$ and appropriate mean vectors yields the Chernoff bound as

$$\begin{aligned} P_e(\theta(t)) \leq |I - s^2\sigma^4 AA^T|^{-1} \exp\Big[\ \mathrm{Re}\ \Big(& -s\vec{a}^H\vec{b} + \frac{s^2\sigma^2}{2}\|\vec{b}\|^2 \\ & + \frac{s^2\sigma^2}{2}(\vec{a} - s\sigma^2\vec{b})^H(I - s^2\sigma^4 AA^T)^{-1}AA^T(\vec{a} - s\sigma^2\vec{b})\Big)\Big] . \end{aligned}$$

The bound is well-defined for $s^2\sigma^4\lambda_{\max}(AA^T) < 1$, where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue.

We now express the bound in terms of fundamental parameters and random variables. Using the definition of phase noise integrals in (8) one obtains after some manipulations

$$P_e(\vec{X}_1, \vec{X}_2) \leq |C| \exp\left[-\frac{2uK}{M}\sqrt{\xi_1\xi_2}t_1 + \frac{u^2K^2}{M}\xi_2 t_2 + \frac{u^2}{M}\xi_1 t_3\right] \tag{9}$$

where $u = s\sigma^2$ is the normalized Chernoff bound parameter, $C = (I - u^2AA^T)^{-1}$, and t_i, $i = 1, 2, 3$, are given as

$$\begin{aligned} t_1 &= \text{Re } \vec{X}_2^H C \vec{X}_1 \\ t_2 &= \vec{X}_2^H C \vec{X}_2 \\ t_3 &= \vec{X}_1^H C A A^T \vec{X}_1 \end{aligned}$$

for $0 \leq u < u_{\max} = \lambda_{\max}^{-1/2}(AA^T)$. In obtaining (9) we have also used the matrix identity $I + (I - B)^{-1}B = (I - B)^{-1}$ with $B = u^2AA^T$.

We now consider some special cases of the conditional Chernoff bound where its tightness may be estimated.

6.2 Special Cases for the Chernoff Bound

The formulation of the Chernoff bound can be applied to a variety of specific problems. We consider three of these problems.

1. Frequency Shift Keying: As explained in Section 5, setting $K = 1$ in (9) results in a bound for the conditional error probability of incoherent double filter reception of wide deviation FSK for which performance results are known. In this case $\vec{X}_1 = \vec{X}_2 = \vec{X}$, and A becomes an M-dimensional unit matrix. Therefore, the bound simplifies to

$$P_e(\vec{X}, K = 1) \leq (1 - u^2)^{-M} \exp\left[\frac{\|\vec{X}\|^2}{M(1 - u^2)}\left(u^2(\xi_1 + \xi_2) - 2u\sqrt{\xi_1\xi_2}\right)\right]$$

with $\xi_1 + \xi_2 = \xi$. The optimal power distribution is symmetric, a fact also observed in Section 5. This distribution results in

$$P_e(\vec{X}, K = 1) \leq (1 - u^2)^{-M} \exp\left[-\frac{u}{1+u}\frac{\xi}{M}\|\vec{X}\|^2\right] \quad 0 \leq u < 1 \, . \tag{10}$$

Note that $\|\vec{X}\|^2$ is now a sum of M independent identically distributed random variables, and $\|\vec{X}\|^2/M \leq 1$ is the reduction in SNR due to phase noise [3]. If we further specialize to the case of no phase noise, where $\|\vec{X}\|^2 = M$, and the optimal value of M becomes 1 resulting in the conventional envelope detector. Thus for phase noise free FSK we obtain the unconditional bound

$$P_e \leq \min_{0 \leq u < 1} (1 - u^2)^{-1} \exp\left(-\frac{u\xi}{1+u}\right) \tag{11}$$

which has the same exponential nature $e^{-\xi/2}$ as the true error probability as $u \to 1$. However since the coefficient also grows with u the bound is not arbitrarily tight as seen in Figure 5 ($\gamma = 0$ curves). Nonetheless it is still within 0.7 dB of the actual performance for low error probabilities. Parametric optimization of (11) for $\xi \gg 1$ yields the bound as $\xi e^{-\xi/2}/8$. (Here we first find the value of ξ for which u is optimal as $\xi = 2u(1+u)/(1-u)$, and we express the bound in terms of ξ when $u \simeq 1$.)

This example shows that Chernoff bound has the potential of retaining the essential features of the actual performance, e.g. the optimal power distribution, the filter bandwidth setting, and the rate of exponential decay of the error probability.

2. Single Filter Receiver: Another interesting special case of the Chernoff bound is the single filter receiver. In this case, A becomes a row vector of K 1's, so $AA^T = K$, and $\vec{X}_1, \vec{X}_2$ become scalars. The bound reduces to

$$P_e(\vec{X}_1, \vec{X}_2) \leq (1 - Ku^2)^{-1} \exp\left[\frac{Ku^2(\xi_1|X_1|^2 + K\xi_2|X_2|^2) - 2uK\sqrt{\xi_1\xi_2}\mathrm{Re}(X_1X_2^*)}{M(1-Ku^2)}\right] .$$

In the case with no phase noise, $X_1 = X_2 = 1$. Then the optimal value of M is 1. Letting $v = \sqrt{K}u$, one finds that the value of v that minimizes the numerator of the exponent is

$$v^* = \frac{\sqrt{K\xi_1\xi_2}}{\xi_1 + K\xi_2} \leq \frac{1}{2}$$

with the resulting numerator being $-K\xi_1\xi_2/(\xi_1 + K\xi_2)$. We can bound the denominator and the leading coefficient using $\frac{3}{4} \leq 1 - (v^*)^2 \leq 1$ to obtain the bound

$$P_e \leq \frac{4}{3} \exp\left(-\frac{K\xi_1\xi_2}{\xi_1 + K\xi_2}\right)$$

which is to be minimized over ξ_1, ξ_2, K and M subject to the constraint $\xi_1 + \xi_2 = \xi$. The minimum occurs with $\xi_2 \to 0$ and $K \to \infty$ such that $K\xi_2 \to \infty$. This results in

$$P_e \leq \frac{4}{3} e^{-\xi}$$

which has the same exponential character as PSK performance. The optimal solution corresponds to sending no power in the reference signal but getting a perfect local reference, which is exactly the case for phase noise free PSK. This example also confirms our expectation that the conclusions reached from the Chernoff bound are likely to correspond to the actual system.

3. Double Filter DPSK: A final application of our bound is the case of double filter DPSK. Here we have the current and previous bits replacing the data and reference signals. Both signals have the same SNR and the filter processing is identical, i.e. $\xi_1 = \xi_2 = \xi$, $K = 1$. The reference vector $\vec{X}_2$ has to be modified so as to include the time delay as

$$X_2(k) = \frac{1}{T_1} \int_{(k-1)T_1 - T}^{kT_1 - T} e^{j\theta(t)}\, dt\ .$$

The conditional Chernoff bound with these parameters is

$$P_e(\vec{X}_1, \vec{X}_2) \leq (1-u^2)^{-M} \exp\left[\frac{1}{1-u^2}\frac{\xi}{M}\left(u^2(\|\vec{X}_1\|^2 + \|\vec{X}_2\|^2) - 2u\,\mathrm{Re}\vec{X}_2^H \vec{X}_1\right)\right]$$

for $0 \leq u < 1$. If we further specialize to the case with no phase noise we obtain

$$P_e \leq \min_{0 \leq u < 1} (1-u^2)^{-1} \exp\left(-\frac{2u\xi}{1+u}\right) \tag{12}$$

which is exactly 3 dB better than the Chernoff bound for FSK in (11), thus retains all the desirable relations to the actual performance.

7 Chernoff–Jensen Approximation

7.1 Noncentered Reference Filter

The conditional Chernoff bound we obtained in Equation (9) depends on quadratic terms of phase noisy vectors $\vec{X}_1$ and $\vec{X}_2$. These quadratics have dependent components; therefore we must know the joint statistics of $\vec{X}_1$ and $\vec{X}_2$ to be able to remove the conditioning on the phase noise process. The statistics of $|X_1(k)|^2$ plays a fundamental role in the performance analysis of incoherent reception of FSK and OOK [2, 3, 4]. In the light of the difficulty of evaluating the probability distribution of this random variable only, the exact removal of the conditioning is likely to be an infeasible task. Even in the single filter case the joint statistics of the random variable pair $|\sqrt{\xi_1}X_1 \pm \sqrt{\xi_2}X_2|^2$ is required to perform the unconditioning. Therefore we pursue an approximate technique to estimate the performance.

Note that the conditional Chernoff bound is of the form

$$P_e(\vec{X}_1, \vec{X}_2) \leq \alpha \exp\left[f(\vec{X}_1, \vec{X}_2)\right]$$

where both the coefficient α and the function $f(\cdot)$ depend on the parameters u, M, K, ξ_1, and ξ_2. Although evaluating the expectation of the right hand side would effectively require the complete statistics (moment generating function) of $f(\cdot)$, the expected value of $f(\cdot)$ can be found with elementary computations. Thus by interchanging the expectation and exponentiation operators, we can obtain an approximation to the error probability P_e. This will not result in an upper bound on P_e as $\exp(\cdot)$ is a convex $\cup$ function.

A similar approach was employed in [3, 4] to find a lower bound to the error probability for envelope detection of FSK and OOK signals. In these cases, the conditional error probabilities could be found exactly and were convex $\cup$ functions, thus resulting in Jensen bounds. It was found that the bounds were very close to more reliable performance estimates. This encourages us to employ the approximation above which will be referred to as *Chernoff–Jensen approximation* to reflect both the method by which the conditional bound is obtained and the method by which the conditioning is removed. It is expected that due to the optimization of the Chernoff bound and the averaging effect of the double filter structure the excursion from the true performance will not be significant. It is immediately seen that Chernoff–Jensen approximation overemphasizes the additive noise by upper bounding its effect, and underemphasizes the phase noise by taking the phase noisy quadratics at their means. Therefore, the values of M that will be predicted by our analysis will be smaller than the optimal values, and conversely the predicted values of K will be larger than the optimal.

A check of the approximation is the FSK case considered in Section 6.2. From (10), the approximation yields

$$P_{e,CJ} = \min_{u,M}(1-u^2)^{-M} \exp\left[-\frac{u}{1+u}\xi\bar{X}(\gamma/M)\right]$$

where $\bar{X}(\cdot)$ is the mean of the squared envelope $|X_1(k)|^2$ and is given by [20, 21, 3]

$$\bar{X}(\gamma) = \frac{4}{\gamma}\left[1-\frac{2}{\gamma}(1-e^{-\gamma/2})\right] .$$

This approximation is shown in Figure 5 together with a more reliable performance estimate from [3]. For each γ value, the Chernoff-Jensen approximation results in a larger error probability than the prediction of [3]. It is seen that the two results are in satisfactory agreement.

Now we apply the Chernoff–Jensen approximation to the double filter transmitted reference system. The conditional Chernoff bound in (9) requires three expectations to be found, $\bar{t_1}$, $\bar{t_2}$ and $\bar{t_3}$. Let's define the correlation matrices of the random vectors $\vec{X_1}$ and $\vec{X_2}$ as

$H_{ij} = E\vec{X}_i\vec{X}_j^H$, $i,j = 1,2$. It is easily seen that the three expectations can be written as $\bar{t_1} = \mathrm{tr}(CH_{21})$, $\bar{t_2} = \mathrm{tr}(CH_{22})$, $\bar{t_3} = \mathrm{tr}(DH_{11})$ where $C = (I - u^2AA^T)^{-1}$ as previously defined, $D = CAA^T$ and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix[5]. The correlation matrices are computed in Appendix, resulting in the following expressions:

$$H_{11}(i,k) = \begin{cases} \frac{2}{\alpha} - \frac{2}{\alpha^2}(1-e^{-\alpha}) & k = i \\ \left(\frac{1-e^{-\alpha}}{\alpha}\right)^2 e^{-\alpha(|k-i|-1)} & k \neq i \end{cases} \tag{13}$$

$$H_{22}(i,k) = \begin{cases} \frac{2}{\mu} - \frac{2}{\mu^2}(1-e^{-\mu}) & k = i \\ \frac{2}{\mu}\left(1 - \frac{|k-i|}{K}\right) + \frac{1}{\mu^2}\left[e^{-\alpha(K-|k-i|)} + (e^{-2\mu} - 2e^{-\mu})e^{\alpha(K-|k-i|)}\right] & 0 < |k-i| < K \\ \left(\frac{1-e^{-\mu}}{\mu}\right)^2 e^{-\alpha(|k-i|-K)} & |k-i| \geq K \end{cases} \tag{14}$$

$$H_{21}(i,k) = \begin{cases} \frac{1}{\mu\alpha}(1-e^{-\mu})(1-e^{-\alpha})e^{-\alpha(k-i-1)} & i < k \\ \frac{2}{\mu} - \frac{1}{\mu\alpha}(1-e^{-\alpha})\left[e^{-\alpha(i-k)} + e^{-\alpha(K-1-(i-k))}\right] & 0 \leq i-k < K \\ \frac{1}{\mu\alpha}(1-e^{-\mu})(1-e^{-\alpha})e^{-\alpha(i-k-K)} & i-k \geq K \end{cases} \tag{15}$$

where $\alpha \triangleq \pi\beta T_1 = \gamma/2M$, $\mu \triangleq \pi\beta T_2 = K\gamma/2M$.

The Chernoff–Jensen approximation is obtained as

$$P_e \simeq |C|\exp\left[-\frac{2uK}{M}\sqrt{\xi_1\xi_2}\,\mathrm{tr}(CH_{21}) + \frac{u^2K^2\xi_2}{M}\,\mathrm{tr}(CH_{22}) + \frac{u^2\xi_1}{M}\,\mathrm{tr}(DH_{11})\right]. \tag{16}$$

This approximation is to be minimized over u, ξ_1, ξ_2, K and M subject to the constraints $\xi_1 + \xi_2 = \xi$, $0 \leq u < \lambda_{\max}^{-1/2}(AA^T)$, $K, M \in \{1,2,\ldots\}$. The results of this optimization will be discussed in Section 8. First we describe the necessary modification for the performance of the centered reference filter.

7.2 Centered Reference Filter

The analysis of Section 7.1 needs only minor modifications to be extended to the case of a centered reference filter which was described in Section 5. The phase noisy reference vector $\vec{X}_2$ is now given by

$$X_2(k) = \frac{1}{T_2}\int_{(k-1/2)T_1-T_2/2}^{(k-1/2)T_1+T_2/2} e^{j\theta(t)}\,dt$$

It is convenient to assume that K is odd since this will ensure that the reference integration window consists of an integral number of chips of the information integration window. Then

[5]The $\mathrm{Re}(\cdot)$ in $\bar{t_1}$ can be omitted since H_{21} is real as will be seen.

with $K = 2K' - 1$, one can write

$$X_2(k) = \frac{1}{K} \sum_{l=-(K'-1)}^{K'-1} X_1(k-l) \, .$$

The conditional Chernoff bound of (9) remains unchanged because the additive noise structure is the same. The only change will be in the Jensen approximation due to new correlation matrices. H_{11} is still given by (13) since $\vec{X}_1$ is not changed. H_{22} is also unchanged and given by (14) since a shift in the time origin does not affect the correlation between $X_2(i)$ and $X_2(k)$. The cross correlation between $\vec{X}_1$ and $\vec{X}_2$ will change however. In fact, this is precisely the reason we introduced this new filter structure. In Appendix we obtain the new cross correlation matrix $\bar{H}_{21}$ as

$$\bar{H}_{21}(i,k) = \begin{cases} \frac{2}{\mu} - \frac{1}{\alpha\mu}(1-e^{-\alpha})\left[e^{-\alpha(K'-1-|k-i|)} + e^{-\alpha(K'-1+|k-i|)}\right] & |k-i| < K' \\ \frac{1}{\alpha\mu}(1-e^{-\alpha})(1-e^{-\mu})e^{-\alpha(|k-i|-K')} & |k-i| \geq K' \, . \end{cases} \tag{17}$$

Note that $\bar{H}_{21}$ is symmetric while H_{21} was not, this is due to the symmetry of the integration windows.

Equation (16) will yield the performance for the system with the centered filter with the substitution of $\bar{H}_{21}$ for H_{21}. We now discuss the performance.

8 Results and Discussions

The Chernoff-Jensen approximation to the bit error probability has been numerically optimized over u, ξ_1, ξ_2, K and M subject to the constraints $u \geq 0$, $\xi_1 + \xi_2 \leq \xi$ and $K, M \in \{1, 2, \ldots\}$ for both the centered and noncentered reference filters. The resulting performance for the noncentered filter is shown in Figure 6 as a function of SNR ξ for various values of the phase noise strength γ. The error probability of double filter FSK obtained by the same approximation is also shown in the same figure for comparison. It is observed that for $\gamma = 0.01$ the transmitted reference system has a 2.5 dB advantage over FSK, this reduces to 1.8 dB for $\gamma = 0.1$. The advantage of noncentered transmitted reference scheme over FSK vanishes for $\gamma \geq 1$ as the respective performance curves become identical. This is because for large values of γ the difference in the bandwidth occupancies of the information and reference signals is small. Hence the respective filters become identical ($K = 1$), this in turn imposes an even power distribution and thus FSK performance.

The performance of the system with a centered reference filter is shown in Figure 7 together with the FSK performance. It is seen that this system has a considerably better phase noise tolerance than its noncentered counterpart. For example, there is a distinct performance improvement over FSK with $\gamma = 4$.

Figure 8 shows the optimal values for K and M for various values of phase noise strength γ at an error probability of 10^{-9}, as well as the optimal distribution of power ξ_1/ξ_2. It is seen that the optimal reference filter bandwidths for the centered system are narrower than those of noncentered system. Consequently a smaller proportion of the total power is allocated to the reference signal. The optimal value of K for the centered system is about twice that of the noncentered one; this is due to the improved correlation between the filter outputs. For both filter types K becomes 1 with increasing γ as expected.

The phase noise induced SNR penalties of the two transmitted reference systems with respect to ideal DPSK are shown in Figure 9. The penalty for double filter FSK, obtained using the same approximation for consistency, is also shown. For small values of the phase noise strength γ, both reference transmission schemes have considerably better performance than FSK. While the system with noncentered filter saturates to FSK performance at about $\gamma = 1$, the system with centered filter has an improved performance up to $\gamma = 12$. For $\gamma = 1$, the centered filter has a gain of 0.8 dB over FSK.

9 Conclusion

The transmission of a reference at an optimal power distribution and a double filter receiver with optimal filter bandwidths and proper filter structures achieves a performance that is better than conventional modulation schemes. The performance gain is particularly pronounced when a centered reference filter is used. At the high linewidth to bit rate ratio regime, Frequency Shift Keying with double filter envelope detection is preferable to a reference transmission scheme, as the two systems achieve the same bit error probability. At the low linewidth to bit rate ratio regime, a properly optimized reference transmission scheme has considerable performance improvement over FSK. As advances in fiber-optic communication technology brings high speed transmission systems with low linewidth lasers into reality, the reference transmission scheme described in this paper may achieve performances that are close to those of ideal Phase Shift Keyed systems. The approach of jointly optimizing the signaling and receiver parameters may also prove useful in other optical communication systems impaired by phase noise and other nonideal phenomena.

Appendix

A Calculation of Correlation Matrices

In this appendix, we obtain the auto- and cross-correlation matrices of the phase noisy vectors $\vec{X}_1$ and $\vec{X}_2$. We start with the autocorrelation matrix of $\vec{X}_1$, H_{11}. From the definition of $X_1(k)$ we have

$$\begin{aligned} H_{11}(i,k) &= E X_1(i) X_1^*(k) \\ &= \frac{1}{T_1^2} \int_{(i-1)T_1}^{iT_1} \int_{(k-1)T_1}^{kT_1} E\, e^{-j(\theta(t)-\theta(\tau))}\, d\tau\, dt\ . \end{aligned}$$

Now using the fact that $\theta(t) - \theta(\tau)$ is Gaussian with variance $2\pi\beta|t-\tau|$, we obtain

$$H_{11}(i,k) = \frac{1}{T_1^2} \int_{(i-1)T_1}^{iT_1} \int_{(k-1)T_1}^{kT_1} e^{-\pi\beta|t-\tau|}\, d\tau\, dt\ .$$

At this point there are two cases to consider: $i = k$ and $i \neq k$. For the first case we get by symmetry

$$\begin{aligned} H_{11}(i,i) &= \frac{2}{T_1^2} \int\int_{t>\tau} e^{-\pi\beta(t-\tau)}\, d\tau\, dt \\ &= \frac{2}{\alpha} - \frac{2}{\alpha^2}(1 - e^{-\alpha}) \end{aligned} \tag{A.1}$$

where we have defined $\alpha = \pi\beta T_1$. For $i \neq k$, let's first assume $i > k$. Then

$$\begin{aligned} H_{11}(i,k) &= \frac{1}{T_1^2} \int_{(i-1)T_1}^{iT_1} e^{-\pi\beta t}\, dt \int_{(k-1)T_1}^{kT_1} e^{\pi\beta\tau}\, d\tau \\ &= \left(\frac{1-e^{-\alpha}}{\alpha}\right)^2 e^{-\alpha(|i-k|-1)} \end{aligned} \tag{A.2}$$

where we have introduced the absolute value in the second line to make (A.2) valid for $i < k$ by symmetry. Equations (A.1) and (A.2) completely define the autocorrelation matrix H_{11} as given in Equation (13).

Now we consider the autocorrelation matrix H_{22} of $\vec{X}_2$. This can be written as the following integral

$$H_{22}(i,k) = \int_{iT_1-T_2}^{iT_1} \int_{kT_1-T_2}^{kT_1} e^{-\pi\beta|t-\tau|}\, dt\, d\tau\ .$$

For $i = k$, this is similar to H_{11}. Thus substituting T_2 for T_1 we get

$$H_{22}(i,i) = \frac{2}{\mu} - \frac{2}{\mu^2}(1 - e^{-\mu}) \tag{A.3}$$

where we have defined $\mu = K\alpha$. For the case $|i-k| \geq K$, the two integrals split with the result easily obtained as

$$H_{22}(i,k) = \left(\frac{1-e^{-\mu}}{\mu}\right)^2 e^{-\alpha(|i-k|-K)} . \tag{A.4}$$

Finally for the case $0 < |i-k| < K$, one has to consider the two regions in which the absolute value in the exponent of the integrand take different values. After a tedious, but straightforward calculation, one gets

$$H_{22}(i,k) = \frac{2}{\mu}\left(1 - \frac{|k-i|}{K}\right) + \frac{1}{\mu^2}\left[e^{-\alpha(K-|k-i|)} + (e^{-2\mu} - 2e^{-\mu})e^{\alpha(K-|k-i|)}\right] . \tag{A.5}$$

Equations (A.3), (A.4) and (A.5) define the autocorrelation matrix H_{22} as given in Equation (14).

Next we want to calculate the autocorrelation matrix H_{21}. Instead of performing a similar calculation to that of H_{11} which would be more involved due to the time overlaps between the defining integrals, we will recall the relation

$$X_2(i) = \frac{1}{K}\sum_{l=0}^{K-1} X_1(i-l)$$

and write the desired correlation as

$$H_{21}(i,k) = \frac{1}{K}E\, X_1(k)\sum_{l=0}^{K-1} X_1^*(i-l) = \frac{1}{K}\sum_{l=0}^{K-1} H_{11}(k, i-l) .$$

The result of the above will depend on whether the diagonal entries of H_{11} enter the summation or not. When $i-k \geq K$, or when $i < k$, the sum consists of nondiagonal entries only, while when $0 \leq i-k < K$ we have both types of terms in the series. For the case $i-k \geq K$, by using (A.2) we obtain

$$\begin{aligned} H_{21}(i,k) &= \frac{1}{K}\left(\frac{1-e^{-\alpha}}{\alpha}\right)^2 \sum_{l=0}^{K-1} e^{-\alpha(i-k-l-1)} \\ &= \frac{1}{\alpha\mu}(1-e^{-\alpha})(1-e^{-\mu})e^{-\alpha(i-k-K)} . \end{aligned} \tag{A.6}$$

Similarly, for the case $i < k$ one gets

$$\begin{aligned} H_{21}(i,k) &= \frac{1}{K}\left(\frac{1-e^{-\alpha}}{\alpha}\right)^2 \sum_{l=0}^{K-1} e^{-\alpha(k-i+l-1)} \\ &= \frac{1}{\alpha\mu}(1-e^{-\alpha})(1-e^{-\mu})e^{-\alpha(k-i-1)} . \end{aligned} \tag{A.7}$$

For the case $0 \leq i-k < K$, one has the consider how the diagonal entries enter the series. Since H_{11} is symmetric and Toeplitz, we can use the convention $H_{11}(i,k) = h(|i-k|)$. With the convention, we can write for the case under consideration

$$H_{21}(i,k) = \frac{1}{K}\left[h(0) + 2\sum_{n=1}^{n_1} h(n) + \sum_{n=n_1+1}^{n_2} h(n)\right]$$

where $n_1 = \min(i-k, K-1-(i-k))$ and $n_2 = \max(i-k, K-1-(i-k))$. This yields

$$H_{21}(i,k) = \frac{2}{\mu} - \frac{1}{\alpha\mu}(1-e^{-\alpha})\left[e^{-\alpha(i-k)} + e^{-\alpha(K-1-(i-k))}\right] . \tag{A.8}$$

Equations (A.6),(A.7) and (A.8) define the cross-correlation matrix H_{21} as given in Equation (15).

Finally we want to calculate the cross-correlation matrix $\bar{H}_{21}$ for the case of the centered reference filter given in Section 7.2. Now the reference vector $\vec{X}_2$ is defined as

$$X_2(i) = \frac{1}{K}\sum_{l=-(K'-1)}^{K'-1} X_1(i-l)$$

so that the elements of $\bar{H}_{21}$ are

$$\bar{H}_{21}(i,k) = \frac{1}{K}\sum_{l=-(K'-1)}^{K'-1} H_{11}(k, i-l) .$$

Note that the only changes from the calculation of H_{21} are the upper and lower limits of the defining series. For $|i-k| \geq K'$, the series contains only the off-diagonal elements of H_{11} and the result is

$$\begin{aligned}\bar{H}_{21}(i,k) &= \frac{1}{K}\left(\frac{1-e^{-\alpha}}{\alpha}\right)^2 \sum_{l=-(K'-1)}^{K'-1} e^{-\alpha(|i-k|-l-1)} \\ &= \frac{1}{\alpha\mu}(1-e^{-\alpha})(1-e^{-\mu})e^{-\alpha(|k-i|-K')} .\end{aligned} \tag{A.9}$$

For the case with $|i-k| < K'$, an accounting of the terms shows that, with the same convention as before, we get

$$\begin{aligned}\bar{H}_{21}(i,k) &= \frac{1}{K}\left[h(0) + 2\sum_{n=1}^{K'-1-|i-k|} h(n) + \sum_{n=K'-|i-k|}^{K'-1+|i-k|} h(n)\right] \\ &= \frac{2}{\mu} - \frac{1}{\alpha\mu}(1-e^{-\alpha})\left[e^{-\alpha(K'-1-|i-k|)} + e^{-\alpha(K'-1+|i-k|)}\right] .\end{aligned} \tag{A.10}$$

Equations (A.9) and (A.10) define the new cross-correlation matrix as given in Equation (17).

References

[1] C. H. Henry, "Theory of linewidth of semiconductor lasers," *IEEE Journal of Quantum Electronics*, vol. QE-18, pp. 259–264, February 1982.

[2] G. J. Foschini, L. J. Greenstein, and G. Vannucci, "Noncoherent detection of coherent lightwave signals corrupted by phase noise," *IEEE Transactions on Communications*, vol. COM-36, pp. 306–314, March 1988.

[3] M. Azizoglu and P. A. Humblet, "Envelope detection of orthogonal signals with phase noise," *Journal of Lightwave Technology*, vol. 9, pp. 1398–1410, October 1991.

[4] M. Azizoglu and P. A. Humblet, "Performance of on-off modulated lightwave signals with phase noise," in *SPIE OE/Fibers '91*, (Boston, MA), September 3–6 1991.

[5] A. Yariv, *Optical Electronics.* Saunders, HRW, fourth ed., 1991.

[6] K. Tamura, S. B. Alexander, V. W. S. Chan, and D. M. Boroson, "Phase-noise canceled differential phase-shift-keying (PNC-DPSK) for coherent optical communication systems," *Journal of Lightwave Technology*, vol. 8, pp. 190–201, February 1990.

[7] S. Betti, F. Curti, G. D. Marchis, and E. Iannone, "Phase noise and polarization state insensitive optical coherent systems," *Journal of Lightwave Technology*, vol. 8, pp. 756–767, May 1990.

[8] I. Bar-David and J. Salz, "On dual optical detection: Homodyne and transmitted-reference heterodyne reception," *IEEE Transactions on Communications*, vol. 36, pp. 1309–1315, December 1988.

[9] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques.* New York: McGraw-Hill, 1966.

[10] H. L. Van Trees, *Detection, Estimation, and Modulation Theory.* New York: John Wiley & Sons, 1968.

[11] S. Benedetto, E. Biglieri, and V. Castellani, *Digital Transmission Theory.* Englewood Cliffs, NJ: Prentice-Hall, 1987.

[12] J. Salz, "Coherent lightwave communications," *AT&T Technical Journal*, vol. 64, pp. 2153–2209, December 1985.

[13] M. Azizoglu, P. A. Humblet, and J. S. Young, "Phase noise induced performance limits for DPSK modulation with and without frequency feedback." Submitted to IEEE Journal of Lightwave Technology, 1992.

[14] G. J. Foschini and G. Vannucci, "Characterizing filtered light waves corrupted by phase noise," *IEEE Transactions on Information Theory*, vol. IT-34, pp. 1437–1448, November 1988.

[15] I. Garrett, D. J. Bond, J. B. Waite, D. S. L. Lettis, and G. Jacobsen, "Impact of phase noise in weakly coherent systems: A new and accurate approach," *Journal of Lightwave Technology*, vol. 8, pp. 329–337, March 1990.

[16] L. G. Kazovsky and O. K. Tonguz, "ASK and FSK coherent lightwave systems: A simplified approximate analysis," *Journal of Lightwave Technology*, vol. 8, pp. 338–352, March 1990.

[17] Y. E. Dallal and S. Shamai (Shitz), "An upper bound on the error probability of quadratic-detection in noisy phase channels," *IEEE Transactions on Communications*, vol. 39, pp. 1635–1650, November 1991.

[18] G. Jacobsen, "Multichannel system design using optical preamplifiers and accounting for the effects of phase noise, amplifier noise, and receiver noise," *Journal of Lightwave Technology*, vol. 10, pp. 367–377, March 1992.

[19] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 1983.

[20] L. G. Kazovsky, P. Meissner, and E. Patzak, "ASK multiport optical homodyne systems," *Journal of Lightwave Technology*, vol. 5, pp. 770–791, June 1987.

[21] J. R. Barry and E. A. Lee, "Performance of coherent optical receivers," *Proceedings of the IEEE*, vol. 78, pp. 1369–1394, August 1990.

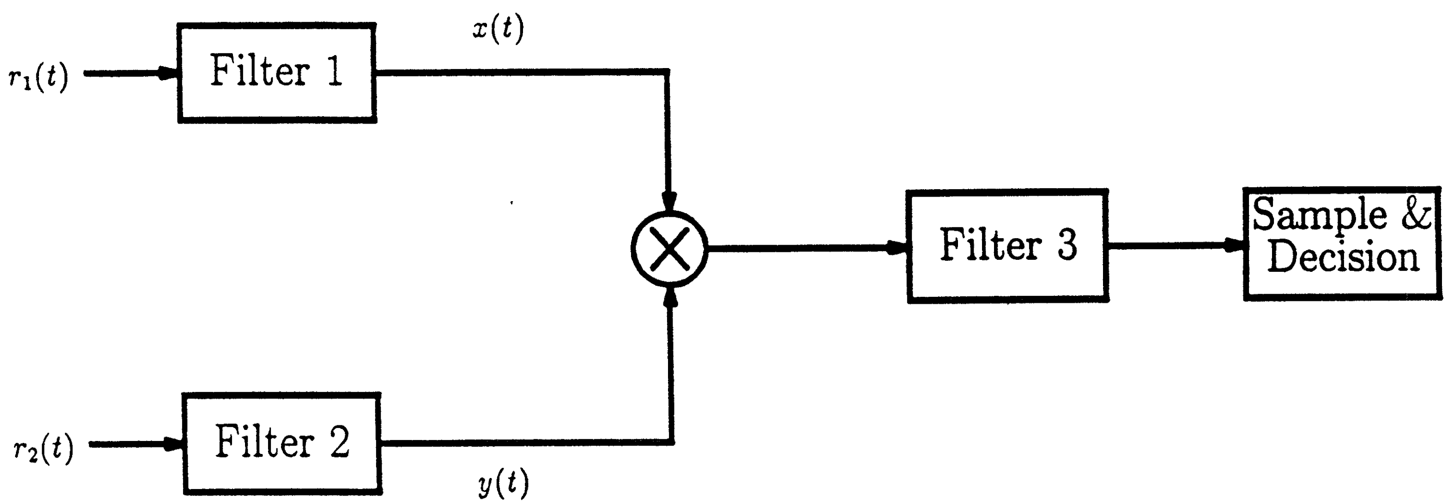

Figure 1: The IF receiver structure.

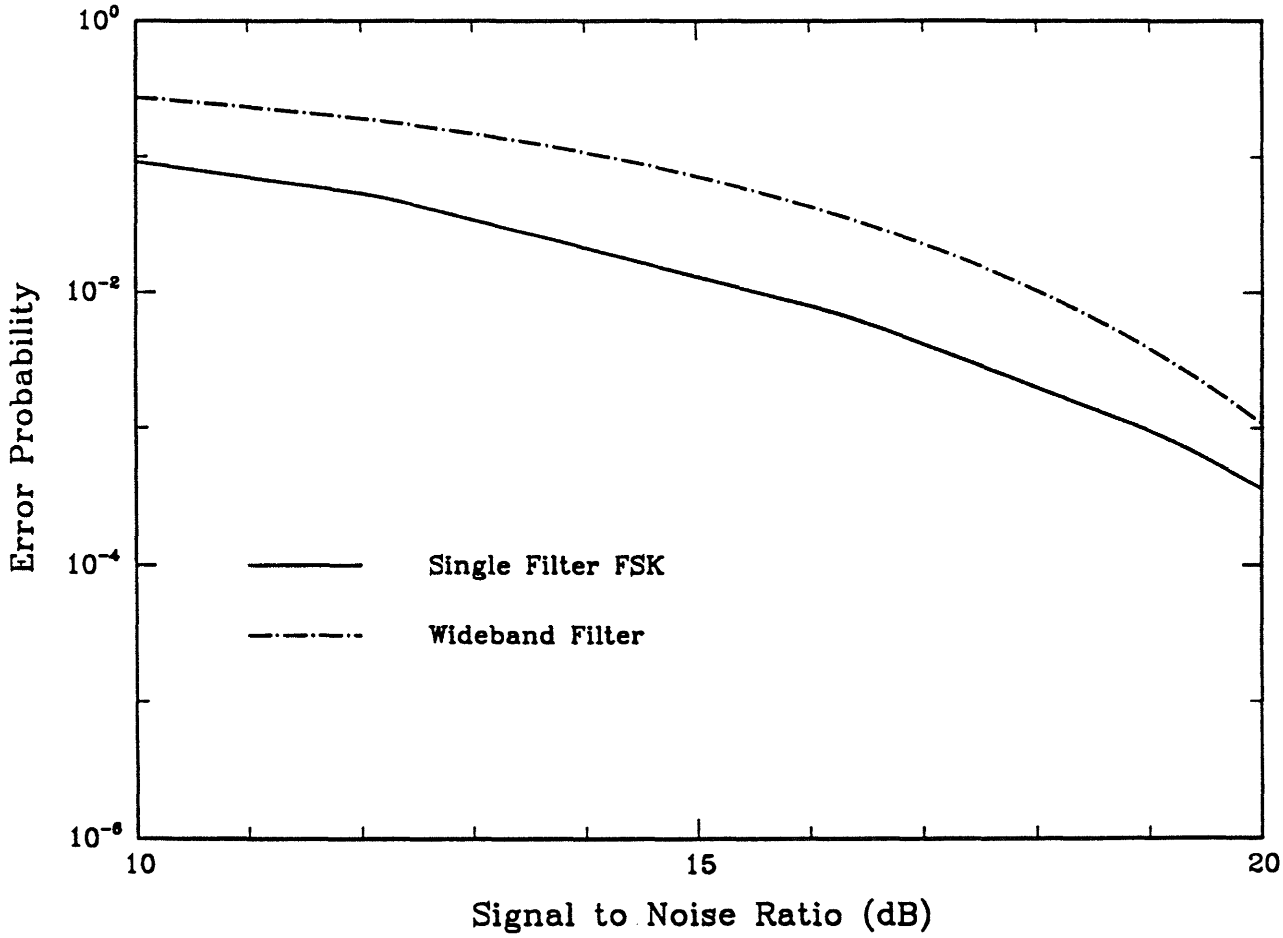

Figure 2: Performance of wideband single filter receiver vs. single filter FSK for $\gamma = 6$.

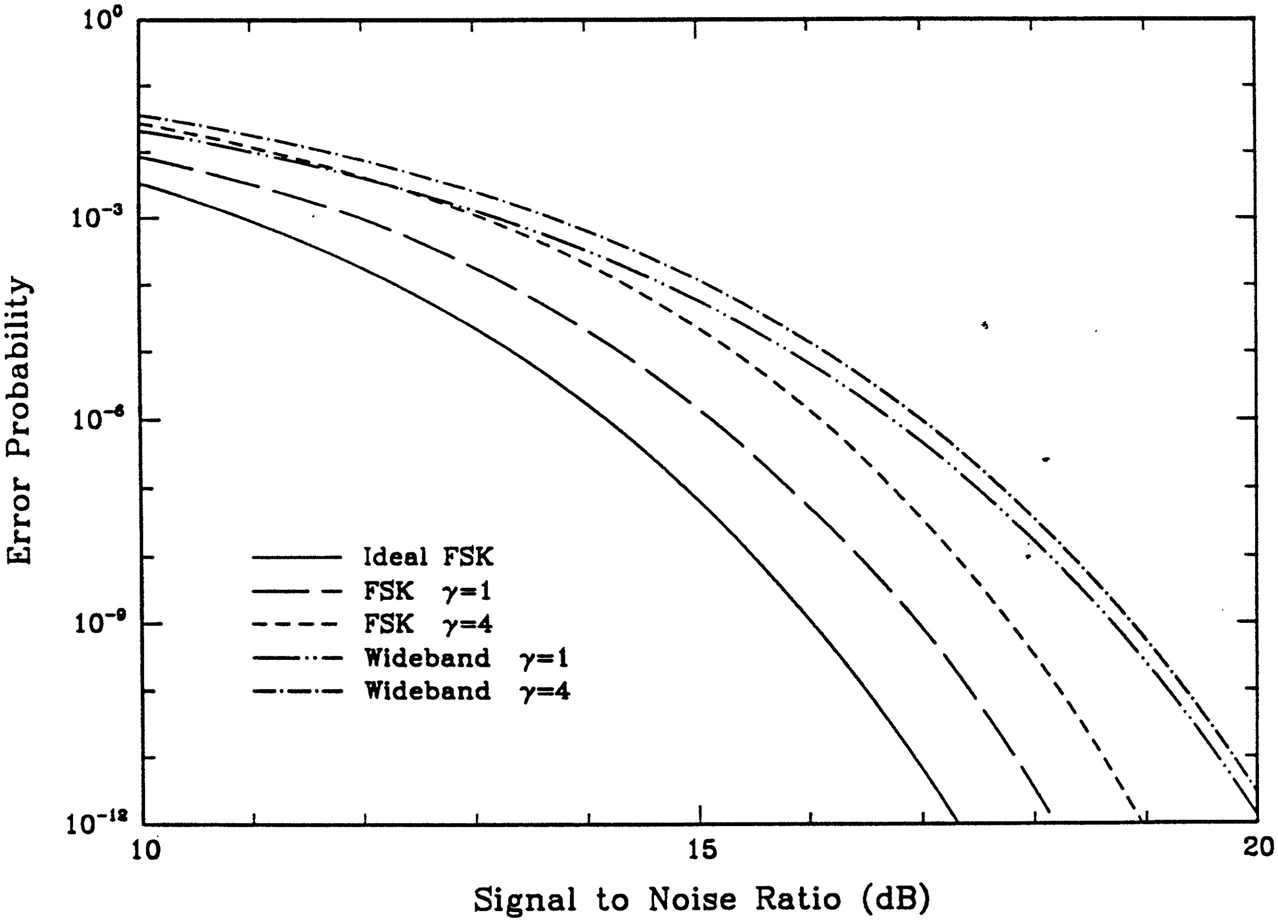

Figure 3: Performance of wideband double filter receiver vs. double filter FSK for $\gamma = 1$ and $\gamma = 4$.

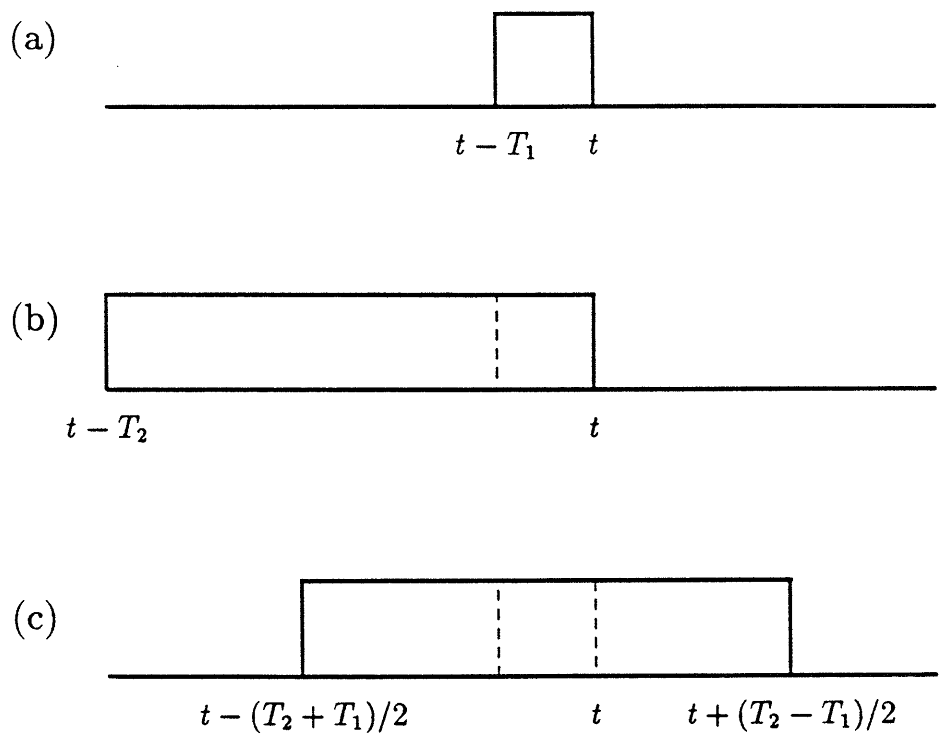

Figure 4: The timing diagram of the integration windows: (a) The information filter, (b) Noncentered reference filter, (c) Centered reference filter.

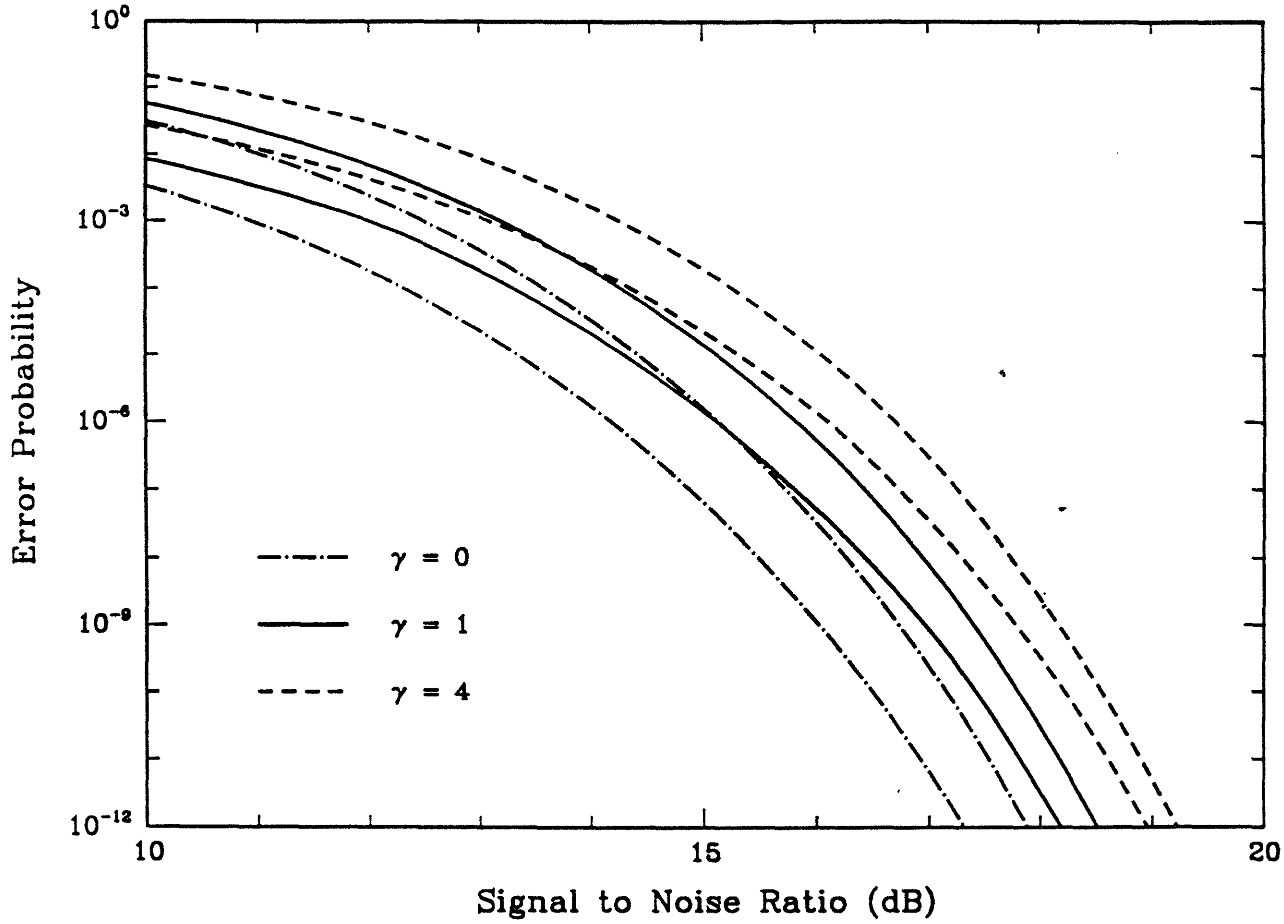

Figure 5: A comparison of Chernoff-Jensen approximation for double filter FSK (top curve for each γ value) with the performance given in [3]. The $\gamma = 0$ curves correspond to the exact error probability and the Chernoff bound for envelope detection of phase noise free FSK.

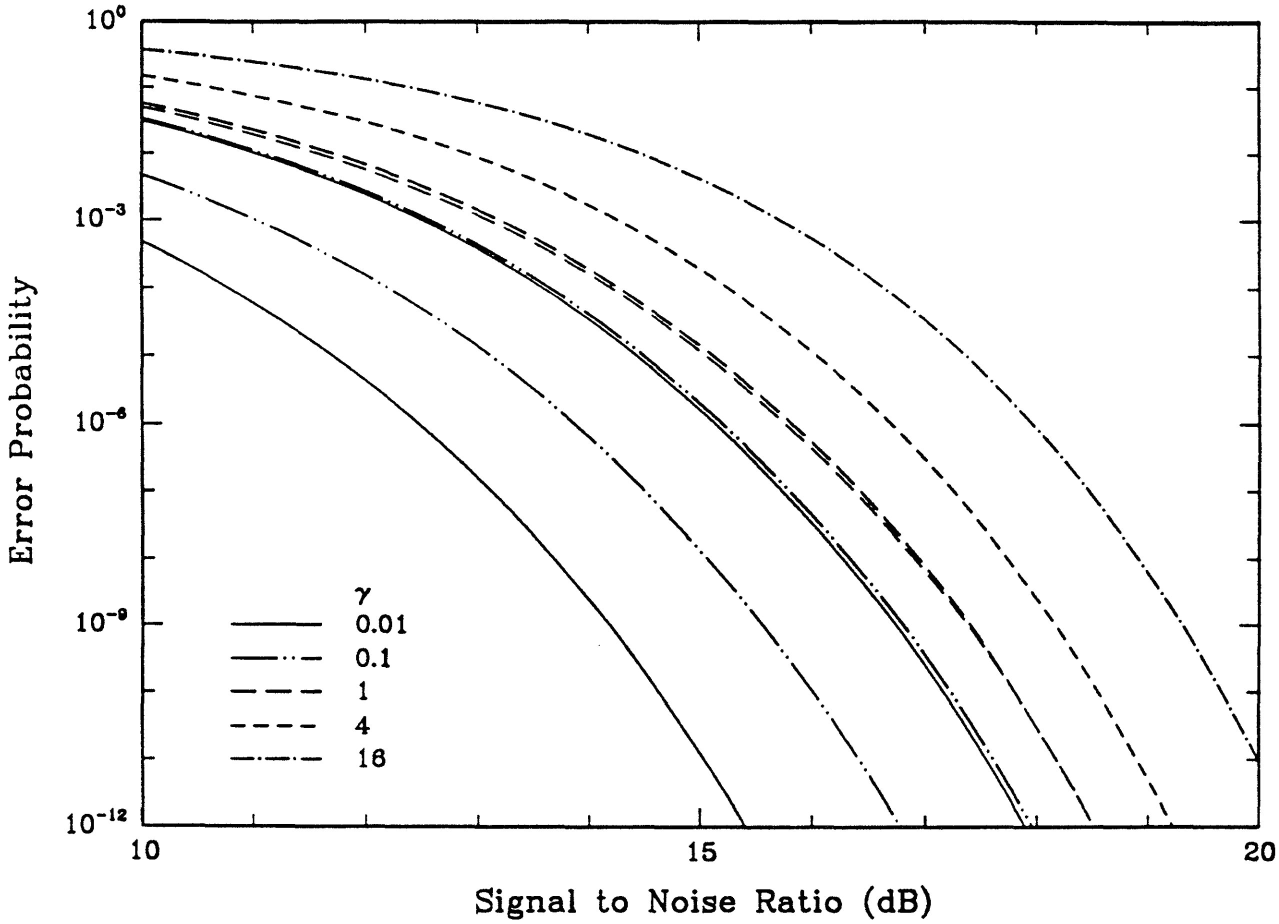

Figure 6: Error probability of transmitted reference system with a noncentered reference filter as well as that of double filter FSK. For each γ value, the former has a lower error probability. With $\gamma = 4, 16$, the performance curves are identical.

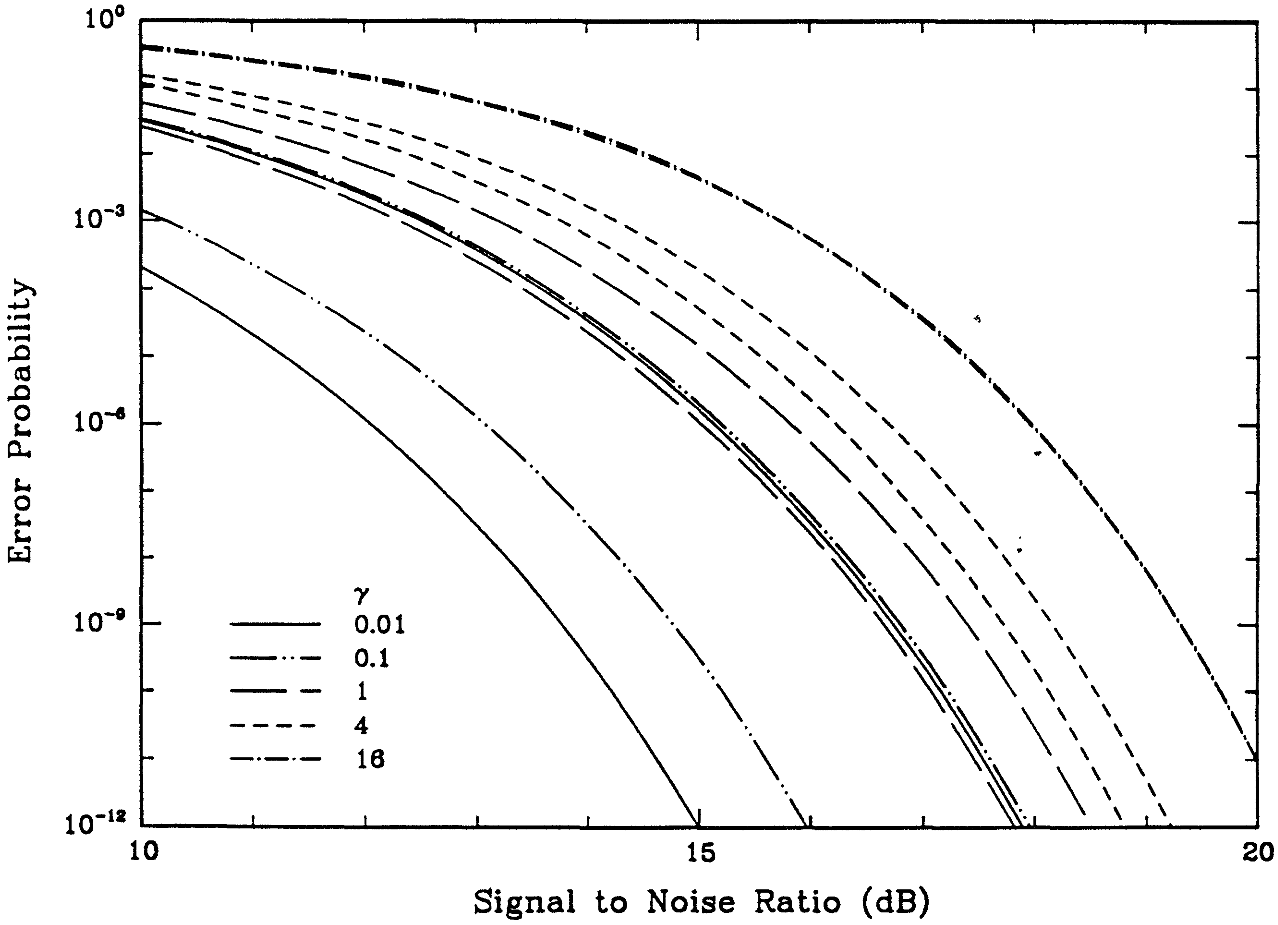

Figure 7: Equivalent of Figure 5 with a centered reference filter. Note the performance gain with $\gamma = 4$.

γ	Noncentered			Centered		
	K	M	ξ_1/ξ_2	K	M	ξ_1/ξ_2
0	∞	1	∞	∞	1	∞
0.006	24	1	10.7	58	1	19.8
0.01	18	1	8.5	41	1	15.8
0.05	7	1	4.1	15	1	7.1
0.1	4	1	2.8	9	1	4.8
0.5	2	1	1.6	7	3	2.0
1	2	2	1.1	5	4	1.5
2	1	2	1	3	4	1.2
4	1	4	1	3	7	1.07
16	1	10	1	1	10	1

Figure 8: Optimal values of K and M for various values of γ at a bit error probability of 10^{-9}.

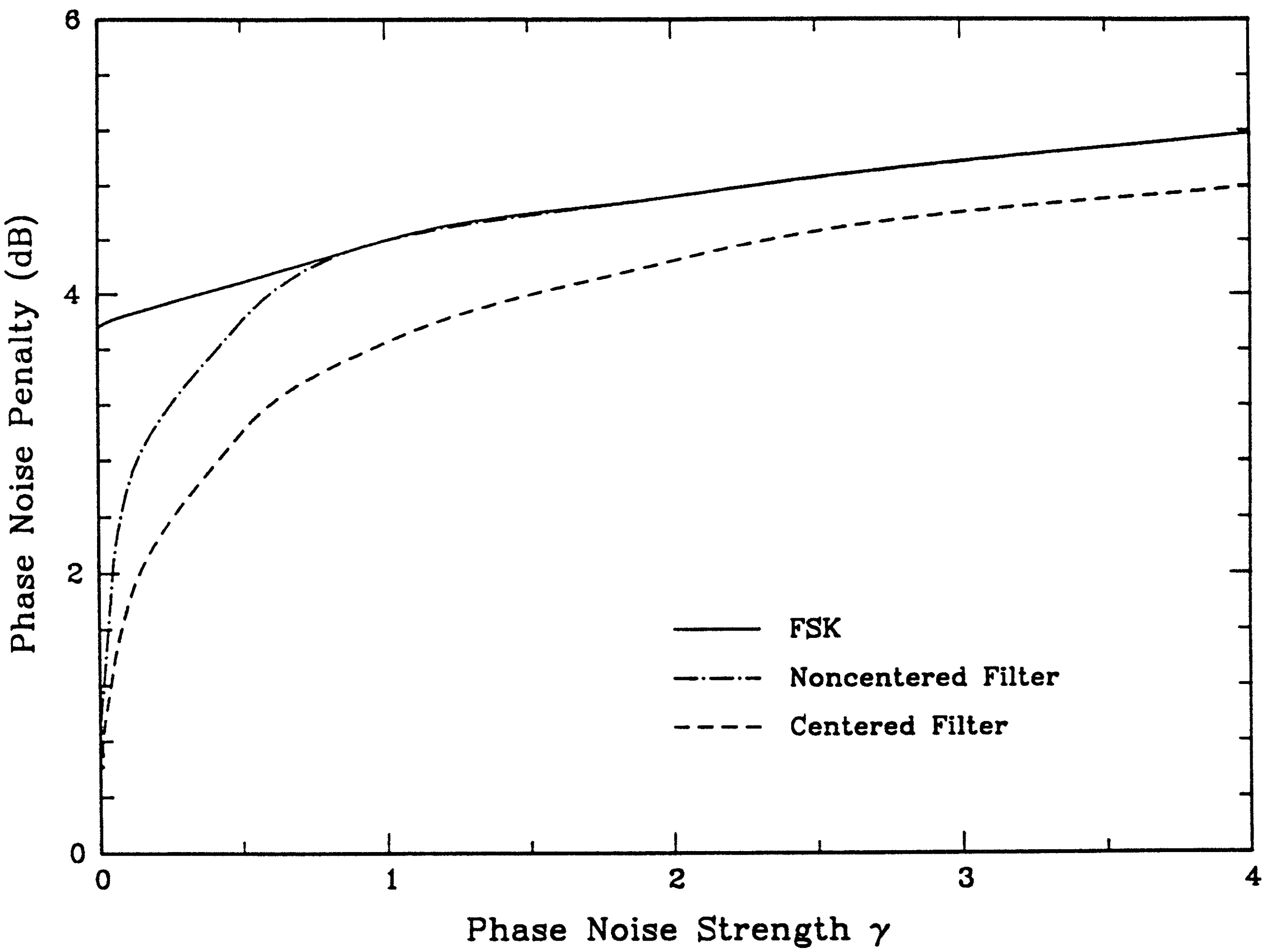

Figure 9: Phase noise induced SNR penalty for noncentered and centered reference transmission systems as well as double filter FSK with respect to ideal DPSK at $P_e = 10^{-9}$.