Presented at National Telesystems Conference, Washington, D.C., May 1992

LIDS-P-2112

Reference Transmission and Receiver Optimization for Coherent Optical Communication Systems [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

March 1, 1992

Abstract

In this paper, we consider the performance of an optical communication system where a reference signal is transmitted along with the information carrying signal to alleviate the effect of phase noise. When the power of the two signals and the receiver bandwidths are optimized, the resulting performance is significantly better than traditional amplitude and frequency modulation schemes. This makes reference transmission with a joint optimization of signal and receiver parameters a promising scheme for coherent optical communications.

[1] This research was supported by NSF under grant NSF/8802991-NCR, by DARPA under grant F19628-90-C-0002 and by ARO under contract DAAL03-86-K-0171.

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 scheme with optimally adjusted power and bandwidth parameters is described in Section 5 and its performance is discussed in Section 6. The conclusions are presented in Section 7.

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 $s(t)$. 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 don't 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.

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. Assuming that Phase Shift Keying (PSK) modulation format is employed, we have the IF signals

$$\begin{aligned} r_1(t) &= A_1 \cos(2\pi f_1 t + \theta(t) + \pi m(t)) + n_1(t) \\ r_2(t) &= A_2 \cos(2\pi f_2 t + \theta(t)) + n_2(t) \end{aligned}$$

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) $\xi_i = A_i^2 T/2$ for $i = 1, 2$. Since $r_1(t)$ and $r_2(t)$ have a common phase noise process, their product contains a term at frequency $f_2 - f_1$ which is free of phase noise, and a term at frequency $f_1 + f_2$ 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. When the SNR is finite, the signals need to be filtered first to limit the additive noise power. 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 the output. The tradeoff between the phase noise and additive noise must be accurately analyzed for transmitted reference systems [6, 7, 8].

4 Performance of Wideband Receiver Structures

We consider two receiver structures in this paper: single and double filter receivers. A single filter receiver has a single stage of filtering that takes place prior to mixing, while a double filter receiver contains an additional filtering stage after the mixing. Similar receivers with wideband filters are analyzed in [6, 7]. In this section, we outline the basic features of these analyses.

In a wideband single filter receiver, the information and reference signals are first filtered around the center frequency to limit the additive noise power. (We assume $f_1 = f_2$ without loss of generality, the two signals can be brought to the same frequency, $f_c = (f_1 + f_2)/2$, by multiplying both with $\cos(\pi(f_2 - f_1)t)$.) The bandpass filter outputs are mixed and sampled at the end of the bit period, the sampled value is compared to 0 to yield 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 error probability can be found using the standard results of [9] as

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

where

$$\begin{aligned} a &= \frac{1}{\sqrt{2}}\left(\sqrt{\frac{\xi_1}{BT}} + \sqrt{\frac{\xi_2}{WT}}\right) \\ b &= \frac{1}{\sqrt{2}}\left(\sqrt{\frac{\xi_1}{BT}} - \sqrt{\frac{\xi_2}{WT}}\right) \end{aligned}$$

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

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

For identical filters considered in [6], it is easy to show that the optimal power distribution is symmetric, $\xi_1 = \xi_2 = \xi/2$. Then (1) simplifies to result in the error probability

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

Since the reference signal $r_2(t)$ is unmodulated, it occupies a smaller bandwidth than the information signal $r_1(t)$ [7]. Hence $r_2(t)$ can be filtered with a narrower passband reducing the additive noise power at the output. For wideband filters and high SNR in both channels the optimal power ratio can be approximated as $\xi_1/\xi = B/(B+W)$ with the resulting error probability

$$P_e = \frac{1}{2} e^{-\xi/(B+W)T} . \tag{3}$$

A comparison of (3) and (2) shows that nonidentical filters have 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 $W/B \ll 1$.

In the analyses outlined above, the effect of phase noise on the performance is hidden in the selection of the filter bandwidths that will pass the phase noisy sinusoids unaltered. A heuristic relation introduced in [10] and used in [6, 7] takes $B = \frac{1}{T} + k\beta$ and $W = k\beta$, where k is a constant, such as 8. The motivation behind this relation is 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.

A wideband double filter receiver is also considered in [6]. The first stage of filtering is the same as the single filter receiver, the product of the filter outputs is integrated over the bit duration prior to sampling. The product contains the baseband PSK modulation, double frequency components, signal cross noise and noise cross noise terms. Assuming that the front end filters are identical with wide passband B, the two signal cross noise terms can be taken to be 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 spectrum over $|f| < B$ with peak $N_0^2 B$. We assume that this term is also Gaussian and replace its spectrum with a flat one with the same average power.

Then the integrator output will have antipodal signal levels corrupted by a Gaussian random variable. The resulting error probability is found as

$$P_e = Q\left(\sqrt{\frac{2\xi_1\xi_2}{\xi + BT}}\right)$$

where $Q(\cdot)$ is the complementary distribution function of a unit Gaussian random variable. The optimal power distribution still satisfies $\xi_1 = \xi_2 = \xi/2$ with the resulting error probability

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

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.

An important observation is that the performance of a transmitted reference system with even power distribution is identical to that of binary orthogonal signaling 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 $(y_1(t) - y_2(t))^2$, 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, e.g. binary FSK with large frequency deviation. Therefore reference transmission without power and filter optimization does not result in a performance improvement over a much simpler scheme. It is with the optimization, as will be shown, that reference transmission becomes a promising alternative.

The analyses outlined above have two deficiencies. First, they don't adequately address the tradeoff between phase noise and additive noise by assuming wideband filters, and in the case of [6] identical filter bandwidths. The nondistortion of phase noisy sinusoids afforded by wideband filters is not necessary for good performance as analyses of single carrier schemes show [2, 3, 4]. The presence of a heuristic bandwidth parameter, the setting of which arbitrarily determines the performance, is also undesired. Secondly, the most promising reference transmission scheme, i.e. that with optimal power distribution and a double filter receiver with optimal filter bandwidths, has not been considered in previous work. In the rest of this paper, we consider the performance of such a system.

5 Optimal Reference Transmission

In this section, we describe a double filter receiver model whose filter parameters will be optimized in conjunction with the distribution of the total power between the information and the reference signals. These signals are first filtered by two bandpass filters. Since the outputs of these filters will be correlated, we may consider the lowpass equivalents of the filter outputs without affecting the decision variable. In particular, we model the lowpass equivalent of a bandpass filter as in-phase and quadrature demodulators followed by finite-time integrators. 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 $K, M \geq 1$ are to be found optimally. 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. The first form has causal filters: 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 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 $(t - (T_1 + T_2)/2, t + (T_2 - T_1)/2)$. This constitutes the second timing form. The integration window of the information filter could not be changed, since intersymbol interference would occur otherwise due to windows crossing bit boundaries at the sampling times.

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 four signals: in-phase and quadrature components of both information and reference signals. These signals are then mixed to obtain $x_I(t)y_I(t)+x_Q(t)y_Q(t)$ where the label x refers to information and y refers to reference, and the subscripts I/Q refers to in-phase/quadrature. 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. Finally the lowpass filter at the end of the bit duration is compared to 0 to reach the decision.

It is convenient to express the real signals in a complex notation as

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

for the noncentered filter, where $d = \pm 1$ depending on the data bit. The centered filter will have the integrals 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 even when the entries of the vectors are statistically independent is an involved task [11]. 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. The bound is of the form

$$P_e(\theta(t)) \le \exp\left[f(M, K, u, \xi_1, \xi_2, \{\theta(t)\})\right]$$

where $u \ge 0$ is the parameter of the Chernoff bound. $f(\cdot)$ is not given here explicitly for brevity, it involves determinants and eigenvalues of matrices that are defined in terms of the arguments. This function also depends on the filter form, centered vs. noncentered, as it affects the correlation structure. (The details of the mathematical development will be provided in another paper.) The dependence of this function on the phase noise process does not allow the exact removal of the conditioning. However one can find the average of $f(\cdot)$ with respect to $\theta(t)$ in closed-form. Thus, by interchanging the exponentiation and the expectation operators, one obtains the approximation

$$P_e \simeq \exp\left[\bar{f}(M, K, u, \xi_1, \xi_2, \{\theta(t)\})\right] \tag{4}$$

where $\bar{f}(\cdot)$ denotes the expectation of $f(\cdot)$ with respect to the phase noise process. The right hand side of (4) is not an upper bound on error probability since $\exp(\cdot)$ is a convex $\cup$ function. We call this the *Chernoff-Jensen approximation* to reflect both the method by which the conditional bound is obtained and the method by which the conditioning is removed.

The analysis outlined here is applicable to a broad range of problems. Envelope detection of FSK signals corresponds to setting $K = 1$ and $\xi_1 = \xi_2 = \xi/2$. Transmitted reference with single filter receivers results when $M = 1$. Performance of Differential Phase Shift Keying (DPSK) can also be obtained. These cases will be considered in detail in an expanded version of this paper.

6 Results and Discussion

The Chernoff-Jensen approximation to the bit error probability has been 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 1 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 2

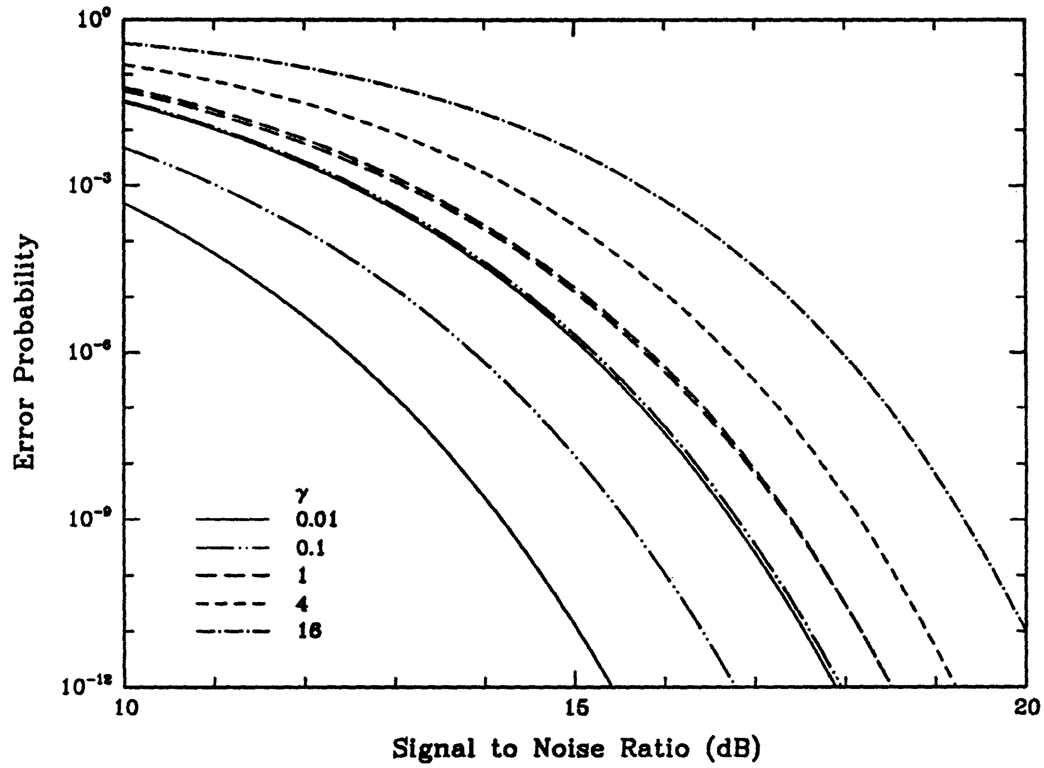

Figure 1: Error probability of transmitted reference system with a noncentered reference filter.

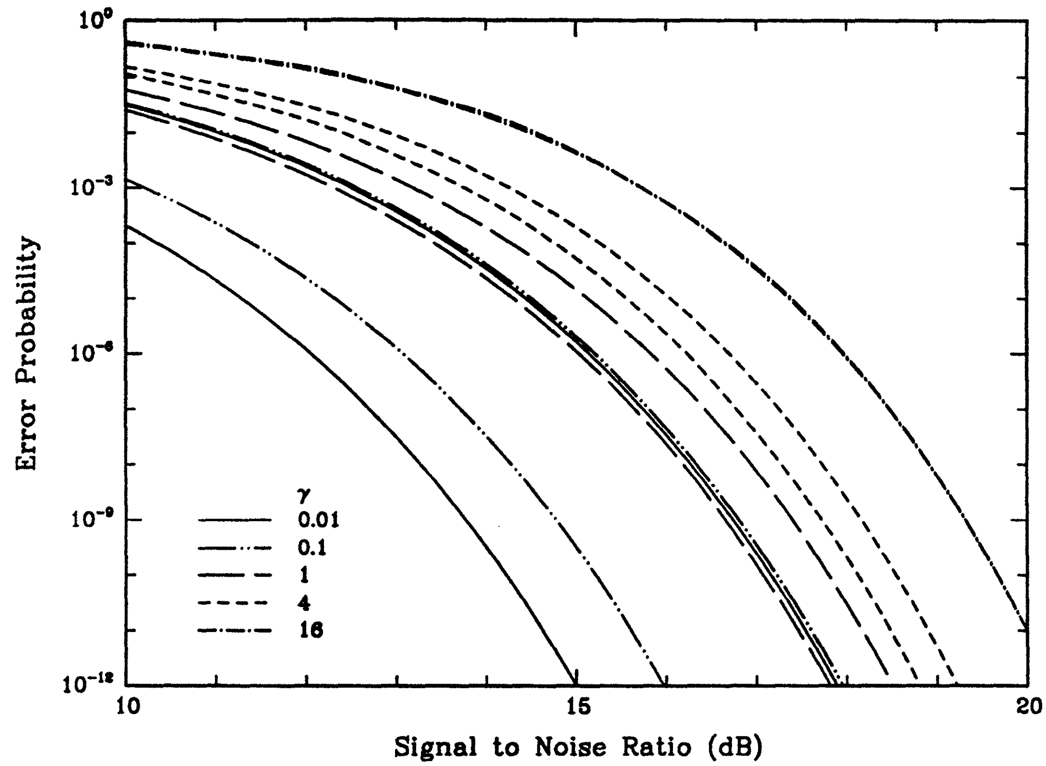

Figure 2: Error probability of transmitted reference system with a centered reference filter.

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 $\gamma = 4$.

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. More features of the performance will be described elsewhere.

The phase noise induced SNR penalties of the two transmitted reference systems with respect to ideal DPSK are shown in Figure 3. The penalty for double filter FSK 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 $\gamma = 0.9$, the system with centered filter has an improved performance up to $\gamma = 8$. For a typical

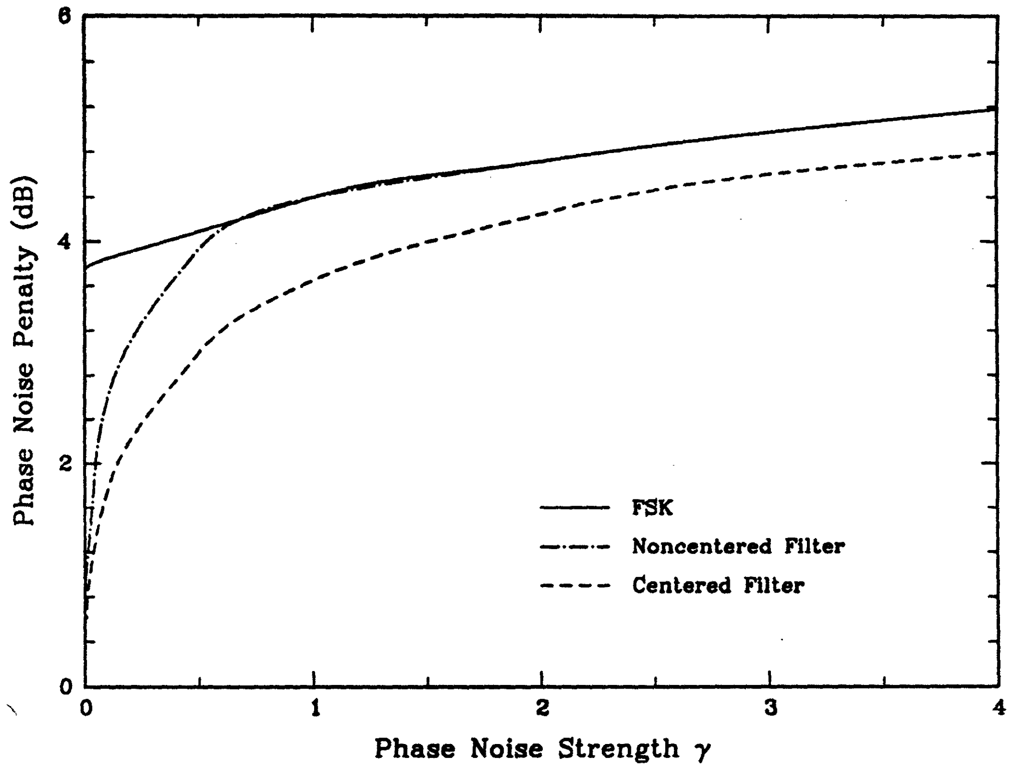

Figure 3: Phase noise penalty for noncentered and centered transmitted reference systems and double filter FSK with respect to ideal DPSK at $P_e = 10^{-9}$.

value of $\gamma = 1$, the centered filter has a gain of 1 dB over FSK.

7 Conclusion

The transmission of a reference at an optical 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. Such a reference transmission scheme is promising for attaining robustness against phase noise for typical linewidth and bit rate values. 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.

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

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