JOURNAL OF LIGHTWAVE TECHNOLOGY, VOL. 10, NO. 12, DECEMBER 1992

Carrier Synchronization for Homodyne and Heterodyne Detection of Optical Quadriphase-Shift Keying

John R. Barry and Joseph M. Kahn

***Abstract*—We investigate, through analysis and simulation, the performance of four carrier-synchronization techniques suitable for both homodyne and heterodyne detection of optical quadriphase-shift keying: the discrete-time decision-directed loop, the analog decision-directed loop, the Costas quadriphase loop, and the fourth-power phase-locked loop. Accounting for shot noise, laser phase noise, and feedback delay, we optimize the loop natural frequency and specify laser-linewidth requirements. The performance discrepancy between the best and worst of these loops is found to be small; accounting for inherent loop delays only, the linewidth requirements range from $\Delta\nu T < 2.5 \times 10^{-5}$ to $\Delta\nu T < 5.2 \times 10^{-5}$, where $\Delta\nu$ is the beat laser linewidth and T is the baud interval. Hence other considerations, such as ease of implementation, will govern the design choice for most practical systems. For the case when propagation delays in the feedback loop are significant, we present a simple and accurate method for estimating the laser-linewidth requirement and corresponding optimal natural frequency.**

I. INTRODUCTION

TO date, both coherent and direct-detection optical communication systems have used binary signaling almost exclusively, primarily because the immense bandwidth of optical fiber has made the high spectral efficiency of quadrature-amplitude modulation (QAM) techniques unnecessary. There are two trends in fiber-optic communication, however, that may lead to the use of QAM techniques such as quadriphase-shift keying (QPSK, or 4-QAM) in future systems. The first trend is the push toward higher speeds, which is driven by the philosophy that, when transmitting a high volume of traffic, a single high-speed link is generally cheaper than several slower-speed links in parallel. As future systems aim for bit rates of 10 Gb/s and higher, the limited electrical bandwidth of the receiver becomes a serious constraint, and QAM becomes a viable alternative. A receiver with an electrical bandwidth of 10 GHz, for example, would be hard pressed to achieve a 20-Gb/s bit rate using binary signaling, but it in principle could achieve a 40-Gb/s bit rate using homodyne QPSK. A second trend in multichannel fiber-optic communications is toward densely packed frequency-division multiplexed (FDM) systems; the high bandwidth efficiency of QAM would allow more FDM channels to be packed into the limited tuning range of tunable receivers. The price paid for this bandwidth efficiency is receiver complexity; QAM requires coherent detection and two electrical processing branches (in-phase and quadrature).

Optical QPSK is the most attractive of all QAM schemes because it offers power efficiency as well as bandwidth efficiency; both heterodyne and homodyne QPSK require only 18 photons/bit in the quantum limit, equaling the sensitivity of synchronous heterodyne binary PSK, the most sensitive of the binary heterodyne schemes.

To perform synchronous detection of QPSK, the receiver must match the phase of the local oscillator (LO) laser to that of the received optical signal using a phase-locked loop (PLL). This PLL can be either optical or electrical; they have identical steady-state performance, and the analysis we present is applicable to both. The optical PLL is preferable, however, due primarily to its wider lock-in range.

In this paper we examine the performance of four carrier-synchronization techniques suitable for both heterodyne and homodyne detection of optical QPSK. The performance of the Costas loop has been examined elsewhere [1], [2]. Here, we analyze the performance of four loops: the discrete-time decision-directed loop, the analog decision-directed loop, the Costas loop, and the fourth-power loop, and compare their relative merits. We follow the approach of Spilker [3], as have others analyzing the performance of carrier synchronization for coherent detection of binary signaling [4], [5]. The goal of our analysis, which accounts for the effects of shot noise, laser phase noise, and feedback delay [6], [7], is to optimize the loop filter and specify laser-linewidth requirements.

In the next section we establish the equivalence between homodyne and heterodyne QPSK [8], [2]. In Section III we describe the loops under consideration. For each loop we present three models: implementation models, which represent the physical system implementation; complex models, which are advantageous for simulation; and linear models, which are used for analysis. Our treatment of the decision-directed and Costas loops is unusual in that we view them both as special cases of a more general quadriphase PLL. This approach yields valuable insight into the relationship between these different loop types. In Section IV we present a unified linear analysis applicable to all of the loops in question.

Manuscript received February 24, 1992; revised July 1, 1992.

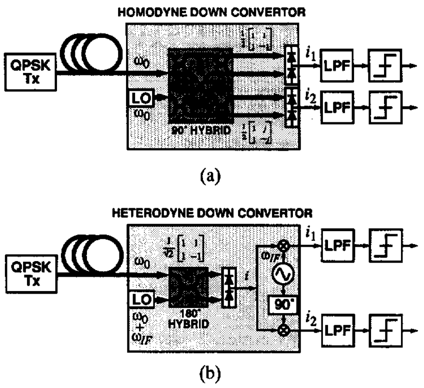
This work was supported in part by NSF, the California State MICRO program, and Sony Corp.

J. R. Barry was with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720. He is now with the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

J. M. Kahn is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

IEEE Log Number 9204531.

0733–8724/92$03.00 © 1992 IEEE



(a)

(b)

Fig. 1. Canonical receiver structure for (a) homodyne and (b) heterodyne QPSK. Carrier synchronization is not shown.

We present procedures for optimizing the loop filter and specifying laser-linewidth requirements when the feedback delay is not negligible. Finally, in Section V, we present numerical simulation results which confirm the accuracy of our theoretical results.

II. Equivalence of Homodyne and Heterodyne QPSK

The receiver structures for homodyne and heterodyne QPSK are compared in Fig. 1. Recall that QPSK can be viewed as two independent binary PSK signals in quadrature. The primary difference between homodyne and heterodyne QPSK detection is the domain in which these two channels are separated. In homodyne detection, the channels are separated in the optical domain using a 90° optical hybrid, which nullifies the familiar 3-dB gain of homodyne over heterodyne [8], [9]. (In fact, there is no sensitivity advantage of homodyne over heterodyne for any QAM technique.) This sensitivity reduction is unavoidable, however, for no single-detector receiver can efficiently recover both the in-phase and quadrature data streams. In heterodyne detection of optical QPSK, the received signal and LO lightwave are coupled using a 180° hybrid, resulting in an intermediate frequency (IF) electrical signal. Channel separation is then performed in the electrical domain.

The heterodyne receiver cannot use the 90° optical-hybrid front end without unnecessarily introducing a 3-dB penalty, and the homodyne receiver cannot use the 180° optical-hybrid front end without destroying information. In this sense, the receiver structures shown in Fig. 1 are canonical.

If the output of the fixed RF oscillator in Fig. 1(b) is $\cos(\omega_{\text{IF}}t)$, where ω_{IF} is the difference between the signal frequency and LO frequency, then the baseband currents $i_1(t)$ and $i_2(t)$ of both the homodyne and heterodyne down convertors of Fig. 1 are given by

$$\begin{aligned} i_1(t) &= A\cos[\theta(t) + \varphi(t) - \psi(t)] + n_1(t), \\ i_2(t) &= A\sin[\theta(t) + \varphi(t) - \psi(t)] + n_2(t) \end{aligned} \qquad (1)$$

where $A \triangleq R\sqrt{P_S P_{\text{LO}}}$ is the baseband signal amplitude, R is the photodiode responsivity, P_S is the power of the received optical signal, P_{LO} is the power of the LO laser, $\theta(t) \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$ is the QPSK modulation and is constant over each symbol interval of duration T, $\varphi(t)$ is the beat laser-phase noise (the sum of the phase noises in the two lasers), $\psi(t)$ is the control phase of the LO laser, and $n_1(t)$ and $n_2(t)$ are shot noises modeled as independent and identically distributed Gaussian processes with two-sided power-spectral densities (PSD's) $qRP_{\text{LO}}/2$, where q is the charge of an electron. We implicity assume that the outputs of the RF mixers in Fig. 1(b) are bandlimited so that the double-frequency components (at $2\omega_{\text{IF}}$) are rejected. With this assumption, the ideal heterodyne and homodyne down convertors of Fig. 1(a) and (b) are completely interchangeable; they take identical inputs and produce identical outputs. As shown in the figure, the signals i_1 and i_2 are filtered, sampled, and quantized to yield the in-phase and quadrature data streams.

Because the homodyne and heterodyne down convertors of Fig. 1 produce identical outputs, we can draw three important conclusions: first, homodyne and heterodyne QPSK have the same sensitivity [8]; second, any PLL suitable for homodyne detection of QPSK can also be applied to heterodyne detection of QPSK; and third, the corresponding PLL design procedure and laser-linewidth requirements are identical for both homodyne and heterodyne QPSK [2].

Only three experimental optical QPSK systems have been reported to date, and all three have used homodyne detection [10]–[12]. There are numerous reasons to consider heterodyne detection of QPSK, however, despite the increased bandwidth requirement:

- Heterodyne QPSK has the same sensitivity as homodyne QPSK.
- The 180° optical hybrid is easier to implement than the 90° optical hybrid.
- The heterodyne receiver requires only one balanced detector.
- The heterodyne receiver, in requiring half as many balanced detectors, requires half as much LO power to achieve shot-noise-limited operation.
- Chromatic dispersion equalization is easier for the heterodyne receiver. This is because equalization in the homodyne receiver requires two baseband equalizers plus cross-processing between the in-phase and quadrature receiver branches, whereas equalization in the heterodyne receiver requires only a single, passband equalizer.

We therefore consider both heterodyne and homodyne detection. This task is simplified greatly by the equivalence of the homodyne and heterodyne down convertors of Fig. 1; most of the loops we consider take i_1 and i_2 as inputs and are thus applicable to both homodyne and heterodyne receivers. The one exception is the fourth-power PLL, which for heterodyne detection has a simple realization based directly on the IF photodetector output.

III. Loop Alternatives

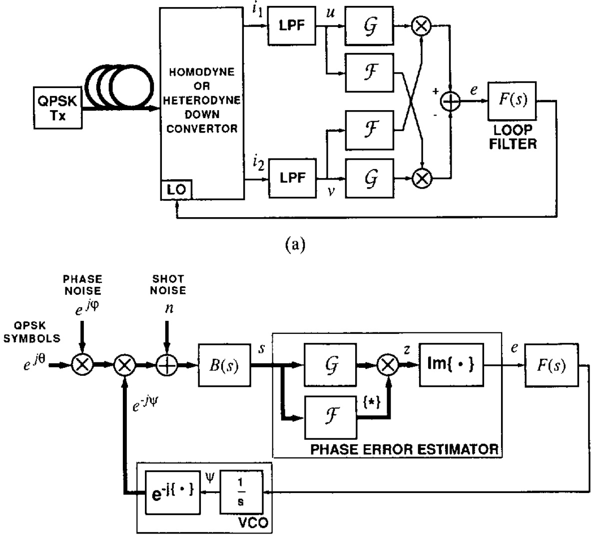
In this section we describe four carrier synchronization techniques suitable for synchronous detection of optical QPSK: the discrete-time decision-directed PLL and its derivatives (the analog decision-directed PLL and the Costas loop), and the fourth-power PLL.

TABLE I
TRANSFER FUNCTIONS $\mathcal{F}$ AND $\mathcal{G}$.

Loop Type	$\mathcal{F}\{x(t)\}$	$\mathcal{G}\{x(t)\}$
Discrete-Time Decision-Directed	Decision: $[x(kT)]$	Sample & Hold: $x(kT)$
Analog Decision-Directed	Decision: $[x(kT)]$	Delay: $x(t-T/2)$
Costas	Hard Limiter: $[x(t)]$	Identity: $x(t)$
4-th Power	Conjugate:† $x^*(t)$	Cube:† $x^3(t)$

†Equation (6) does not apply.

Note: shown are the outputs of the $\mathcal{F}$ and $\mathcal{G}$ blocks during the kth baud interval, $t \in [kT, (k+1)T]$, subject to an input $x(t)$.



(a)

(b)

Fig. 2. The decision-directed and Costas loops: (a) implementation; (b) complex model.

A. The Discrete-Time Decision-Directed Loop and its Derivatives

We first consider the discrete-time decision-directed loop and its closely related derivatives, the analog decision-directed loop and the Costas loop. Each can be applied to both homodyne and heterodyne QPSK. We describe first their implementation, then their complex model.

1) Implementation: The discrete-time decision-directed loop and its derivatives have an identical structure, as shown in Fig. 2(a). The outputs $i_1(t)$ and $i_2(t)$ of the homodyne or heterodyne down convertor are passed through identical low-pass filters with rectangular impulse response $(1/A)b(t)$, where:

$$b(t) \triangleq \begin{cases} 1/T & \text{for } t \in (0, \mathrm{T}] \\ 0 & \text{for } t \notin (0, T]. \end{cases} \tag{2}$$

The low-pass filters are normalized by A to simplify later notation. The low-pass-filter outputs $u(t)$ and $v(t)$ are then passed through blocks labeled $\mathcal{F}$ and $\mathcal{G}$. The final error estimate $e(t)$ is found by a cross multiplication and a subtraction, as shown schematically in Fig. 2(a). As we will see, $e(t)$ closely approximates the true phase error $\varepsilon(t) \triangleq \varphi(t) - \psi(t)$.

The contents of the $\mathcal{F}$ and $\mathcal{G}$ blocks are determined by the loop type per Table I. The square brackets in the table denote a hard-limiting (for a time-varying input) or a quantizing (for a discrete-time input) operation:

$$[x] \triangleq \begin{cases} -1/\sqrt{2} & \text{for } x < 0 \\ 1/\sqrt{2} & \text{for } x \geq 0. \end{cases} \tag{3}$$

From Fig. 2(a), the phase error estimate $e(t)$ for $t \in [kT, (k+1)T]$ is given by:

$$\begin{aligned} e(t) &= \mathcal{G}\{u(t)\}\mathcal{F}\{v(t)\} - \mathcal{F}\{u(t)\}\mathcal{G}\{v(t)\} & (4) \\ &= \operatorname{Im}\{\mathcal{G}\{s(t)\}\mathcal{F}\{s(t)\}^*\} & (5) \end{aligned}$$

where $s(t) \triangleq u(t) + jv(t)$. We extend the definitions of $\mathcal{F}$ and $\mathcal{G}$ to complex-valued inputs by assuming independent action on each of the real and imaginary parts, so that:

$$\mathcal{F}\{u + jv\} \triangleq \mathcal{F}\{u\} + j\mathcal{F}\{v\} \tag{6}$$

and similarly for $\mathcal{G}$. This is a natural extension, based on the following observations regarding the functions described in the first three rows of Table I:

- Independently delaying the real and imaginary parts of a complex signal amounts to delaying the complex signal itself.
- Independently sampling and holding the real and imaginary parts of a complex signal amounts to sampling and holding the complex signal itself.
- Independently quantizing (limiting) the real and imaginary parts of a complex number (signal) amounts to a two-dimensional quantization (limiting) of the complex number (signal) itself.

This exhausts the functions listed for the first three loops in Table I, because a decision device is a concatenation of a sample-and-hold device with a quantizer, and the identity function is a zero-delay device.

In the next subsection we introduce a complex model for the decision-directed and Costas loops which not only facilitates analysis and simulation, but also clarifies why (4) provides a good estimate of the true phase error $\varepsilon = \varphi - \psi$.

2) Complex Model: The relationship in (5) leads to a mathematically equivalent complex model for the decision-directed loops of Fig. 2(a), as shown in Fig. 2(b). We assume the LO laser acts as an ideal current-controlled oscillator, so that the laser control phase ψ is proportional to the integral of the loop filter output. The local laser is thus represented in Fig. 2(b) by the block labeled VCO, which consists of an integrator followed by a complex exponentiator. Without loss of generality, we can combine the laser phase noise of the

LO laser with that of the received lightwave, with the result being that the phase noise φ of the input signal in Fig. 2(b) is actually the sum of the phase noises in both lasers, and the VCO in Fig. 2(b) has no phase noise. The noise $n(t)$ in Fig. 2(b) represents the additive shot noise and is modeled as a complex white Gaussian noise process with a PSD of T/M, where T is the symbol interval and $M \triangleq RP_S T/q$ is the number of detected photons per symbol [13].

As described in Table I, the discrete-time decision-directed loop uses a decision device for $\mathcal{F}$ and a sample-and-hold for $\mathcal{G}$. With the aid of Fig. 2(b), we will now explain the intuition behind these choices. Once this is understood, the workings of the other loops listed in Table I will become clear.

Suppose the complex representation of the signal during the kth baud interval, $t \in [kT, (k+1)T]$, is given by:[1]

$$s_k = e^{j(\theta_k + \varepsilon_k)}. \tag{7}$$

Here, $\theta_k \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$ represents the QPSK modulation during the kth baud interval, and ε_k represents the phase error, which for now we assume varies slowly enough that it is approximately constant during one baud interval. Note that, for the purposes of this illustrative example, we are neglecting shot noise. In Fig. 3 we illustrate the QPSK constellation along with a typical s_k, which is marked with an X.

The problem faced by the carrier synchronizer is to estimate ε_k based on s_k. Observe from Fig. 3 that multiplying s_k by a complex-conjugated and quantized version of itself, $[s_k]^*$, yields $e^{j\varepsilon_k}$, provided that $|\varepsilon_k| < \pi/4$. This leads to the following estimate of the phase error:

$$e(t) = \text{Im}\{s_k[s_k]^*\}, \tag{8}$$

which in this example reduces to:

$$e(t) = \text{Im}\{e^{j\varepsilon_k}\} = \sin\varepsilon_k \approx \varepsilon_k \tag{9}$$

the last equality being valid for small ε_k.

Of course, in practice there is additive shot noise, and the phase error is not constant during each baud interval. To counter these complications, a receiver can pass the signal through a matched filter and then sample-and-hold the output. The output of the sample-and-hold device during the kth baud interval is then:

$$s_k = e^{j(\theta_{k-1} + \varepsilon_{k-1})} + n_{k-1} \tag{10}$$

which is similar to (7), except for an effective delay of one baud interval, and here $e^{j\varepsilon_k}$ is the time average of the phase-error phasor over the kth baud interval:

$$e^{j\varepsilon_k} \triangleq \frac{1}{T}\int_{kT}^{(k+1)T} e^{j\varepsilon(t)}\,dt \tag{11}$$

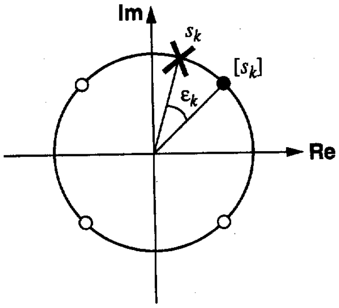


Fig. 3. The QPSK constellation with a typical slicer input, s_k, and corresponding slicer output, $[s_k]$.

where $\varepsilon(t) = \varphi(t) - \psi(t)$ is the true phase error, and n_k is the shot noise averaged over the k-th baud interval:

$$n_k \triangleq \frac{1}{T}\int_{kT}^{(k+1)T} n(t)\,dt. \tag{12}$$

The noise n_k has the effect of perturbing s_k off of the unit circle of Fig. 3. Nevertheless, (8) still provides an excellent estimate of the phase error. And this is precisely the estimate generated by the discrete-time decision-directed loop of Fig. 2 per (4). To wit, the s_k factor in (8) is generated by $\mathcal{G}$, which from Table I is a sample-and-hold device, and the $[s_k]$ factor in (8) is generated by $\mathcal{F}$, which from Table I is a decision device (sample-and-hold plus quantizer). In other words, the discrete-time decision-directed loop estimates the phase error by comparing the phase of the complex-valued slicer input (s_k) with that of the complex-valued slicer output ($[s_k]$).

a) Analog Decision-Directed Loop: What we call the analog decision-directed loop is frequently referred to as simply the decision-directed PLL [3], [14]. According to Table I, the analog decision-directed loop replaces the sample-and-hold device of the discrete-time decision-directed loop by a half-baud delay. This substitution does not have significant impact on loop performance, because the low-frequency component of the sample-and-hold input is approximately equal to the low-frequency component of the sample-and-hold output, and also the sample-and-hold device has an effective delay of $T/2$ (the front-end filter of (2) has an additional group delay of $T/2$, so they combine to make a total delay of T). Because the high-frequency components are rejected by the low-pass nature of the loop transfer function, the performance penalty due to the substitution of the half-baud delay for the sample-and-hold device is minimal, and thus the analog decision-directed loop is only slightly inferior to the discrete-time decision directed loop. As we see later, this conclusion is borne out by our simulations.

b) Costas Loop: The Costas loop is still simpler than the analog decision-directed loop: it replaces the decision device of the analog decision-directed loop by a hard limiter, and the half-baud delay of the analog decision-directed loop by an identity function. The low-frequency component of the hard-limiter output will be equal to the correct decision. Of course, the high-frequency component will be different, but it should be rejected by the low-pass nature of the loop. Since

---

[1] In this paper we will consistently denote the piecewise-constant continuous-time function $T\sum_k x_k b(t-T)$, where $b(t)$ is the rectangular pulse defined in (2), by the values $\{x_k\}$ it takes on during each interval.

TABLE II
INHERENT DELAYS DUE TO FILTERING AND OTHER PROCESSING

Loop Type	Low-Pass Filter	Other Processing	Total Delay
Discrete-Time Decision-Directed	$T/2$	$T/2$	T
Analog Decision-Directed	$T/2$	$T/2$	T
Costas	$T/2$	0	$T/2$
Fourth-Power	τ_4	0	τ_4

there is no intrinsic delay in the hard limiter, the additional half-baud delay of the analog decision-directed loop can be eliminated. The Costas loop thus has an inherent delay of only one half-baud interval, as compared to the full-baud delays of the decision-directed loops; the inherent delays of each of the loops under consideration are listed in Table II. The Costas loop, being only an approximation to the analog decision-directed loop, performs worse than the analog and discrete-time decision-directed loops.

3) *Linear Model:* In this section we develop a linear model for the decision-directed and Costas loops shown in Fig. 2. In fact, we consider only the discrete-time decision-directed PLL. As discussed above, however, the analog decision-directed loop and the Costas loop are both approximations of the discrete-time PLL. To the degree that these approximations are accurate, then, the analysis we present here is applicable to all three loops.

Consider the complex model for the decision-directed loops shown in Fig. 2(b). Assume that the front-end filter has unity dc gain, and that it is matched to the rectangular pulse shape of the transmitter, so that $b(t)$ is given by (2). The output $s(t)$ of the front-end filter is then given by

$$s(t) = b(t) \otimes \left\{ T e^{j\varepsilon(t)} \sum_m e^{j\theta_m} b(t - mT) + n(t) \right\} \qquad (13)$$

where the symbol $\otimes$ denotes convolution. According to Table I, $\mathcal{G}$ is a sample-and-hold device, so the output of the $\mathcal{G}$ block in Fig. 2(b) during the kth baud interval $t \in [kT, (k+1)T]$ is constant, equal to $s(kT)$:

$$\mathcal{G}\{s(t)\} = s(kT) = e^{j(\theta_{k-1}+\varepsilon_{k-1})} + n_{k-1} \qquad (14)$$

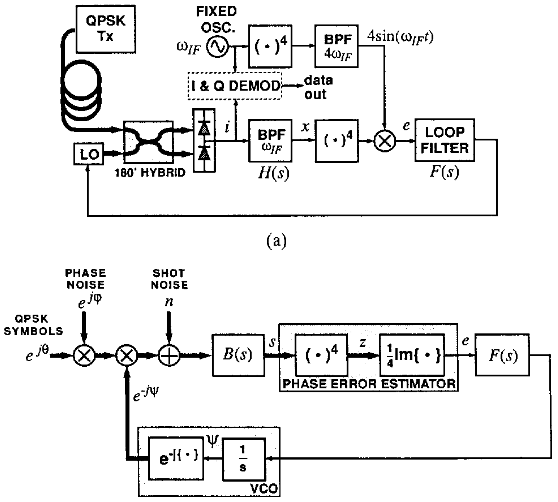
where $e^{j\varepsilon_k}$ is given by (11) and n_k is given by (12). The $\mathcal{F}$ block is a decision device, which means its output during the kth baud interval is a two-dimensionally quantized version of the sampled-and-held signal

$$\mathcal{F}\{s(t)\} = [s(kT)] = [e^{j(\theta_{k-1}+\varepsilon_{k-1})} + n_{k-1}] = e^{j\theta_{k-1}}. \qquad (15)$$

The last equality assumes that the noise perturbation n_{k-1} and the phase error ε_{k-1} are small enough that the decision device does not make an error. In a practical system, errors are rare (error rates below 10^{-9} are common), so this is a good assumption.

From (14) and (15), the error signal during the kth baud interval $t \in [kT, (k+1)T]$ becomes:

$$\begin{aligned} e(t) &= \mathrm{Im}\{\mathcal{G}\{s(t)\}\mathcal{F}\{s(t)\}^*\} && (16) \\ &= \mathrm{Im}\{e^{j\varepsilon_{k-1}} + n_{k-1}e^{-j\theta_{k-1}}\} && (17) \\ &= \sin\varepsilon_{k-1} + w(t-T) && (18) \end{aligned}$$



(a)

(b)

Fig. 4. Fourth-power optical PLL: (a) implementation for heterodyne detection; (b) complex model.

where

$$w(t) \triangleq \mathrm{Im}\{n_k e^{-j\theta_k}\} \quad \text{for} \quad t \in [kT, (k+1)T]. \qquad (19)$$

In Appendix A we show that $w(t)$ is well-approximated by a white random process with power spectrum $S_w(\omega) = \eta T/M$, where $\eta = 1/2$. If $\varepsilon(t)$ is approximately constant over each baud interval, then from (11) we have $\varepsilon_k \approx \varepsilon(t)$, and (18) reduces to:

$$e(t) \approx \varepsilon(t-T) + w(t-T) \qquad (20)$$

for small ε. This is the linear model which we use in Section IV to analyze the performance of this PLL.

B. The Fourth-Power PLL

We now consider the fourth-power PLL. We consider first the traditional passband architecture which, unlike the decision-directed loops considered in Section III-A, is specific to heterodyne detection and cannot be used in homodyne. We describe its passband system implementation, then introduce a baseband model more amenable for analysis and simulation. We also introduce a homodyne realization of the fourth-power PLL.

1) *Implementation:* In Fig. 4(a) we illustrate the structure of a fourth-power PLL for heterodyne detection of optical QPSK. The IF current $i(t)$ is given by:

$$i(t) = 2A\cos[\omega_{\mathrm{IF}}t + \theta(t) + \varphi(t) - \psi(t)] + n_{sh}(t), \qquad (21)$$

where all of the parameters were introduced in Section II except for $n_{sh}(t)$, which represents the IF shot noise and is modeled as a Gaussian noise with two-sided PSD qRP_{LO}. Hereafter we assume the modulating phase $\theta(t)$ in (21) takes on values from $\{0, \pi/2, \pi, 3\pi/2\}$.

To reject out-of-band noise, the IF signal is first passed through a band-pass filter with frequency response:

$$H(\omega) = \frac{1}{A}[B(\omega - \omega_{\mathrm{IF}}) + B(\omega + \omega_{\mathrm{IF}})] \qquad (22)$$

where $B(\omega)$ is a baseband low-pass filter with unity dc gain; ideally this band-pass filter passes the signal-with-phase-noise term undistorted, while at the same time minimizing the amount of shot noise admitted. The low-pass filter given by (2), although suitable for the decision-directed loops of Section III-A, is not necessarily appropriate in this case. To wipe off the QPSK modulation, the output $x(t)$ of the band-pass filter is raised to the fourth power, yielding a signal centered at frequency $4\omega_{\text{IF}}$. This is mixed with $4\sin(4\omega_{\text{IF}}t)$, which is produced by band-pass-filtering the fourth-power of the fixed reference signal. The output $e(t)$ of the mixer is a good estimate of the phase error $\varepsilon(t) = \varphi(t) - \psi(t)$.

At this point it is useful to represent $x(t)$ by:

$$x(t) = \text{Re}\{s(t)e^{j\omega_{\text{IF}}t}\} \tag{23}$$

where $s(t) \triangleq u(t) + jv(t)$ is the complex baseband envelope of $x(t)$ [15]. Based on the schematic diagram of Fig. 4(a), it is straightforward to show that $e(t)$ is given by:

$$e(t) = u(t)v(t)\big(u^2(t) - v^2(t)\big) = \frac{1}{4}\text{Im}\{s^4(t)\}. \tag{24}$$

We can work backward from (24) and reformulate a homodyne realization of the fourth-power loop based on the quadriphase structure of Fig. 2(a). The resulting loop, as listed in the last (shaded) column of Table I, uses two cubing devices. This loop is unlike the other quadriphase loops for two reasons: first, the identity in (6) does not hold, and second, its error estimate approximates 4ε rather than ε. Nevertheless, the resulting loop is legitimate and has performance identical to that of the passband fourth-power loop. In fact, this loop is closely related to the low-SNR maximum likelihood QPSK carrier synchronizer [14].

The second equality in (24) leads to a convenient baseband model, which we describe next.

2) Complex Model: In Fig. 4(b) we show a mathematically-equivalent complex baseband model that simplifies considerably the analysis of the fourth-power PLL. It also simplifies simulation, because baseband signals require smaller sampling rates than passband signals; the fourth-power-loop simulation results of Section V are based on this model. Again, the noise $n(t)$ in Fig. 4(b) is modeled as a complex Gaussian process with a PSD of T/M.

3) Linear Model: As we did for the decision-directed loop, we now develop a linearized model for the fourth-power loop, using an approach similar to that in [3]. This model leads to analytical results that agree closely with accurate computer simulations.

Assume that the bandwidth of $B(s)$, the front-end filter in Fig. 4(b), is large enough to pass the signal-with-phase-noise term undistorted. The input to the fourth-power block is then $s(t) = x(t) + y(t)$, where $x(t)$ is the undistorted signal-with-phase-noise:

$$x(t) = e^{j(\theta(t)+\varepsilon(t))} \tag{25}$$

and $y(t)$ is the filtered shot noise, which is a Gaussian random process with PSD

$$S_y(\omega) = \frac{T}{M}|B(\omega)|^2. \tag{26}$$

The output of the fourth-power block is then given by

$$z(t) = (x(t) + y(t))^4 \tag{27}$$

$$= e^{j4\varepsilon(t)} + m(t) \tag{28}$$

where, because $e^{j4\theta(t)} = 1$ for all t, $x^4(t) = e^{j4\varepsilon(t)}$. The additive distortion $m(t)$ in (28) represents all of the noise-cross-noise and signal-cross-noise terms. The error signal $e(t)$ is simply:

$$e(t) = \frac{1}{4}\text{Im}\{z(t)\} \approx \varepsilon(t) + w(t) \tag{29}$$

whenever $\sin\varepsilon \approx \varepsilon$, where $w(t) \triangleq \frac{1}{4}\text{Im}\{m(t)\}$. In Appendix B we show that $w(t)$ is accurately modeled as a zero-mean random process, approximately independent of $\varepsilon(t)$, with a PSD of $\eta T/M$, where $\eta \triangleq \eta(M, W)$ is specified by (59) in Appendix B and W is the normalized bandwidth of $B(s)$, defined by (50). The best justification for this approximation is the close agreement found between the analytical results and the computer simulations of Section V.

The above derivation assumed that the front-end filter had no effect on the signal. In fact, a realizable filter will have at least some group delay τ_4, in which case (29) becomes:

$$e(t) \approx \varepsilon(t - \tau_4) + w(t - \tau_4). \tag{30}$$

IV. UNIFIED LINEAR ANALYSIS

The complex models for both the decision-directed PLL and the fourth-power PLL, as illustrated in Fig. 2(b) and Fig. 4(b), have the same general form; the only difference lies in how they estimate the phase error.

A. Common Linear Model

Comparing (20) and (30), we see that the linearized models for the decision-directed loops and the fourth-power loop are nearly identical, as shown schematically in Fig. 5. A control phase $\psi(t)$ is subtracted from the input phase $\varphi(t)$, and then a white noise $w(t)$ with PSD $S_w(\omega) = \eta T/M$ is added. The result is delayed by τ, which represents the total loop delay, including inadvertent system feedback delays as well as the inherent delays listed in Table II. The output of the delay is then passed through the loop filter and integrated to generate the control phase $\psi(t)$.

The linear model shown in Fig. 5 is identical to the linear models presented elsewhere in the context of carrier synchronization loops for binary PSK [4], [6], [7], [16]. We may therefore borrow results from previously published works. Conversely, the new results we present here, such as the method in Section IV-C for estimating the laser-linewidth requirement with arbitrary loop delay, are applicable to binary PSK as well as to QPSK systems.

The decision-directed loops and the fourth-power loop differ only in the noise factor η; as shown in Appendices A and B, $\eta = 1/2$ for the decision-directed loops, while $\eta \approx 0.552$ for the fourth-power loop (with $M = 45.3$ and $W = 3$). We therefore expect the impact of shot noise on the phase error

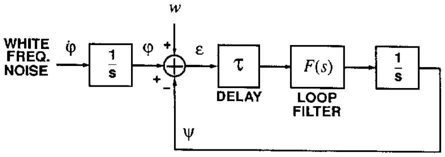


Fig. 5. Linearized PL model for all optical PLL's.

variance for the fourth-power loop to be slightly greater than for the decision-directed loops.

We model the phase noise $\varphi(t)$ as a Wiener process with parameter $2\pi\Delta\nu$, where $\Delta\nu$ is the beat (IF) laser linewidth, so that the time derivative $\dot{\varphi}(t)$ is a Gaussian random process with white PSD $2\pi\Delta\nu$ [13]. This is shown explicitly in Fig. 5, where $\varphi(t)$ is the output of an integrator whose input $\dot{\varphi}$ is Gaussian white noise with PSD $2\pi\Delta\nu$.

Because of its superior performance when compared to other filters with regard to lock-in range, frequency tracking, and stability, we assume a proportional-plus-integral loop filter [17]:

$$F(s) = 2\zeta\omega_n + \frac{\omega_n^2}{s}, \tag{31}$$

where ω_n is defined as the loop natural frequency in rad/s, and ζ is the loop damping coefficient. In this paper we assume $\zeta = 1/\sqrt{2}$, a common choice because it tends to balance the conflicting requirements of quick transient response and low steady-state variance.

The key parameter for measuring PLL performance is the total variance σ^2 of the steady-state phase error ε. A linear analysis of the model in Fig. 5, as presented in [6], [7], [16], yields the following expression:

$$\sigma^2 = \frac{\pi\Delta\nu}{2\zeta\omega_n}\Gamma_{PN}(\omega_n\tau) + \frac{(1+4\zeta^2)\omega_n T\eta}{4\zeta M}\Gamma_{SN}(\omega_n\tau) \tag{32}$$

where the "bandwidth expansion factors" $\Gamma_{PN}(\omega_n\tau)$ and $\Gamma_{SN}(\omega_n\tau)$ represent the increased variances due to nonzero delay, and are defined by:

$$\Gamma_{\rm PN}(\omega_n\tau) \triangleq \frac{2\zeta\omega_n}{\pi}\int_{-\infty}^{\infty}\left|j\omega + e^{-j\omega\tau}F(\omega)\right|^{-2}d\omega \tag{33}$$

$$\Gamma_{\rm SN}(\omega_n\tau) \triangleq \frac{2\zeta}{\pi(1+4\zeta^2)\omega_n}\int_{-\infty}^{\infty}\left|\frac{F(\omega)}{j\omega + e^{-j\omega\tau}F(\omega)}\right|^2 d\omega. \tag{34}$$

Both are unity at $\tau = 0$ and both are increasing functions of their arguments. The purpose of the functional notation is to emphasize the fact that these factors depend on ω_n and τ only through the product $\omega_n\tau$.

B. Loop Optimization

The loop filter is wholly specified by a single parameter: the natural frequency ω_n. We will optimize the loop natural frequency to minimize σ^2, the phase-error steady-state variance. There are of course other transitory considerations in the design of a PLL, such as the pull-in range, hold-in range, and acquisition time, which we ignore; discussion of these issues can be found elsewhere in the literature [17].

Assuming a zero steady-state mean phase error, σ^2 completely specifies the power penalty and bit-error rate (BER) floor. Using the results of Prabhu [(41) in [18]], we find that standard deviations of $\sigma_{\max} = 2.97°$ and $\sigma_{\max} = 3.78°$ correspond to power penalties of 0.5 and 1 dB at 10^{-9} BER, respectively. We derive laser-linewidth requirements by requiring that $\sigma < \sigma_{\max}$.

We first optimize the loop natural frequency assuming zero delay, and define the parameters α_0 and β_0. We then consider the more general problem of loop optimization assuming nonzero loop delay, and define the parameters α_1 and β_1. Together, these four parameters comprise a simple yet accurate model for specifying the laser-linewidth requirements and corresponding optimal natural frequency for arbitrary delays; this model will be presented in Section IV-C.

1) Zero Delay: When the loop delay is zero, $\Gamma_{PN}(0) = \Gamma_{SN}(0) = 1$, and the optimum natural frequency that minimizes σ^2 in (32) is given by [16]:

$$\omega_{n,0}T = \sqrt{\frac{2\pi\Delta\nu TM/\eta}{1+4\zeta^2}} \triangleq \beta_0 \tag{35}$$

where we introduce β_0 for later use. The resulting minimized variance (with $\omega_n = \omega_{n,0}$) is:

$$\sigma^2 = \sqrt{\frac{1+4\zeta^2}{4\zeta^2}2\pi\Delta\nu T\eta/M}\,. \tag{36}$$

With (36) we can solve $\sigma < \sigma_{\max}$ for $\Delta\nu T$ to arrive at the following zero-delay laser-linewidth requirement:

$$\Delta\nu T < \frac{\sigma_{\max}^4 4\zeta^2 M}{(1+4\zeta^2)2\pi\eta} \triangleq \alpha_0 \tag{37}$$

where again we introduce α_0 for later use.

2) Nonzero Delay: There are no closed-form expressions for the bandwidth expansion factors Γ_{PN} and Γ_{SN} when $\tau > 0$, and so we must resort to numerical techniques to compute the optimal natural frequency. In Fig. 6 we plot σ versus ω_n for feedback delays of 0, $T/2$, and T, and $3.26T$; other parameters are specified in the caption. As pointed out in [6], the effects of the loop delay are two-fold; first, the delay reduces the optimal natural frequency from its zero-delay value, and second, the delay increases the resulting minimized phase error standard deviation. For any set of system parameters $\{\Delta\nu, \tau, T, M\}$, the optimal natural frequency $\omega_{n,\text{opt}}$ can be found numerically by computing the minimum of a curve like those in Fig. 6.

In Fig. 7 we plot the normalized optimal natural frequencies $\omega_{n,\text{opt}}T$ versus normalized laser linewidth for the loops under consideration, with the delay τ taken from Table II. Here and in Section V we assume that $\tau_4 = T$, which makes it easier to compare the performance of the fourth-power loop with the other loops. Also illustrated in Fig. 7 are the optimal natural frequencies assuming zero delay ($\tau = 0$), as specified by (35). These curves emphasize the importance of taking the delay into account when optimizing the loop natural frequency; ignoring loop delay can lead to overestimating the optimal loop natural frequency by as much as 24% for $\Delta\nu T = 10^{-4}$. (In

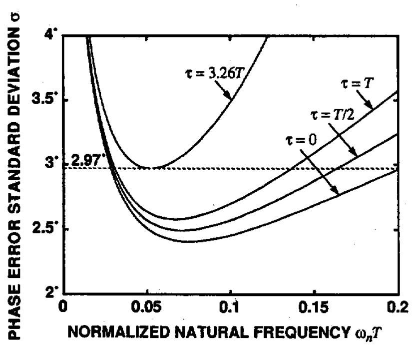


Fig. 6. Phase error standard deviation σ versus $\omega_n T$ for $\tau \in \{0, T/2, T, 3.26T\}$, assuming $M = 45.3$, $\eta = 1/2$, and $\Delta\nu T = 3 \times 10^{-5}$.

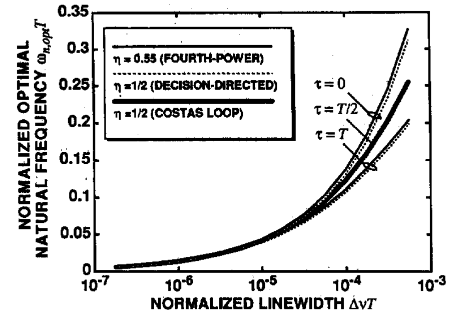


Fig. 7. Optimal (normalized) loop natural frequency versus $\Delta\nu T$.

Section IV-C we will see that larger values of $\Delta\nu T$ are not practical due to an unacceptably-large power penalty.) Delays larger than T will, of course, increase this effect.

3) Large Delay: The inherent loop delay of the decision-directed loop is $\tau = T$. For very high bit-rate systems, however, it is likely that this single-baud delay will be negligible when compared to electronics processing and propagation delays. For this reason, we consider below the loop optimization problem when $\tau \gg T$ [7].

When $\tau = 0$, the shot noise and phase noise contribute equally to the total phase error, so that the first and second terms in (32) are equal. When $\tau > 0$, however, this is no longer true. There are two reasons for this: first, the optimum natural frequency $\omega_{n,\text{opt}}$ decreases as τ increases, enhancing the phase-noise term while diminishing the shot-noise term; and second, $\Gamma_{\text{PN}}(\omega_n\tau)$ grows faster than $\Gamma_{\text{SN}}(\omega_n\tau)$ as τ increases [6], [7]. The net effect is that, for large delays, the phase-noise term dominates and the shot-noise term is negligible.

In the limit of large delay, therefore, the second term of (32) can be neglected, and the total variance approaches:

$$\sigma^2 \to \frac{\pi\Delta\nu\tau}{2\zeta\omega_n\tau}\Gamma_{\text{PN}}(\omega_n\tau) \tag{38}$$

in which case the loop optimization reduces to choosing $\beta = \omega_n\tau$ to minimize $\Gamma_{\text{PN}}(\beta)/\beta$. The optimal value was found numerically to be

$$\beta_1 \triangleq 0.34. \tag{39}$$

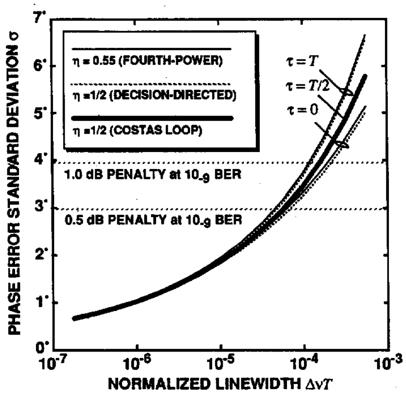


Fig. 8. Theoretical phase error standard deviation versus linewidth-to-symbol-rate ratio.

This value was reported first in [7]. Substituting (38) with $\omega_n\tau = \beta_1$ into $\sigma < \sigma_{\max}$ and solving for $\Delta\nu\tau$ yields the following laser-linewidth requirement for large delays:

$$\Delta\nu\tau < \frac{2\zeta\sigma_{\max}^2}{\pi\Gamma_{\text{PN}}(\beta_1)/\beta_1} \triangleq \alpha_1 \tag{40}$$

where again α_1 will be used later. Numerical calculations show that $\Gamma_{\text{PN}}(\beta_1) \approx 2.26$ [6], [7]. Note that, whenever $\tau \gg T$, the symbol rate $1/T$ is irrelevant and the laser-linewidth requirement depends only on τ.

In the next section we address the problem of specifying laser-linewidth requirements for moderate delays.

C. Analysis Results

The bit-error rate (BER) for an ideal shot-noise limited optical QPSK system is $\text{BER} = Q(\sqrt{M})$ [18], from which we find that the quantum limit of sensitivity for a 10^{-9} BER is $M = 36$ photons/symbol (or 18 photons/bit). We assume that the system is operating 1 dB from the quantum limit, so that $M = 45.3$ photons/symbol. We will allot 0.5 dB of this penalty to system margin, leaving the remaining 0.5 dB for the penalty due to phase error. The designer of the phase-locked loop is thus faced with the problem of keeping the steady-state phase error standard deviation less than 2.97° with a detected signal power of $M = 45.3$ photons per symbol.

We will specify laser-linewidth requirements by plotting σ versus $\Delta\nu T$ with $M = 45.3$. At each point on the curve, we optimize the loop natural frequency, as discussed in Section IV-B, so that the resulting standard deviation σ is as small as possible. The resulting curve will be strictly increasing. The linewidth requirement can then be read off directly from the graph: linewidths yielding $\sigma < 2.97°$ are acceptable, while other linewidths are not.

In Fig. 8 we plot the standard deviation σ of the phase error versus the linewidth-to-symbol-rate ratio $\Delta\nu T$, as predicted by our linear analysis. The thin lines correspond to the fourth-power loop, the dashed lines correspond to the decision-directed loops, and the thick line corresponds to the Costas loop. To emphasize the importance of accounting for delay in the analysis, we include the zero-delay theoretical curves

TABLE III
MINIMAL-DELAY LASER-LINEWIDTH REQUIREMENTS

Loop Type	Maximum Allowable $\Delta\nu T$ Theory, ignoring delay (α_0)	 Theory, with inherent delay	 Simulation, with inherent delay
Discrete-Time Decision-Directed	6.9×10^{-5}	4.9×10^{-5}	5.2×10^{-5}
Analog Decision-Directed	6.9×10^{-5}	4.9×10^{-5}	3.7×10^{-5}
Costas Loop	6.9×10^{-5}	5.7×10^{-5}	2.5×10^{-5}
Fourth-Power	6.3×10^{-5}	4.6×10^{-5}	3.9×10^{-5}

as well, which were computed from (36). We see that the zero-delay analysis underestimates the phase-error standard deviation, particularly at large linewidth-to-symbol-rate ratios. The nonzero-delay curves were computed using (32) with ω_n chosen optimally at each point. The intersection of these curves with the horizontal line at $\sigma = 2.97°$ is the maximum allowable linewidth-to-symbol-rate ratio $\Delta\nu T$.

In the first two columns of Table III we summarize the theoretical results. We emphasize that these numbers depend on our initial assumption of operating at 1 dB above the quantum limit with 0.5 dB of this penalty due to phase error, where both penalties are measured at a 10^{-9} BER. A different signal power, system margin, or BER would lead to different linewidth requirements. In fact, these linewidth requirements can be considered conservative; a practical system is likely to be further than 1 dB from the quantum limit, in which case the higher signal power would reduce the effects of shot noise, thus easing the carrier synchronization task. The damping coefficient is assumed to be $\zeta = 1/\sqrt{2}$. The normalized front-end filter bandwidth for the fourth-power loop is assumed to be $W = 3$, which from (59) with $M = 45.3$ yields $\eta = \eta(M, W) = 0.552$. Also included on the curve is a line at $\sigma = 3.95°$, which corresponds to a 1-dB penalty; this limit can be used if one does not allow for a system margin.

As pointed out by Norimatsu *et al.* in [7], the electrical delays due to amplification and propagation within the loop will not be negligible for high bit-rate systems, and hence must be taken into account when specifying laser-linewidth requirements and optimizing ω_n. To illustrate the effects of increased propagation delay on the laser-linewidth requirement, we plot in Fig. 9 the maximum allowable $\Delta\nu T$ versus τ/T as predicted by the theory of Section IV-B.2 with $\eta = 1/2$ and $M = 45.3$ photons/symbol. At each point on the curves we optimize the loop natural frequency.

As shown in the figure, the laser-linewidth requirement near $\tau = 0$ is $\Delta\nu_{\max} = \alpha_0/T$, as specified by (37). For large loop delays ($\tau \gg T$), on the other hand, the linewidth requirement approaches $\Delta\nu_{\max} = \alpha_1/\tau$, as specified by (40). The exact linewidth requirement for moderate delays cannot be expressed in analytical form; it must be found numerically. Fortunately, however, we have found the following approximation to be surprisingly accurate for all delays:

$$\Delta\nu_{\max} \approx \frac{\alpha_0}{T} \parallel \frac{\alpha_1}{\tau} \tag{41}$$

where $a \parallel b \triangleq ab/(a+b)$ represents a "parallel" addition.

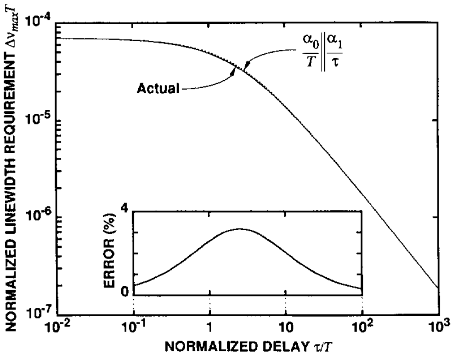


Fig. 9. Effect of propagation delay on laser-linewidth requirement ($\eta = 1/2$); the inset shows the percentage error due to the approximation of (41).

This approximation is shown as a dashed curve in Fig. 9, and is barely discernible from the actual curve. It is asymptotically precise at the two extremes as $\tau \to 0$ and as $\tau \to \infty$, while it provides a maximum error of 3.2% at $\tau \approx 2.6T$.

We found a similar expression for $\omega_{n,\text{opt}}$. When $\tau = 0$ and $\Delta\nu = \alpha_0/T$, the optimal natural frequency from (35) is $\omega_{n,\text{opt}} = \beta_0/T$. On the other hand, when $\tau \gg T$, the optimal natural frequency from (39) is $\omega_{n,\text{opt}} = \beta_1/\tau$. The following approximation is quite accurate for all delays:

$$\omega_{n,\text{opt}} \approx \frac{\beta_0}{T} \parallel \frac{\beta_1}{\tau}. \tag{42}$$

The error approaches zero as $\tau \to 0$ and as $\tau \to \infty$, while the maximum error is 4.3% at $\tau \approx 2.5T$.

The procedure for estimating the laser-linewidth requirement can be summarized as follows. First, estimate the total feedback delay τ. Second, calculate α_0 using (37) and α_1 using (40). Finally, use (41) to calculate the linewidth requirement. To determine the optimal natural frequency at this linewidth, calculate β_0 from (35), β_1 from (39), and use (42).

An alternate interpretation of Fig. 9 is as follows. As shown in Fig. 6, the decision-directed loop with $\Delta\nu T = 3 \times 10^{-5}$ and $M = 45.3$ just barely meets the phase-error requirement (i.e., $\sigma < 2.97°$) when $\tau = 3.26T$. Larger delays yield a larger minimized phase error standard deviation. Therefore, the maximum allowable delay for $\Delta\nu T = 3 \times 10^{-5}$ is $\tau_{\max} = 3.26T$. Different values of $\Delta\nu T$ will yield different values of $\tau_{\max}$. Fig. 9 generalizes this result for the decision-directed loop by showing how $\tau_{\max}$ varies with $\Delta\nu T$; we see that the maximum allowable loop delay (normalized by a baud interval) decreases as the linewidth-to-symbol-rate ratio increases. As an example, consider a 10-Gb/s system: with a 10-kHz beat linewidth, it can tolerate a total loop delay of at most 18 ns, whereas with a 100-kHz beat linewidth, it can tolerate a delay of at most 1.2 ns.

V. SIMULATION RESULTS

To test the validity of our linearized analysis, we have performed accurate computer simulations. These simulations have done more than just verify our analysis, however; they also helped discern the performance discrepancies between the various loops, because they included all of the higher-order effects our analysis ignored. We were able to analyze accurately only the discrete-time decision-directed and the fourth-power loop. We saw how the analog decision-directed loop and the Costas loop were rough approximations to the discrete-time decision-directed loop, and thus we expect their performance to be inferior, but there is no way to determine how inferior using our analysis. The classical approach to this problem is to examine the performance of the loops in the limit of either large or small signal-to-noise ratio (M). For example, Stiffler shows that, as $M \to \infty$, the performance of the Costas loop approaches that of the discrete-time decision-directed loop, while as $M \to 0$, the variance σ^2 for the Costas loop approaches 3/2 times that of the discrete-time decision-directed loop [14]. However, this analysis considered neither laser phase noise nor loop delay. It is likely that, if it did, the asymptotic results would no longer hold. When M is neither zero nor infinite, and when laser phase noise as well as shot noise is affecting the loop performance, a complete analysis seems intractable, and thus computer simulations are quite valuable.

The numerical simulations presented in this paper were performed using the Ptolemy simulation environment [19]. The signals in the complex models presented in Fig. 2(b) and Fig. 4(b) were modeled by their sampled values, where the sampling rate $1/\Delta t$ was chosen as $30/T$ for the decision-directed and Costas loops, and $100/T$ for the fourth-power loop ($1/T$ is the baud rate). The continuous-time filters were modeled in discrete time using the impulse-invariant method. A continuous-time white noise process with PSD N_0 was modeled as a discrete-time white random process with PSD $N_0/\Delta t$.

In Fig. 10 we present our simulation results for the fourth-power loop by plotting σ versus $\Delta\nu T$. For each point on the curve, ω_n was optimized using the procedure described in Section IV-B (see Fig. 7), taking into account the inherent loop delay in each case. The solid curve represents the theoretical predictions, while the squares mark the simulation results. They are in good agreement. The front-end filter $B(s)$ was taken to be Gaussian with $W = 3$, as given in (49), with a total group delay of $\tau_4 = T$.

The simulation results for the decision-directed and Costas loops are presented in Fig. 11. The solid line represents the linearized theory for the discrete-time decision-directed loop. There is again good agreement between this theory and the simulation results for the discrete-time loop, which are marked with triangles. The analog decision-directed loop, marked with circles, is seen to yield a slightly higher phase-error standard deviation than the discrete-time decision-directed loop, as expected. The Costas loop, marked by a cross, performs even worse.

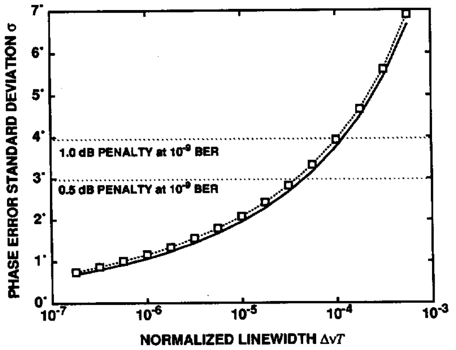


Fig. 10. Fourth-power loop: simulation and theory.

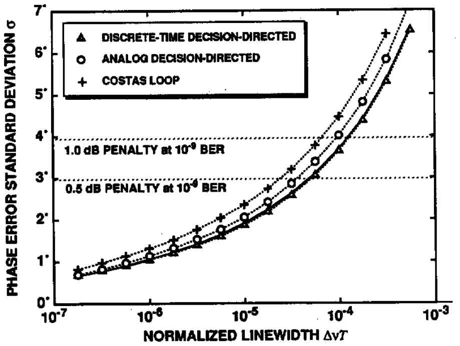


Fig. 11. Decision-directed and Costas loops: simulation and theory.

In Fig. 12 we present all of the simulation results on the same graph, to facilitate comparison. There, we see that the fourth-power loop is inferior to the analog decision-directed loop at low linewidth-to-symbol-rate ratios, while it is superior at high ratios. Overall, the performance discrepancy between these loops is not great. For most noncritical applications, then, it is best to choose the loop based on other considerations, such as ease of implementation.

Based on the theory of Section III-A.3, one might conclude that, because of its smaller inherent delay, the Costas loop would outperform the decision-directed loops. This is not the case, however, for the theory did not account for the performance degradation due to the poor phase-error estimate provided by the Costas loop. Actually, this performance degradation outweighs any performance gain due to decreased delay, with the net effect being a reduced overall performance. Our simulation results support this claim. Simulation results do seem to indicate that the Costas loop behaves as a noisy half-baud delay PLL, rather than as a less-noisy full-baud delay loop. This effect is illustrated in Fig. 12, where we see that the standard deviation of the Costas loop is about 18% larger than that of the analog-directed loop at narrow linewidths

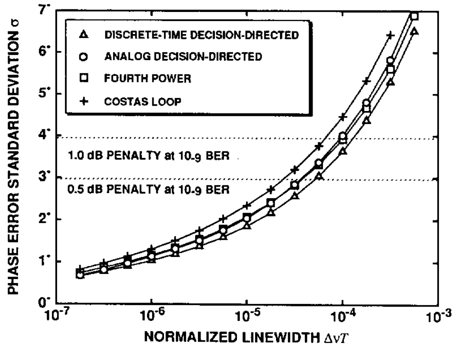


Fig. 12. Comparison of different simulation results.

(near $\Delta\nu T = 10^{-7}$), while it is only about 10% larger at wider linewidths (near $\Delta\nu T = 10^{-3}$). Thus, the performance discrepancy between the analog decision-directed loop and the Costas loop diminishes as the laser linewidth increases.

VI. SUMMARY AND CONCLUSIONS

We have presented a detailed analysis of four phase-locked loops suitable for synchronous detection of optical QPSK: the discrete-time decision-directed loop, analog decision-directed loop, the Costas loop, and the fourth-power loop. We showed how the analog decision-directed loop and the Costas loop can be viewed as rough approximations to the discrete-time decision-directed loops, thus explaining their slightly inferior performance.

Our linear analysis used a common model for all four loop types. The only difference in each case was the noise factor η, which represents the effects of the phase-error estimator on the additive shot noise. For the fourth-power loop, we found that $\eta \approx 0.55$ for a typical front-end filter, while for the decision-directed and Costas loops, we found that $\eta = 1/2$. Assuming a proportional-plus-integral loop filter, we found a closed-form expression for the optimal loop natural frequency $\omega_{n,\mathrm{opt}}$ when the loop delay was zero, and we presented an analytical procedure for estimating the optimal natural frequency for arbitrary loop delays.

To corroborate our linearized analysis, and to examine the effects of those factors beyond its grasp, we also presented computer simulations. The agreement between these simulations and our analysis was good. In some sense, the simulations represent a more thorough treatment of the problem, since they account for the higher order effects neglected in the linear analysis. We thus view the minimal-delay linewidth requirements arrived at through simulation, which are summarized in Table III, as more accurate than those based on theory. Since our simulations considered only the inherent delays in the loop, (41) should be used to determine the laser-linewidth requirement when feedback delays are significant.

All four loops have very similar performance and laser-linewidth requirements. The system designer may therefore use other criteria, such as ease of implementation, when choosing a loop type.

APPENDIX A
DERIVATION OF NOISE FACTOR η FOR DECISION-DIRECTED LOOP

The PSD of the shot noise $n(t)$ in the decision-directed loop in Fig. 2(b) is white with PSD $S_n(\omega) = T/M$, while we claimed in Section III-A3 that the additive noise $w(t)$ in Fig. 5 is approximately white with PSD $S_w(\omega) = \eta T/M$, where (for the decision-directed loops) $\eta = 1/2$. In this Appendix we justify this approximation, and derive $\eta = 1/2$.

Our starting point is (19) in Section II, where we defined $w(t)$ by:

$$w(t) = \mathrm{Im}\{\tilde{n}(t)\} \tag{43}$$

where $\tilde{n}(t)$ is the piecewise-constant function defined as follows:

$$\tilde{n}(t) \triangleq n_k e^{-j\theta_k} \quad \text{for} \quad t \in [kT, (k+1)T] \tag{44}$$

with n_k defined by (12), and $\theta_k \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$ representing the modulation data. In other words, $\tilde{n}(t)$ represents the noise-cross-signal term that arises due to the baseband mixer in the phase-error estimator of Fig. 2(b). From (12) we see that the discrete-time sequence $\{n_k\}$ is a white sequence with variance $1/M$. Similarly, assuming the data symbols are all equally likely and chosen independently, then $\{e^{-j\theta_k}\}$ is a white sequence with unity variance. Furthermore, because they are mutually independent, the produce sequence $\{n_k e^{-j\theta_k}\}$ is white with variance equal to $1/M$; the product of the two variances. Now $\tilde{n}(t)$ is the piecewise-constant continuous-time function taking the value $n_k e^{-j\theta_k}$ during the k-th baud interval; therefore, its autocorrelation function is given by [20]:

$$R_{\tilde{n}}(\tau) = \frac{1}{M}\Lambda(\tau) \tag{45}$$

where the triangle function $\Lambda(\tau)$ is zero for $|\tau| > T$ and $\Lambda(\tau) \triangleq (1 - |\tau|/T)$ for $|\tau| \le T$. Taking the Fourier transform yields a PSD of

$$S_{\tilde{n}}(\omega) = \frac{T}{M}\left(\frac{\sin(\omega T/2)}{\omega T/2}\right)^2. \tag{46}$$

In Fig 13 we plot this PSD versus frequency. Also indicated on the graph are the optimal (zero-delay) loop bandwidths for linewidth-to-symbol-rate ratios of $\{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$. As we see, the power spectrum is nearly flat for frequencies within all bandwidths; the variation of the PSD over even the largest bandwidth, which corresponds to a beat linewidth of $\Delta\nu T = 10^{-3}$, is less than 0.16 dB. A linewidth this large will not be used in most cases, however. More likely, the linewidth will satisfy the requirements specified in Table III, and thus $\Delta\nu T$ will be no larger than about 3×10^{-5}.

Since the PSD of $\tilde{n}(t)$ is nearly flat for all frequencies of interest, we can approximate it by its value at zero frequency

$$S_{\tilde{n}}(\omega) \approx S_{\tilde{n}}(0) = T/M. \tag{47}$$

Finally, $w(t)$ is defined as $\mathrm{Im}\{\tilde{n}(t)\}$, and therefore it is also approximately white, but with half the PSD

$$S_w(\omega) = \frac{1}{2}S_{\tilde{n}}(\omega) \approx \frac{1}{2}T/M \tag{48}$$

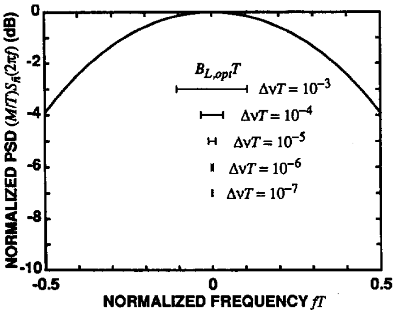


Fig. 13. Normalized power spectrum $(M/T)S_{\tilde{n}}(\omega)$.

from which we conclude that $\eta = 1/2$.

APPENDIX B
DERIVATION OF NOISE FACTOR η FOR FOURTH-POWER LOOP

The PSD of the shot noise $n(t)$ in Fig. 4(b) is white with PSD $S_n(\omega) = T/M$, while, as stated in Section III-B3, the additive noise $w(t)$ in Fig. 5 is approximately white with PSD $S_w(\omega) = \eta T/M$. In this appendix we justify this approximation, and specify η for the fourth-power loop.

Before proceeding, we must specify the front-end filter. In [3] an ideal brick-wall low-pass filter is assumed, which is not very practical. Here, we assume a Gaussian filter

$$B(\omega) = e^{-\omega^2/(2\beta^2)} \tag{49}$$

with a normalized bandwidth of

$$W \triangleq \beta T/\sqrt{4\pi}. \tag{50}$$

It is related to the -3-dB bandwidth f_c by the following relationship: $f_cT = W\sqrt{(\log 2)/\pi}$. In Section V we assume that $W = 3$, which corresponds to a -3-dB bandwidth of about 1.4 times the symbol rate. For now, we assume that the group delay of this front-end filter is zero, i.e., $\tau_4 = 0$.

The fourth-power loop is difficult to linearize; we will use an approach similar to that given in [3]. From (27), the output of the fourth-power block, $z(t)$, is given by:

$$z(t) = x^4 + \sum_{i=0}^{3} \binom{4}{i} x^i y^{4-i} \tag{51}$$

where $x(t)$ is defined by (25) and $y(t)$ is the filtered noise with PSD given in (26). We will find the autocorrelation function $R_z(\tau)$ of $z(t)$ and decompose it into two terms:

$$R_z(\tau) = R_{\text{signal}}(\tau) + R_m(\tau), \tag{52}$$

with the first term corresponding to the "signal" term x^4 and the second "noise" term $R_m(\tau)$ corresponding to all of the noise and signal-cross-noise terms. The bold assumption we then make is to decompose $z(t)$:

$$z(t) \approx x^4 + m(t) \tag{53}$$

where we assume the additive noise $m(t)$, which has autocorrelation function $R_m(\tau)$, is independent of the signal term x^4. The best justification for this assumption is found in Section V, where we see that our analysis, based on this assumption, accurately predicts the results of our numerical simulations.

If we assume that $E\left[e^{jm(\varepsilon(t)-\varepsilon(t+\tau))}\right] \approx 1$ for $m \in \{0, 1, 2, 3, 4\}$ and all τ, then substituting from (51) yields the following expression for $R_z(\tau) = E\left[z(t)z(t+\tau)^*\right]$:

$$R_z(\tau) = \alpha_0 g_4(\tau) + \Lambda(\tau)\sum_{k=1}^{3}\alpha_k\rho^k(\tau) + \alpha_4\rho^4(\tau) \tag{54}$$

where $\alpha_k \triangleq \binom{4}{k}^2 k!(W/M)^k$, where $\Lambda(\tau)$ is the triangular autocorrelation of the QPSK data signal $e^{j\theta(t)}$ (assuming equally likely and independent symbols) [20], $\Lambda(\tau)$ is zero for $|\tau| > T$ and $\Lambda(\tau) \triangleq (1 - |\tau|/T)$ for $|\tau| \le T$, where $(W/M)\rho(\tau)$ is the autocorrelation of the filtered shot noise $y(t)$, and where, from (26) and (49), $\rho(\tau)$ is given by

$$\rho(\tau) = e^{-\beta^2\tau^2/4}. \tag{55}$$

All of the terms but the first in (54) make up the "noise" autocorrelation

$$R_m(\tau) = \Lambda(\tau)\sum_{k=1}^{3}\alpha_k\rho^k(\tau) + \alpha_4\rho^4(\tau). \tag{56}$$

Since $w(t) = \frac{1}{4}\text{Im}\{m(t)\}$, the autocorrelation function of $w(t)$ is $R_w(\tau) = R_m(\tau)/32$, or

$$R_w(\tau) = \Lambda(\tau)\sum_{k=1}^{3} a_k(W/M)^k\rho^k(\tau) + a_4(W/M)^4\rho^4(\tau), \tag{57}$$

where $a_k \triangleq \binom{4}{k}^2 k!/32$.

$S_w(\omega)$ is the Fourier transform of the autocorrelation function in (57), and hence is not white. As shown in Fig. 14, however, $S_w(\omega)$ is nearly white for frequencies less than the loop bandwidth, so we will approximate it by its value at zero frequency

$$S_w(\omega) \approx S_w(0) = \frac{T}{M}\eta(M, W) \tag{58}$$

where from (57) we find

$$\begin{aligned}\eta(M, W) = \int_{-\infty}^{\infty} \frac{M}{T}R_w(\tau)\,d\tau &= \frac{3}{8}(W/M)^3 + 3(W/M)^2 \\ &\cdot \left[\frac{1}{\sqrt{3}}\left[1 - 2Q\left(W\sqrt{6\pi}\right)\right] - \frac{1}{3\pi W}\left(1 - e^{-3\pi W^2}\right)\right] \\ &+ \frac{9}{4}(W/M)\left[\frac{1}{\sqrt{2}}\left[1 - 2Q\left(W\sqrt{4\pi}\right)\right] - \frac{1}{2\pi W}\left(1 - e^{-2\pi W^2}\right)\right] \\ &+ \frac{1}{2}\left[\left[1 - 2Q\left(W\sqrt{2\pi}\right)\right] - \frac{1}{\pi W}\left(1 - e^{-\pi W^2}\right)\right].\end{aligned} \tag{59}$$

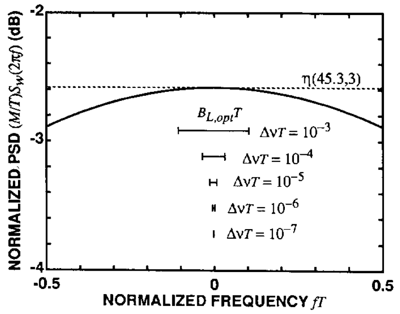


Fig. 14. Normalized power spectrum $(M/T)S_w(\omega)$ and its approximation $\eta(M,W) = (M/T)S_w(0)$ with $M = 45.3$ and $W = 3$.

Here, $Q(x)$ is the standard Gaussian error function (equal to the probability that a zero-mean unit-variance Gaussian random variable is greater than x). We thus see that the noise factor for the fourth-power loop is an increasing function of bandwidth W and a decreasing function of signal-to-noise ratio M. Substituting $M = 45.3$ and $W = 3$ yields $\eta \approx 0.552$. □

ACKNOWLEDGMENT

The authors are grateful to M. Audeh and A. Porter for their contributions in the initial stages of the analysis and simulation, to S. Norimatsu for pointing out the related work in [2], and to the Ptolemy group at U.C. Berkeley for providing the simulation environment.

REFERENCES

[1] S. Yamazaki and K. Emura, "Feasibility study on QPSK optical-heterodyne detection system," *J. Lightwave Technol.*, vol. 8, no. 11, pp. 1646–1653, Nov. 1990.

[2] S. Norimatsu and K. Iwashita, "Linewidth requirements for optical synchronous detection systems with nonnegligible loop delay time," *J. Lightwave Technol.*, vol. 10, no. 3, pp. 341–349, Mar. 1992.

[3] J. J. Spilker, *Digital Communications by Satellite.* New York: Prentice Hall, 1977.

[4] L. G. Kazovsky, "Decision-directed phase-locked loop for optical homodyne receivers: Performance analysis and laser linewidth requirements," *J. Lightwave Technol.*, vol. LT-3, no. 6, pp. 1238–1247, December 1985.

[5] L. G. Kazovsky, "Balanced phase-locked loops for optical homodyne receivers: Performance analysis, design considerations, and laser linewidth requirements," *J. Lightwave Technol.*, vol. LT-4, no. 2, pp. 182–195, Feb. 1986.

[6] M. A. Grant, W.C. Michie, and M.J. Fletcher, "The performance of optical phase-locked loops in the presence of nonnegligible loop propagation delay," *J. Lightwave Technol.*, vol. LT-5, no. 4, pp. 592–597, Apr. 1987.

[7] S. Norimatsu and K. Iwashita, "PLL propagation delay-time influence on linewidth requirements of optical PSK homodyne detection," *J. Lightwave Technol.*, vol. 9, no. 10, pp. 1367–1375, Oct. 1991.

[8] F. Derr, "Comparison of electrical and optical BPSK and QPSK systems," *J. Opt. Commun.*, vol. 10, no. 4, pp. 127–130, 1989.

[9] L. G. Kazovsky, "Optical heterodyning versus optical homodyning: A comparison," *J. Opt. Commun.*, vol. 6, no. 1, pp. 18–24, 1985.

[10] F. Derr, "Optical QPSK homodyne transmission of 280 Mb/s," *Electron. Lett.*, vol. 26, no. 6, pp. 401–403, Mar. 1990.

[11] F. Derr, "Optical QPSK transmission system with novel digital receiver concept," *Electron. Lett.*, vol. 27, no. 23, pp. 2177–2179, Nov. 1991.

[12] S. Norimatsu, K. Iwashita, and K. Noguchi, "Eight-Gb/s quadrature-PSK optical homodyne detection experiment using external cavity laser diodes," *OFC '92* (San Jose), Feb. 1992.

[13] J. R. Barry and E. A. Lee, "Performance of coherent optical receivers," *Proc. IEEE*, vol. 78, no. 8, pp. 1369–1394, Aug. 1990.

[14] J. J. Stiffler, *Theory of Synchronous Communications.* Englewood Cliffs, New Jersey: Prentice Hall, 1971.

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

[16] L. G. Kazovsky, "Performance analysis and laser linewidth requirements for optical PSK heterodyne communications systems," *J. Lightwave Technol.*, vol. LT-4, no. 4, pp. 415–425, Apr. 1986.

[17] F. M. Gardner, *Phaselock Techniques.* New York: John Wiley & Sons, 1979.

[18] V. K. Prabhu, "PSK Performance With Imperfect Carrier Phase Recovery," *IEEE Trans. Aerospace and Electron. Syst.*, vol. AES-12, no. 2, pp. 275–285, Mar. 1976.

[19] J. Buck, S. Ha, E. A. Lee, and D. G. Messerschmitt, "Ptolemy: A mixed-paradigm simulation/prototyping platform in C++," *C++ Conf.* (Santa Clara, CA), Nov. 1991.

[20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, 1984.

John R. Barry received the B. S. degree summa cum laude and with departmental honors in 1986 from the State University of New York at Buffalo, the M.S. degree in 1987 from the University of California at Berkeley (U. C. Berkeley), and the Ph.D. degree in 1992 from U. C. Berkeley, all in electrical engineering. His dissertation investigated the feasibility of broadband wireless communications using non-directed infrared radiation. He was awarded the 1992 David J. Griep Memorial Prize by U. C. Berkeley for scholarly achievement.

Since 1992 he has been an Assistant Professor with the School of Electrical Engineering at the Georgia Institute of Technology, where his research interests include digital communication theory, fiber-optic communications, and broadband wireless communications.

Joseph M. Kahn is Assistant Professor in the Department of Electrical Engineering and Computer Sciences at UC Berkeley. He received the AB Physics (1981), MA Physics (1983), and Ph.D. Physics (1986) from UC Berkeley. His Ph.D. thesis was entitled "Hydrogen-Related Acceptor Complexes in Germanium."

From 1987 to 1990 he was a Member of Technical Staff in the Lightwave Communications Research Department of AT&T Bell Laboratories in Holmdel, NJ, where he performed research on multi-gigabit-per-second coherent optical fiber transmission systems and related device and subsystem technologies. He demonstrated the first BPSK-homodyne optical fiber transmission system, and achieved world records for receiver sensitivity in multi-gigabit-per-second systems. He joined the faculty of UC Berkeley in 1990, where his research interests include optical fiber communication networks and transmission systems, indoor local-area networks using free-space optical communication, and optical interconnects in digital systems.

Dr. Kahn is a recipient of the National Science Foundation Presidential Young Investigator Award, and is a member of IEEE.